Disparities in Geographic Accessibility of National Cancer Institute Cancer Centers in the United States

Yanqing Xu[a], **Cong Fu**[b], **Xun Shi**[c], **Tracy Onega**[d], **Fahui Wang**[e]

[a]Department of Geography & Planning, University of Toledo, Toledo, OH 43606

[b]Metro Engineering Solutions, Corby Energy Services, 2021 S. Schaefer Hwy, Detroit, MI 48217

[c]Department of Geography, Dartmouth College, Hanover, NH 03755

[d]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, 1 Rope Ferry Rd, Hanover, NH 03755

Abstract

The National Cancer Institute (NCI) Cancer Centers form the backbone of the cancer care system in the United States since their inception in the early 1970s. Most studies on their geographic accessibility used primitive measures, and did not examine the disparities across urbanicity or demographic groups. This research uses an advanced accessibility method, termed "2-step floating catchment area (2SFCA)" and implemented in Geographic Information Systems (GIS), to capture the degree of geographic access to NCI Cancer Centers by accounting for competition intensity for the services and travel time between residents and the facilities. The results indicate that urban advantage is pronounced as the average accessibility is highest in large central metro areas, declines to large fringe metro, medium metro, small metro, micropolitan and noncore rural areas. Population under the poverty line are disproportionally concentrated in lower accessibility areas. However, on average Non-Hispanic White have the lowest geographic accessibility, followed by Hispanic, Non-Hispanic Black and Asian, and the differences are statistically significant. The "reversed racial disadvantage" in NCI Cancer Center accessibility seems counterintuitive but is consistent with an influential prior study; and it is in contrast to the common observation of co-location of concentration of minority groups and people under the poverty line.

Keywords

Geographic access; Cancer care; National Cancer Institute Cancer Centers; Disparity; 2-step floating catchment area (2SFCA) method; weighted ordinary-least-squares regression

Introduction

Cancer is a leading cause of death in the United States, only second to heart disease [1]. Adequate geographic access to cancer care services is a major factor for ensuring timely and appropriate care and improving outcomes for patients [2,3]. Research suggests that longer

[e]*Corresponding author*; Department of Geography & Anthropology, Louisiana State University, Baton Rouge, LA 708083.

travel time to cancer care services increases risk of advanced cancer [e.g., 4], reduces utilization of certain therapy [e.g., 5], and limits enrollment in clinical trials [e.g., 6]. Access barriers are especially pronounced among minorities, individuals with low socioeconomic status, and residents in certain geographic areas [7]. Regardless of quality and effectiveness, individuals with fewer resources are inclined to choose care according to geographic proximity and accessibility [8].

There is a great deal of variability in geographic accessibility of quality cancer care [9]. Uneven distributions of cancer care facilities and population result in geographic disparity in accessibility, exemplified by presence of plentiful services in some areas and paucity or lack of service in others [10]. Inner city neighborhoods tend to have greater physician supply, shorter distances to facilities, and more transportation options when compared to suburban and rural ones [11]. People who live in rural areas may suffer from fewer preventive and screening services, lack of quality care during and after treatment [12], and a rising trend in rural hospital closures [13]. This is termed "*geographic disparity*." Furthermore, studies have reported that disproportionally higher numbers of racial and ethnic minorities often suffer from poor access to health care including cancer care [14], commonly referred to as "*racial disparity*." Disparities in cancer care accessibility among socio-demographic groups contribute to corresponding differences in cancer care and outcomes [15]. Geographic and racial disparities are also interconnected, and both contribute to disparities in cancer outcomes such as late-stage cancer risk and mortality [16,17].

The Cancer Centers designated by the National Cancer Institute (NCI) in the U.S. (hereafter referred to as "NCI Cancer Centers") have demonstrated "scientific excellence and the capability to integrate a diversity of research approaches to focus on the problem of cancer." There are currently a total of 69 NCI Cancer Centers, located in 35 states and the District of Columbia [18]. The NCI Cancer Centers are not the only cancer care providers in the U.S. Other specialized cancer care facilities also include (1) the NCI Cancer Center satellite facilities, (2) the Community Clinical Oncology Programs (CCOPs), and (3) Academic Medical Centers (AMCs) in the Council of Teaching Hospitals and Health Systems (COTH). Nevertheless, the NCI Cancer Centers provide the highest quality in cancer care settings, including cancer control, clinical care, and a combination of basic, population sciences, and clinical research [19]. Patients cared for by the NCI Cancer Centers have significantly lower mortality rates [20]. Furthermore, given the $281 million allocated to the NCI Cancer Center program in 2014 [21], maximizing the potential benefits to populations should be a key goal. Therefore, it is important to examine access issues including geographic accessibility of the NCI Cancer Centers and identify underserved areas and demographic groups.

The origin of NCI Cancer Centers stems from the 1971 National Cancer Act, but more formally came into being in 1975 when the Senate approved the establishment of at least 30 comprehensive cancer centers "to permit an estimated 80 percent of the U.S. population access within a reasonable driving distance", which was not specified [22]. Thus, geographic accessibility was a stated federal intention. A common measure of geographic accessibility is travel time or distance from the patient's location to the nearest facility [2,23–25]. A recent study [26] used travel time to the nearest NCI Cancer Centers to measure their geographic access and significant disparities among racial groups and between rural and

urban areas. However, travel time or travel distance does not capture the number of facilities from which the patient can choose [27], and how available each facility is. Alford-Teaster et al. [28] argued that the common assumption using travel time to the closest facility to assess geographic accessibility may depend on the abundance of the healthcare service. Henry et al. [29] concluded that travel time to diagnosing facility was not a determinant of late stage of breast cancer and other factors such as facility capacity, insurance acceptance, and travel costs can affect access. When the capacity of service providers is less a concern, such an approach is acceptable. In most cases, accessibility measures need to account for both supply and demand because of scarcity of supply.

