


Rural representation of the surveillance, epidemiology, and end results database

Joshua Herb^{1,4}  · Rachael Wolff² · Philip McDaniel² · Mark Holmes^{3,4} · Jennifer Lund³ · Karyn Stitzenberg¹

Received: 31 March 2020 / Accepted: 21 November 2020 / Published online: 3 January 2021

Abstract

Purpose SEER data are widely used to study rural–urban disparities in cancer. However, no studies have directly assessed how well the rural areas covered by SEER represent the broader rural United States.

Methods Public data sources were used to calculate county level measures of sociodemographics, health behaviors, health access and all cause cancer incidence. Driving time from each census tract to nearest Commission on Cancer certified facility was calculated and analyzed in rural SEER and non-SEER areas.

Results Rural SEER and non-SEER counties were similar with respect to the distribution of age, race, sex, poverty, health behaviors, provider density, and cancer screening. Overall cancer incidence was similar in rural SEER vs non-SEER counties. However, incidence for White, Hispanic, and Asian patients was higher in rural SEER vs non-SEER counties. Unadjusted median travel time was 53 min (IQR 34–82) in rural SEER tracts and 54 min (IQR 35–82) in rural non-SEER census tracts. Linear modeling showed shorter travel times across all levels of rurality in SEER vs non-SEER census tracts when controlling for region (Large Rural: 13.4 min shorter in SEER areas 95% CI 9.1;17.6; Small Rural: 16.3 min shorter 95% CI 9.1;23.6; Isolated Rural: 15.7 min shorter 95% CI 9.9;21.6).

Conclusions The rural population covered by SEER data is comparable to the rural population in non-SEER areas. However, patients in rural SEER regions have shorter travel times to care than rural patients in non-SEER regions. This needs to be considered when using SEER-Medicare to study access to cancer care.

Keywords Rural · SEER · Travel time · Cancer

Introduction

The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI) is a population based cancer database that currently covers approximately 34.6% of the United States population [1]. It was started in 1973 with 9 registries but has expanded to include

data from 21 registries as of 2019. Registries are carefully selected to be representative of the country as a whole with oversampling of certain minority populations to allow adequate sample size for analysis [2]. The SEER database is a widely used resource for studies of cancer incidence, mortality, outcomes, treatment patterns, and disparities [3, 4].

Relevant to this study, the SEER database is commonly used to study rural–urban cancer disparities. When linked to Medicare claims data, the SEER-Medicare database is also used to study access to care issues such as travel distance [5, 6]. However, there are limited data on how the rural population covered in the SEER database compares to the broader US rural population, and this is important for understanding generalizability of these studies. We do know that on the whole the SEER population is more urban than the non-SEER population from a recent study by of the SEER-18 database [7]. However, no studies specifically examine how the rural population covered in SEER represents the broader US rural population. SEER registries are

✉ Joshua Herb
Joshua.herb@unchealth.unc.edu

¹ Department of Surgery, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

² University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

³ Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁴ Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

not randomly chosen, so it may not represent the considerable heterogeneity of the US rural population with respect to demographic/socioeconomic measures and health behaviors [8, 9]. The primary objective of this study is to compare rural areas included in the SEER database with rural areas in non-SEER areas with respect to demographics, socioeconomic measures, health behaviors, health access, and cancer incidence.

Methods

Study population

We examined demographic, socioeconomic, health behaviors, and cancer incidence measures from all counties in the United States ($n = 3,142$) and compared travel times from their respective populated census tracts ($n = 72,615$) in the 2010 census.

Definition of rurality, SEER and non-SEER areas

The primary exposure was rural county/census tract SEER participant. The most recent SEER-21 data release contains cancer registries covering the entire state for 13 states (Utah, New York, New Jersey, New Mexico, Massachusetts, Louisiana, California, Kentucky, Iowa, Idaho, Hawaii, Georgia and Connecticut), as well as data from 13 counties in the Seattle-Puget Sound registry in Washington and from three counties in the Detroit metropolitan registry in Michigan [10]. The counties from these registries, as well as their corresponding census tracts, were considered SEER areas. The remainder of the geographic areas contained within the 50 states of the United States were designated non-SEER areas. For the purposes of this study, geographic areas for which SEER data only include data from Native American tumor registries, rather than from the entire population, (Alaska, Arizona and Oklahoma) were also considered non-SEER areas.

Rurality was defined at the county level using the 2017 Office of Management and Budget's (OMB) guidelines [11]. Metropolitan counties were considered urban while micropolitan and non-core counties were considered rural. Metropolitan areas are defined by a core urban county (or counties) of 50,000 or more people, along with the surrounding counties that are highly economically and socially integrated, while micropolitan areas have a similar definition but for a core urban population of 10,000–49,999 people. Those not designated as either metropolitan or micropolitan are non-core [12]. Rurality was additionally defined at the census tract level using secondary rural urban commuting area (RUCA) codes [13]. RUCA codes are based on the OMB guidelines but also incorporate daily commuting flows. They consist of 10 primary codes which refer to the

primary commuting destination but are further subdivided into 21 secondary codes to also indicate the secondary commuting destinations [13]. We used the secondary codes to categorize rurality into four levels (urban, large rural, small rural, and isolated rural) based on previous recommendations [14].