Prior to the widespread use of GIS, the simple supply-demand ratio method computes the ratio of supply vs. demand in an area (usually an administrative unit such as township or county) to measure accessibility. For example, the U.S. Department of Health and Human Services (DHHS) uses the population-to-physician ratio within a "rational service area" (most as large as a whole county or a portion of a county or established neighborhoods and communities) as a basic indicator for defining physician shortage areas [30]. The simple supply-demand ratio method has at least two shortcomings. First, it cannot reveal the detailed spatial variations within an area unit (usually large). Secondly, it assumes that the boundaries are impermeable, i.e., residents do not seek care beyond the area unit's borders [31]. The past decade has seen major advances in GIS-based models of geographic accessibility along this line of work. The most popular methods are the "two-step floating catchment area (2SFCA) method" [32] and its various refinements. In essence, the 2SFCA method measures accessibility as a ratio of supply capacity (e.g., hospital bed size) to demand amount (e.g., number of patients), and the two interact only within a reasonable distance or travel time range. It first assesses each hospital's "availability" as the ratio of its capacity to its surrounding (within-the-spatial-range) demand, and then sums up these ratios for hospitals around (within the range of) a residential location to represent the accessibility for that location. By doing so, it accounts for the match ratio between supply and demand as well as their interaction within a spatial range. More recently, various methods along this line of work have been synthesized into a generic model, termed "generalized 2SFCA or G2SFCA" [33; see the methods section for detail]).

The primary objective of this study is to assess geographic accessibility of NCI Cancer Centers in the conterminous U.S. by using the improved 2SFCA method, and examine the disparities across geographic areas and socio-demographic groups. The results may shed light on possible mitigation strategies of adjusting related resource allocation and maximizing the benefits of NCI Cancer Centers.

Study area, data sources, and variables

The study area is the 48 contiguous states and Washington D.C. excluding Hawaii and Alaska (hereafter referred to as "conterminous U.S."). In Hawaii or Alaska, the accessibility issue is based on substantively different travel modes. Specifically, rather than travel based on road networks, air and water travel are dominant in Hawaii and Alaska, and these travel time measures are beyond the scope of this study to accurately measure. The accessibility measures of the conterminous U.S. need to account for complex interaction between supply

and demand across state borders, and thus can benefit more from the advanced methods to be discussed in the study.

Three variables are needed in defining geographic accessibility to cancer care: supply, demand and the geographic relationship between them. In this study, the supply is the NCI Cancer Centers and their corresponding capacities (i.e., numbers of beds), the demand for cancer care is potential cancer patients across geographic areas (e.g., counties or census tracts) in the U.S., and the link is travel time between them.

On the supply side, this study considers the 58 NCI Cancer Centers that currently provide care to patients in the conterminous U.S. as existing NCI Cancer Centers. Addresses of the NCI Cancer Centers were based on the NCI website [34]. Data of the number of staffed hospital beds were extracted from the American Hospital Directory (http://www.ahd.com/ state_statistics.html) and websites of individual hospitals. The integrated data set, including their geographic locations and existing beds counts, is shown in Figure 1. Most of the NCI Cancer Centers (NCI Cancer Centers) are in the east part of the country, and there is a cluster of NCI Cancer Centers in the New England Area. A few centers are in west and the Great Plains expect for the clusters around San Francisco and Los Angeles in California. There are three centers in the three largest cities in Texas (Houston, Dallas and San Antonio). The sizes of these centers vary a great deal. Among the 58 NCI Cancer Centers, the smallest one is the USC Norris Comprehensive Cancer Center in Los Angeles with 26 staffed beds, and the largest one is the Albert Einstein Cancer Center at the Montefiore Medical Center in New York with 1409 staffed beds.

On the demand side, we use the analysis units of both county and census tract to examine possible presence of *modifiable areal unit problem (MAUP)* [35]. In other words, we are interested in assessing whether the study results are stable when the analysis unit changes from one to another. For the purpose of illustrating the methodology, the number of cancer patients was estimated using the 2010 Census data (http://www.census.gov/) and the nationwide standardized cancer prevalence rates by age, sex, and race and ethnicity. The number of potential patients in each demographic group in a census tract was calculated by multiplying the population in that category by the corresponding cancer rate. The total estimate was the sum of patients from all categories [9]. To more accurately represent the location of each county, we calculated population-weighted centroid for each county based on the 2010 census tract data. Similarly, we used population-weighted centroid for each census tract based on the 2010 census block data. The study area contains 3,138 counties, 22 of which have zero population and thus not included in the analysis. The remaining 3,109 counties represent the demand side at the county level. There are 72,238 census tracts in the contiguous U.S., 210 of which have no population and thus no potential cancer patients. The 72,028 census tracts with non-zero patients form the demand side at the census tract level (Figure 1).

We are aware of the spatial and temporal variability of cancer rates. More accurate estimate of cancer care demand will require the cancer prevalence data by age, sex, and race and ethnicity at sharper geographic resolutions from the North American Association of Central

Cancer Registries (NAACCR) (www.naaccr.org) data and SEER*Stat (http://seer.cancer.gov/seerstat/). Future research will provide refined and more in-depth analysis of the issue.

The travel time between the centroid of each census tract/county and each hospital was computed using the ArcGIS Network Analyst module. The road network data for this task were extracted from the 2012 ESRI StreetMap USA data that came with the ArcGIS 10.1 release. We considered only the major roads, including interstate, the U.S. and state highways, due to the computational limitation of both software and hardware. The estimation of travel time assumes that travelers take the shortest path and follow the speed limit posted on each road section. This approach is adequate for capturing the travel distance between patients and hospitals at the national scale for planning and public policy analysis. In addition, it assumes that patients seeking the specialized cancer care in NCI Cancer Centers travel by private vehicles. Conceivably, people may also choose other transportation modes such as by air or railway. The former incurs considerably high financial cost and the latter is very limited in the U.S., and neither is considered by this study.

Demographic information including racial-ethnic composition is extracted from the 2010 Census. According to the categories classified from the ESRI Demographic and Business Data List, non-Hispanic White, non-Hispanic Black, non-Hispanic Asian and Hispanic are included in this research [36]. Poverty data is from the 2008–2012 American Community Survey (ACS) 5-Year Summary File. Poverty rate is the estimated percent of people of all ages in poverty, with the average poverty rate is 16.34% at the county level and 15.93% at the census tract level.

2013 NCHS Urban–Rural Classification Scheme for Counties prepared by the National Center for Health Statistics [37] was used to define urbanicity at the county level. There are six urban–rural categories in a descending order of urbanicity such as large central metro, large fringe metro, medium metro, small metro, micropolitan and noncore. For the analysis at the census tract level, this research uses the 2010 Census Urban and Rural Classification [38]: a census tract or census block is in an urbanized area (UA) (50,000 or more people), an urban cluster (UC) (at least 2,500 and less than 50,000 people), or rural. Here a census tract is classified as UA, UC or rural if its centroid falls within an UA, UC or rural area, respectively. To examine possible variability of disparities across geographic regions, we further divided the study areas into four census regions: Northeast, Midwest, South, and West (see Figure 2), and repeated the above analysis at the census tract level in each region. Since the study area was the conterminous U.S., the West region in this study did not include Hawaii or Alaska.

Methods: measure of geographic accessibility

As discussed in the introduction, the most recent advancement in measures of geographic accessibility is synthesized into the generalized 2SFCA method that count for complex interactions between supply and demand. The method is summarized as:

$$A_i = \sum_{j=1}^{n} \left[s_j f(d_{ij}) / \left(\sum_{k=1}^{m} D_k f(d_{ij}) \right) \right] \tag{1}$$

where A_i is the accessibility to NCI Cancer Centers at county or census tract i; S_j is the number of staffed beds representing the capacity of an NCI Cancer Center (supply) j; D_k is the estimated number of patients at county or census tract (demand) k; d is the travel time between them; and n and m are the total numbers of supply locations and demand locations, respectively. For the distance decay function $f(d)$, this research adopts a decay schema with discrete values based on a Gaussian function as suggested by Shi [9]. More specifically, the interaction between a patient and a hospital declines with travel time, and six discrete values (1.00, 0.91, 0.69, 0.43, 0.22, 0.10, 0.015) are used to capture the interaction intensity within various travel time ranges (0–30, 30–60, …, 150–180 minutes). The three-hour cap for travel time is almost the limit for a patient to travel to a hospital (one way), obtain some service and return home in a single day. The chosen Gaussian function resembles a bell-shaped curve, widely believed to capture the spatial behavior of hospital visits [9]. Future work will analyze the actual cancer center visitation data to derive the best fit distance decay function[e.g., 39].

In essence, Equation (1) first captures the availability of a facility as the ratio of its capacity and surrounding demand, and the latter is discounted by the distance (travel time) decay effect; and then measures the accessibility of patients in a residential location by summing up the total availability of their surrounding facilities, and the latter is once again discounted by the distance (travel time) decay effect. The derived accessibility value is basically the ratio of supply capacity (i.e., number of beds) to demand (i.e., number of potential patients). In fact, it is proven that the weighted average accessibility score across the whole study area is the ratio of total supply to total demand [40,31]. In practice, since the ratios in our case are very small, the values are multiplied by 1,000 and can be interpreted as number of hospital beds per 1,000 potential patients. A higher A_i value indicates better accessibility.

Figure 2 shows the geographic accessibility at the census tract level in the conterminous U.S. The accessibility index for every census tract is based on the access to all the NCI cancer centers, not only to the closest facility or facilities in the states, which overcome the limitation of impermeable boundaries mentioned in the introduction. For the most part, accessibility exhibits a concentric pattern around the cancer centers, however, the scores vary for areas within similar travel ranges from individual hospitals as a result of (1) specific supply-demand match ratio and (2) whether the effect overlaps with adjacent hospitals.

Results

Variation of accessibility by urbanicity

Examining the variability of health care environment across geographic areas of different urbanicity is a common theme in public health studies [e.g., 41]. This section begins with analyzing how geographic accessibility of NCI Cancer Centers differs across areas of various urbanicity levels.

We first examine the variability at the county level across the six NCHS categories discussed in the previous section. A simple regression model with dummy variables is formulated to examine whether accessibility varies significantly across urban-rural categories [42]. The variable of interest, accessibility value, defines the dependent variable in the regression; and

the independent variables are the dummy variables that code the urban-rural categories. For instance, five dummy variables are used to code six urbanicity categories at the county level: the category "large central metro" is coded as $x_1=x_2=x_3=x_4=x_5=0$, referred to as *reference category*; the category "large fringe metro" is coded as $x_1=1$, $x_2=x_3=x_4=x_5=0$; the category "medium metro" is coded as $x_2=1$, $x_1=x_3=x_4=x_5=0$; and so on. The model is written as

$$A = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 \tag{2}$$

where A is accessibility score in a county, the constant term b_0 is the average score in areas of *reference category* (i.e., "large central metro"), the coefficient b_1, b_2...or b_5 is the difference of average scores between reference category and each of the other five categories, and the t values for the corresponding coefficients indicate whether the average of accessibility score of a category differs from that of the reference category significantly. It is equivalent to the traditional ANOVA, but considered more straightforward in implementation and interpretation.

The results are reported in Table 1. Taking the model for the county level as an example, the intercept from the regression model (8.85) is the average accessibility for the reference category (i.e., large central metro counties), and the coefficient for a category is the difference between the average accessibility of the reference category and this category. For instance, the coefficient −1.12 for large fringe metro counties indicates that the average accessibility of large fringe metro is 1.12 below that of large central metro; moreover, its associated t-statistic value (−3.74) indicates that the difference is statistically significant. As the corresponding coefficient becomes more and more negative from large fringe metro, to medium metro, small metro, micropolitan, and non-core, the average accessibility keeps declining. Similarly, the other t-values all indicate that the difference in average accessibility between the reference category (large central metro counties) and any of the five other categories is statistically significant.

For the central tracts, there are only three categories (Urbanized areas as reference category, urban clusters, and rural). A similar regression (with two dummy explanatory variables) is implemented. Based on the result from Table 1, average accessibility in Urbanized Areas (UA) is higher than that in Urban Clusters (UC) or rural areas, and the differences are statistically significant. However, the gap between UA and UC is larger than the gap between UA and rural areas, and that is to say, the average accessibility in rural areas is rather higher than that in UC, i.e., and the order of average accessibility is UA > rural > UC. The reversal in accessibility order between UC and rural (thought the difference is small) is somehow surprising at the census tract level, while the accessibility declines consistently with decreasing urbanicity at the county level. The analysis by region indicates that the aforementioned findings are consistent across all four regions (results not presented here). In other words, in each region, the average accessibility declines from large fringe metro, to medium metro, small metro, micropolitan, and non-core; and the highest average accessibility is in UA, and then rural and finally UC.

Disparities of geographic accessibility among demographic groups

We now turn the attention to examining disparities of accessibility across major racial-ethnical groups: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian (hereafter referred to as White, Black, and Asian, receptively), and Hispanic; and by socioeconomic status (here poverty status is chosen as an example). Note that the accessibility index is based on location such as census tract, so it has an ecological nature. For example, all racial-ethnical groups and both people above and below the poverty line may be present in a census tract. It is the variability of their various concentrations (i.e., percentages) across geographic areas that lead to disparity. Our approach is to assess whether one group is disproportionally represented in areas of different levels of accessibility.

The analysis begins with comparing the average accessibility values for different groups to gain some preliminary understanding of the issue. Specifically, the weighted average for each group is calibrated by using the number of population in that group in an area as the weight to average across the whole study area. Census tract is a finer geographic unit than county, and is chosen for the analysis of disparities by demographic groups. Both units are used in the previous section on variability by urbanicity since the two definitions of urbanicity use different area units.

As reported in Table 2 (2nd column), Whites have the lowest average score for accessibility to the NCI cancer centers among the ethnic groups, followed by Hispanics, Blacks, and Asians at the census tract level. This may be attributable to that a disproportionally high number of Whites tend to locate in the suburbia or rural areas, and thus suffer from the poorest geographic accessibility [43]. Onega et al. [2] also found that African Americans have a shorter median travel time from their nearest NCI Cancer Centers (69 minutes) than the Whites (86 minutes), and Asians were found to have the shortest median travel time (28 minutes). The finding implies that when it comes to geographic accessibility, minorities such as Hispanics, Blacks and Asians actually come ahead of Whites. We may term this as "*reversed racial disadvantage.*"

Table 2 also shows that population below the poverty line have a lower weighted average accessibility value than the overall population in general (either at the census tract level). Unlike the aforementioned "reversed racial disadvantaged," this suggests that population below poverty remain at a higher risk of low accessibility to NCI cancer centers than their counterpart. The disagreement in accessibility disparities between the racial-ethnic groups and socioeconomic groups is in contrast to the common observation of co-location of concentration of minority groups and people under the poverty line. This disconnection calls for more in-depth examination and validation.

Are the aforementioned differences in average accessibility scores among demographic groups statistically significant? One way to formulate the null hypothesis (H_0) is such as: the ratios of a racial-ethnic (or demographic) group in areas with above-average accessibility values are the same as those areas with below-average accessibility values. If this null hypothesis for the one-tailed test is rejected, the alternative that the ratios of this particular population group are significantly higher (lower) in areas of above-average accessibility is accepted. This may be measured by conducting a pooled t-test to compare the sample mean

of the group's percentages in areas with accessibility higher than average with those below the average. For simplicity and easy interpretation, we follow the weighted regression model proposed by Irkam et al. [44] to implement the test. The weighted ordinary-least-squares (OLS) regression model is formulated as

$$Y = a + b * Flag \tag{3}$$

where the dependent variable Y stands for ratios of ethnic groups in various areas, the independent variable $Flag$ is a binary dummy variable (= 0 or 1, flagging whether a tract has an accessibility value above or below the average), and a and b are parameters to be estimated. By employing a weight term (i.e., population in each tract) in the regression, the error term is weighted heavier in a census tract (or county) with more population than one with less population.

For example, using 72,028 census tracts, the average accessibility value is 7.03. All the tracts are split into two parts: tracts in Part 1 are coded as "Flag=0" in which the accessibility values are larger than or equal to 7.03, the rest in Part 2 with value less than 7.03 are coded as "Flag=1". The regression results are presented in Table 3. For example, the model result for Whites indicates that the sample mean of White ratios in above-average-accessibility tracts is 0.2652 (when Flag=0), whereas the sample mean of White ratios in below-average-accessibility tracts is 0.2652 + 0.1959 = 0.4611 (when Flag = 1). The corresponding t value (= 33.45) indicates that the difference is statistically significant (p<0.001).

Table 3 reported the statistical difference in percentages of demographic groups between the areas with below-average accessibility and those with above-average accessibility. Based on Table 3 (column 4), percentages of White are higher in the census tracts with above-average accessibility values than the census tracts with below-average accessibility scores, and the opposite is observed for Black, Hispanic and Asian. By poverty status, percentages of population below poverty line are higher in tracts with below-average accessibility values than those above-average accessibility tracts. These findings are consistent with the results from Table 2. In other words, the reverse racial disadvantages for Black, Hispanic and Asian are validated by the statistical test, and the higher concentration of population below the poverty line in poorer accessibility areas is also confirmed as statistically significant.

Variation of accessibility across census regions

Similarly, we now turn attention to examining possible variability across the four census regions. First of all, the average accessibility for all population is the highest in the Northeast, and declines to Midwest, South and then West. The advantage in the Northeast is evident while the differences among the other three regions are much smaller.

As reported in Table 2 (columns 3–6), the observations on disparities by demographic groups in each region are largely consistent with those at the national level. There are two discrepancies: one is the higher average accessibility score for below-poverty-line population in the Northeast region (10.10 vs. the region's average 9.72), another is the lower average accessibility score for Hispanic in the South region (6.35 vs. the region's average

6.53). Both differences are small, and we will revisit them after the statistical tests as formulated by the weighted regression model in Equation (3) are conducted.

The results of the statistical tests by regions are outlined in Table 3 (columns 5–8). To save space, only the difference in average percentages of a demographic group in tracts with above-average accessibility vs. average percentages of that group in tracts with below-average accessibility is presented, along with its corresponding t-value. The examination begins with focusing on the sign of "difference in %" for each model, especially in whether a regional model yields a different sign from the national model and whether the difference is statistically significant. Three observations (also highlighted in the table) merit noticing and discussion. Note that both the differences for poverty status in the Northeast region and for Hispanic in the South region are not statistically significant. Recall the two discrepancies (trend in regional averages vs. national averages) raised from our survey of Table 2, as discussed in the previous paragraph: higher average accessibility for below-poverty-line population in the Northeast and lower average accessibility score for Hispanic in the South. Neither is substantiated by the statistical test. The other "outlier" is the negative sign (−1.05) for difference of population % in poverty status in the Midwest, and the difference is statistically significant. In other words, the Midwest model indicates that 1.05% more population in poverty are in above-average-accessibility tracts than those in below-average-accessibility tracts, in contrast to the finding at the national level. The conjunction of concentration of population in poverty status and NCI Cancer Center accessibility may indeed exhibit a unique pattern than the rest of the country. The difference, though statistical significant, is minor (a 1.05% difference) and needs to be interpreted with caution. This highlights the value of rigorous statistical analysis that interpretation from simple comparison of average accessibility scores may be misleading. Overall the regional models confirm that the findings at the national level are largely consistent across the four census regions.

Another value of deriving the regional models is to provide insight on the contribution of individual regions to the national trend. For example, based on the values of "difference in %" reported in Table 3 (also their corresponding t-values), the disparities in relative concentrations of various demographic groups in poorer access areas vs. better access areas vary in degree in each region: the "White disadvantage" and "Hispanic advantage" are most pronounced in the Northeast, the "Black advantage" is more evident in the Midwest and Northeast, the "Asian advantage" are the strongest in the West, and the disadvantage for people under the poverty line is most severe in the South.

Conclusion

In public health studies, geographic accessibility is widely used to assess the convenience of people seeking health care. This research uses the NCI Cancer Centers as a case study to examine both geographic and racial disparities in geographic accessibility. The results show that the accessibility generally declines with increment of travel time away from the centers, but the pattern varies as a result of complex interaction between the centers and population, both of which are distributed far from uniformly. Urban advantage is pronounced at the county level as the average accessibility is highest in large central metro areas, declines to

large fringe metro, medium metro, small metro, micropolitan and noncore rural areas. When a different urban-rural definition based on a finer geographic unit (i.e., census tract) is used, the urban advantage remains largely intact as urbanized areas (of 50,000 or more people) enjoy higher average accessibility than urban clusters (of 2,500–50,000 people) or rural areas. However, accessibility in urban clusters, on average, is slightly lower than that in rural areas. Analysis on the racial disparity reveals that on average White have the lowest geographic accessibility, followed by Hispanic, Black and Asian, and the differences are statistically significant. This "reversed racial disadvantage" seems counterintuitive, but is consistent with the finding reported in an influential prior study [17]. Population under the poverty line are disproportionally concentrated in lower accessibility areas. These findings are largely consistent across the four census regions.

Caution is advised in interpreting the observation of "reversed racial disadvantage" in geographic accessibility to the NCI Cancer Centers from this study. First, the accessibility measure is location-based and ecological in nature, and does not necessarily transfer to advantage in actual utilization of the services provided by these facilities. When such an advantage in accessibility is indeed realized in a higher admission rate, as it is the case for African American reported by Onega et al. [17], it leads to better outcomes for the group and contributes to reduction in health disparity. Secondly, there is geographic variability in this finding. For instance, the "Hispanic advantage" is absent in the South Region in this study. This suggests that the "Hispanic paradox" (i.e., better health status including cancer mortality rate for Hispanics), especially in the southwest [42,43], could not be attributable to the favorable access to quality cancer care such as NCI Cancer Centers in the region. Finally, it is important to uncover the complex interaction between racial-ethnic groups and other socioeconomic variables, which would require data of individuals, and thus not a feasible task for this study.

Additional limitations in this research need to be acknowledged. First, the focus on a single segment of cancer care delivery systems, such as NCI Cancer Centers, is very limited in scope as many other hospitals outside of the system also provide quality cancer care, including satellite facilities of NCI Cancer Centers. Satellite facilities tend to only provide outpatient cancer care, so although important to include in comprehensive studies of cancer accessibility, only including parent facilities in this study isolates inpatient accessibility of NCI Cancer Centers. Secondly, in implementing the geographic accessibility measure, we used an arbitrarily defined distance decay function to characterize the patient behavior in seeking specialized cancer care. Future research needs to use data of actual hospital visits to capture the best fitting distance decay function. Thirdly, the current estimate of cancer patients at the small geographic areas (census tracts) is coarse and needs to be improved as data of finer geographic resolution will increasingly become available. Similarly, the number of physicians/oncologists may be a better measure of a hospital's capacity than its bed size used in this study, and we only consider the parent not satellite sites of NCI Cancer Centers. Lastly, in analysis of accessibility disparity between demographic groups, our finding of the so-called "reversed racial advantage" is preliminary and more importantly, ecological in nature. Future work may use data of individual patients to validate such a preliminary observation.

Acknowledgements

References

1. CDC (2017) Statistics for Different Kinds of Cancer. https://www.cdc.gov/cancer/dcpc/data/types.htm (last accessed on 10–24-2017)

2. Onega T, Duell EJ, Shi Xun, Wang Dongmei, Demidenko E, Goodman D (2008) Geographic access to cancer care in the U.S. Cancer 112 (4):909–918 [PubMed: 18189295]

3. Ananthakrishnan AN, Hoffmann RG, Saeian K (2010) Higher Physician Density is Associated with Lower Incidence of Late-stage Colorectal Cancer. Journal of General Internal Medicine 25 (11):1164–1171 [PubMed: 20658268]

4. Gumpertz ML, Pickle LW, Miller BA, Bell BS (2006) Geographic patterns of advanced breast cancer in Los Angeles: associations with biological and sociodemographic factors (United States). Cancer Causes Control 17 (3):325–339 [PubMed: 16489540]

5. Celaya MO, Rees JR, Gibson JJ, Riddle BL, Greenber ER (2006) Travel distance and season of diagnosis affect treatment choices for women with early-stage breast cancer in a predominantly rural population (United States). Cancer Causes Control 17 (6):851–856 [PubMed: 16783613]

6. Avis NE, Smith KW, Link CL, Hortobagyi GN, Rivera E (2006) Factors Associated With Participation in Breast Cancer Treatment Clinical Trials. American Society of Clinical Oncology 24 (12):1860–1867

7. Agency for Healthcare Research and Quality (2015) 2014 National Healthcare Quality and Disparities Report. U.S. Department of Health and Human Services, Rockville, MD

8. Finlayson SRG, Birkmeyer JD, Tosteson ANA, Nease RF (1999) Patient Preferences for Location of Care: Implications for Regionalization. Medical Care 37 (2):204–209 [PubMed: 10024124]

9. Shi X, Alford-Teaster J, Onega T, Wang D (2012) Spatial access and local demand for major cancer care facilities in the United States. Annals of the Association of American Geographers 102 (5):1125–1134

10. Wang F, Fu C, Shi X (2015) Planning Towards Maximum Equality in Accessibility of NCI Cancer Centers in the U.S. Spatial Analysis in Health Geography. Henry Ling Limited, United Kingdom, pp 261–274

11. Meilleur A, Subramanian S, Plascak JJ, Fisher JL, Paskett ED, Lamont EB (2013) Rural Residence and Cancer Outcomes in the US: Issues and Challenges. Cancer Epidemiol Biomarkers Prev 22 (10). doi:10.1158/1055-9965.EPI-13-0404

12. Charlton M, Schilchiting J, Chioreso C, Ward M, Vikas P (2015) Challenges of Rural Cancer Care in the United States. Oncology

13. UNC (2017) 82 Rural Hospital Closures: January 2010 – Present. Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill, Chapel Hill, NC

14. National Cancer Institute (2008) Cancer Health Disparites.

15. Esnaola NF, Ford ME (2012) Racial Differences and Disparities in Cancer Care and Outcomes: Where's the Rub? Surgical Oncology Clinics of North America 21 (3):417–437 [PubMed: 22583991]

16. Wang F, McLafferty S, Escamilla V, Luo L (2008) Late-stage breast cancer diagnosis and health care access in Illinois. Professional Geographer 60:54–69

17. McLafferty S, Wang F, Luo L, Butler J (2011) Rural-urban inequalities in late-stage breast cancer: spatial and social dimensions of risk and access. Environment and Planning B-Planning & Design 38:726–740

18. National Cancer Institute (2017) http://cancercenters.cancer.gov/Center/CancerCenters (last access on 10–24-2017).

19. In H, Neville BA, Lipsitz SR, Corso KA, Weeks JC, Greenberg CC (2012) The Role of National Cancer Institute–Designated Cancer Center Status. Ann Surg 255 (5):890–895 [PubMed: 22504278]

20. Onega T, Duell EJ, Shi X, Demidenko E, Gottlieb D, Goodman DC (2009) Influence of NCI cancer center attendance on mortality in lung, breast, colorectal, and prostate cancer patients. Medical Care Research and Review 66 (5):542–560 [PubMed: 19454624]

21. National Cancer Institute (2017) NCI Budget and Appropriations. National Cancer Institute at the National Institutes of Health

22. Lewis BA, Gibson D, Cunningham NA (2013) NCI Comprehensive Cancer Centers. Springer Health Advisory Summit 2013. https://library.mskcc.org/sites/library.mskcc.org/media/pictures/about/staff/SpringerHealthSummit_2013.pdf (last accessed on 10–24-2017)

23. Lin CC, Bruinooge SS, Kirkwood MK, Olsen C, Jemal A, Bajorin D, Giordano SH, Goldstein M, Guadagnolo BA, Kosty M, Hopkins S, Yu JB, Arnone A, Hanley A, Stevens S, Hershman DL (2015) Association Between Geographic Access to Cancer Care, Insurance, and Receipt of Chemotherapy: Geographic Distribution of Oncologists and Travel Distance. Journal of Clinical Oncology 33 (28):3177–3185 [PubMed: 26304878]

24. Massarweh NN, Chiang Y-J, Xing Y, Chang GJ, Haynes AB, You YN, Feig BW, Cormier JN (2014) Association Between Travel Distance and Metastatic Disease at Diagnosis Among Patients With Colon Cancer. American Society of Clinical Oncology 32 (942–948)

25. Wang EH, Mougalian SS, Soulos PR, Rutter CE, Evans SB, Haffty BG, Gross CP, Yu JB (2014) Adoption of Hypofractionated Whole-Breast Irradiation for Early-Stage Breast Cancer: A National Cancer Data Base Analysis. International journal of Radiation Oncology 90 (5):993–1000

26. Onega T, Alford-Teaster J, Wang F (2017) Population-based geographic access to National Cancer Institute (NCI) Cancer Center parent and satellite facilities. Cancer 123 (17):3305–3311 [PubMed: 28464212]

27. Guagliardo MF (2004) Spatial accessibility of primary care: concepts, methods and challenges. International Journal of Health Geographics 3 (3)

28. Alford-Teaster J, Lange JM, Hubbard RA, Lee CI, Haas JS, Shi X, Carlos HA, Henderson L, Hill D, Tosteson ANA, Onega T (2016) Is the closest facility the one actually used? An assessment of travel time estimation based on mammography facilities. International Journal of Health Geographics 15 (8)

29. Henry KA, Boscoe FP, Johnson CJ, Goldberg DW, Sherman R, Cockburn M (2011) Breast Cancer Stage at Diagnosis: Is Travel Time Important? Journal of Community Health 36 (6):933–942 [PubMed: 21461957]

30. GAO (1995) Health Care Shortage Areas: Designation Not a Useful Tool for Directing Resources to the Underserved. vol HEHS-95–200. General Accounting Office, Washington, DC

31. Wang F (2015) Quantitative Methods and Socio-Economic Applications in GIS. CRC Press

32. Luo W, Wang F (2003) Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. Environment and Planning B: Planning and Design 30:865–884

33. Wang F (2012) Measurement, Optimization, and Impact of Health Care Accessibility: A Methodological Review. Ann Assoc Am Geogr 102 (5):1104–1112 [PubMed: 23335813]

34. National Cancer Institute (2016) NCI-Designated Cancer Centers

35. Fotheringham AS, Wong DWS (1991) The modifiable areal unit problem in multivariate statistical analysis. Environment & Planning A 23 (7):1025–1044

36. ESRI (2016) ESRI Demographic and Business Data List. Census 2010.

37. National Center for Health Statistics (2013) 2013 Urban-Rural Classification Scheme for Counties. https://www.cdc.gov/nchs/data_access/urban_rural.htm (last accessed on 10–24-2017)

38. United States Census Bureau (2015) 2010 Census Urban and Rural Classification and Urban Area Criteria.

39. Delamater PL, Messina JP, Grady SC, WinklerPrins V, Shortridge AM (2013) Do more hospital beds lead to higher hospitalization rates? a spatial examination of Roemer's Law. PLoS ONE 8 (2).

40. Shen Q (1998) Spatial technologies, accessibility, and the social construction of urban space. Computers, Environment and Urban Systems 22 (5):447–464

41. Larson E, Lin SX, Gomez-Durate C (2003) Antibiotic Use in Hispanic Households, New York City. Emerg Infect Dis 9 (9):1096–1102 [PubMed: 14519246]

42. Xu Y, Wang F (2015) Built environment and obesity by urbanicity in the U.S. Health & Place 34:19–29 [PubMed: 25864019]

43. Fu C (2015) Planning Towards Equal Spatial Accessibility of NCI Cancer Centers Across Geographic Areas and Demographic Groups in the U.S, LSU Doctoral Dissertation

44. Ikram SZ, Hu Y, Wang F (2015) Disparities in Spatial Accessibility of Pharmacies in Baton Rouge, Louisiana. Geographic Review 105 (4):492–510
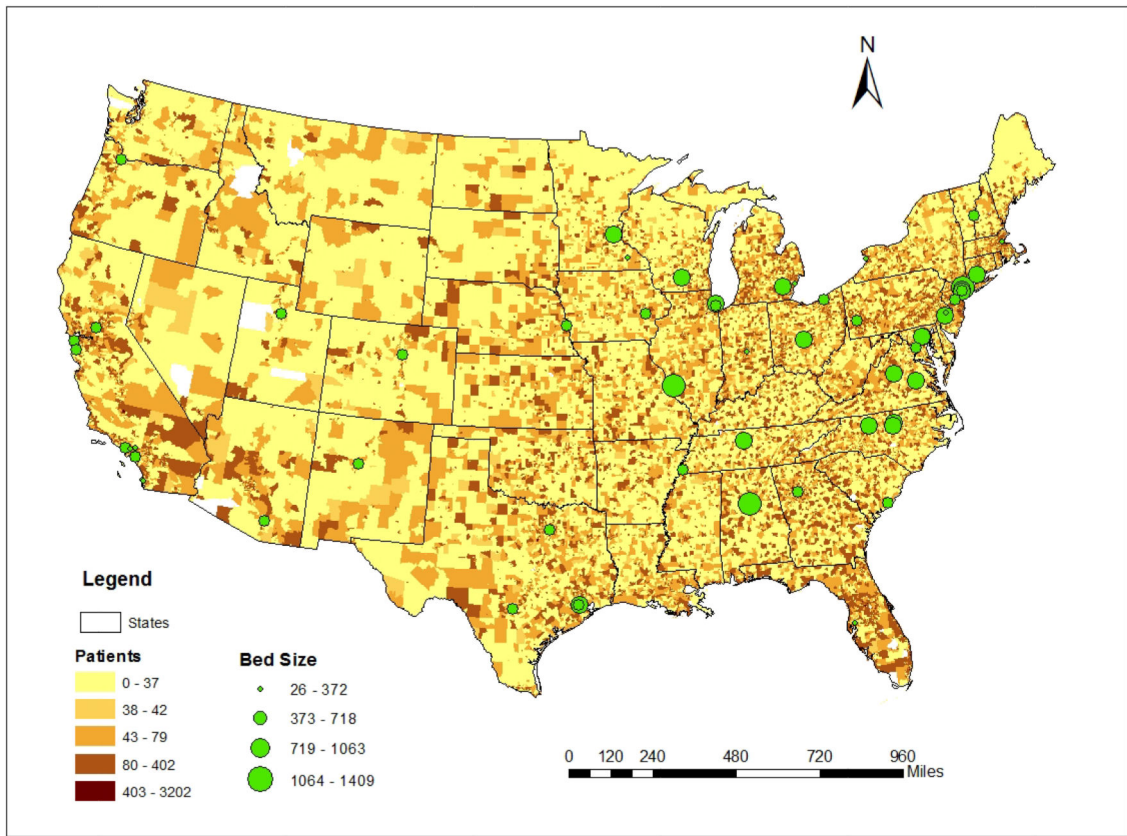
Figure 1.
NCI Cancer Centers and potential cancer patients at the census tract level in the contiguous U.S.
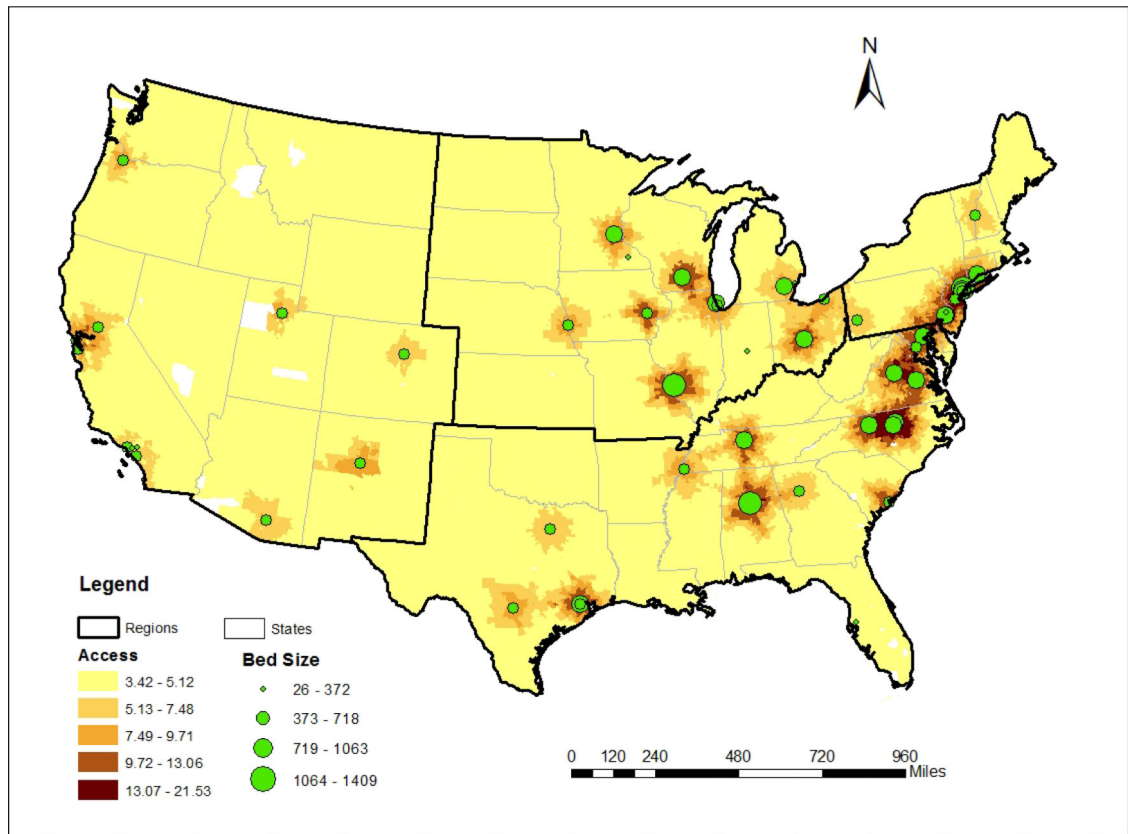
Figure 2.
Geographic accessibility of NCI Center Centers at the census tract level

Table 1.

Variation of accessibility index across urban-rural classifications at county and census tract levels

County Level			Census tract level		
	Coefficient (t-value)	SE		Coefficient (t-value)	SE
Large central metro (reference category)	*8.85*** (32.12)*	0.28	*Urbanized areas (reference category)*	*7.95*** (473.02)*	0.02
Large fringe metro	−1.12*** (−3.74)	0.30	Urban clusters	−2.92*** (−45.35)	0.06
Medium metro	−3.47*** (−11.56)	0.30	Rural	−2.66*** (−89.53)	0.03
Small metro	−4.08*** (−13.56)	0.30			
Micropolitan	−4.41*** (−15.22)	0.29			
Non-core	−4.83*** (−17.06)	0.28			
Adjusted R^2	0.25		0.11		
F-statistic	202.67		4535.69		

Note

***
$p <= 0.001$

Table 2.

Weighted average accessibility for demographic groups (census tract level)

	Contiguous U.S.	Census Region			
		Northeast	Midwest	South	West
All population	*7.03*	*9.72*	*6.73*	*6.53*	*5.99*
By racial-ethnic group:					
White	6.67	8.80	6.50	6.23	5.76
Black	8.17	12.50	8.03	7.18	6.53
Hispanic	7.31	12.82	7.62	6.35	6.20
Asian	8.76	12.70	7.99	8.36	7.22
By poverty status:					
Population below poverty line	6.79	10.10	6.69	6.13	5.84

Table 3.

Test of statistical difference in percentages of demographic groups (census tract level)

	Conterminous U.S.			Census Region			
				Northeast	Midwest	South	West
	% > average accessibility	% < average accessibility	Difference in %	Difference in %	Difference in %	Difference in %	Difference in %
By racial-ethnic group:							
White	26.52	46.11	19.59 *** (33.45)	23.60 *** (54.04)	16.51 *** (43.81)	11.82 *** (35.39)	12.75 *** (39.16)
Black	6.57	6.11	−0.46 *** (−39.38)	−12.25 *** (−33.55)	−12.83 *** (−35.71)	−8.11 *** (−25.22)	−2.15 *** (−17.04)
Hispanic	9.26	7.15	−2.11 *** (−12.07)	−12.30 *** (−43.91)	−4.42 *** (−23.86)	−0.46 (−1.66)	−8.18 *** (−20.75)
Asian	2.82	1.79	−1.03 *** (−55.32)	−5.58 *** (−37.54)	−2.09 *** (−31.56)	−2.13 *** (−36.73)	−7.88 *** (−43.01)
By poverty status:							
Population below poverty line	5.65	9.46	3.81 *** (16.28)	−0.05 (−0.24)	−1.05 *** (−5.12)	2.74 *** (16.05)	1.80 *** (9.61)

Note: t-value in parentheses

***
$p <= 0.001$