County level demographic, socioeconomic and health access measures

The 2018–2019 county level Area Health Resource File (AHRF) was used to obtain county level demographic, socioeconomic and health access measures. The AHRF compiles these measures from over 50 data sources. The latest available year across measures in the AHRF may differ depending on the latest year available in the source data. For our purpose, the most recent year was used in order to estimate the most updated representation of the populations in each area. Thus, the majority of data comes from 2017, but for some measures the data are older. Factors that were analyzed for each county included age distribution (latest year 2016), education (2013–2017 percent less than high school, at least high school, and at least college educated), unemployment rate, per capita income, percent in poverty, and percent under 65 who are uninsured. The number of primary care physicians, surgeons, internal medicine subspecialists, and radiation oncologists in each county were used to calculate physician density.

Health behaviors

The 2019 County Health Rankings (CHR) data source, published by the Robert Wood Johnson foundation and the University of Wisconsin, was utilized to obtain county level measures health behaviors (percent tobacco users, percent obese, percent excessive drinking, and percent physically inactive). The CHR compiles county level data from a variety of national data sources. Tobacco and alcohol use rates in 2019 CHR are compiled from the 2016 Behavioral Risk Factor Surveillance System (BRFSS), while obesity and physical inactivity rates are compiled from the 2015 Centers for Disease Control (CDC) Diabetes Interactive Atlas [15].

Cancer screening, incidence and mortality

County level cancer screening and incidence rates were obtained from the National Cancer Institute (NCI) and CDC's State Cancer Profiles tool. Data from the BRFSS and National Health Interview Survey (NHIS) are combined to provide a county level modeled estimate of screening rates from 2008 to 2010 (latest time period available) [16]. The rate of breast cancer screening (defined as mammography within the past 2 years for women over 40 years old), and the

rate of colorectal cancer screening (defined as ever having colorectal endoscopy or having fecal occult blood testing within the past 2 years) were analyzed.

County level age-standardized cancer incidence rates were also captured using the same tool, which uses the cancer incidence rates from the National Program of Cancer Registries Cancer Surveillance System and the SEER program [17]. All cause average cancer incidence for the latest 5 years (2012–2016) for both sexes and all ages were evaluated. Incidences were analyzed overall by SEER and non-SEER rural areas and then further stratified by race/ethnicity (White, Black, Hispanic, Asian). The states of Kansas and Minnesota were excluded from this analysis as they do not allow release of county level cancer incidence data. Additionally, if a county had fewer than 16 cases during the time period or the number of persons in a county of a particular race were low, then the county level estimates were suppressed to protect confidentiality [18].

Travel distance

To measure travel time to care, Commission on Cancer-certified (CoC) facilities (National Cancer Institute Designated, Academic, Integrated Network, Comprehensive Community, and Community Cancer Program) ($n = 1,185$) were geocoded using Google maps. The geographic centroid of each census tract was created using ArcGIS (Environmental Systems Research Institute, Redlands, California). Network travel time was calculated by determining the fastest route (by time) between each census tract centroid and the nearest CoC facility using Network Analyst (Environmental Systems Research Institute, Redlands, California) and US road network data (2012). Some pairs had the same origin and destination; therefore, travel time was recorded as 0 and is underestimated in those tracts.

Statistical analysis

The population adjusted county level means for each measure above were calculated and compared across SEER and non-SEER rural areas. A sub-analysis was performed by stratifying the rural group into its component micropolitan and non-core categories to compare measures across SEER and non-SEER larger rural areas with more economic integration (micropolitan) and those that are isolated (non-core). Given the large numbers of counties, statistical significance would be reached with even small differences. Therefore, we chose to calculate standardized differences between the groups to evaluate their differences. This method uses the difference in means between two groups divided by the pooled standard deviations [19]. The result is the difference in terms of standard deviation units and is often called the “effect size.” This measure is

less sensitive to sample size and has been used in similar studies in the past [7]. There is no agreed upon definition of a meaningful effect size, but standard differences < 0.1 are generally considered negligible [20]. Therefore, values of 0.1 or greater were considered meaningful differences between groups.

Median travel times by SEER and non-SEER tracts are reported overall, by RUCA rural categories, and by census region. Linear regression was used to measure the overall effect of SEER status on travel time while controlling for census region (Northeast, South, Midwest, West) given that CoC facilities are not evenly distributed in the United States, and travel time can vary significantly by region [21]. Census division was not used due to small or empty cell sizes in some divisions. The regression was also weighted by 2010 census tract population. This was performed first by using census tracts from all RUCA rural categories (large rural, small rural, and isolated rural) combined and then by stratifying each rural RUCA category. A sensitivity analysis was performed in which travel time was log-transformed and coefficients were exponentiated to give a meaningful interpretation (reported as percent change in travel time). An additional sensitivity analysis was performed by omitting the state of Alaska and repeating the regressions. All statistical analyses were performed in STATA (v16.0). This study was deemed exempt by the University of North Carolina Institutional Review Board as only deidentified secondary data sources were used.

Results

Rural designations

Overall, 80% of counties in SEER areas (584 of 732), and 79% of counties (1898 of 2410) in non-SEER areas are rural. Yet, the population of these rural counties represent only 8.5% of the SEER population and 17.1% of the non-SEER population.

Demographic and socioeconomic measures

Table 1 displays results of the county-level rural SEER and rural non-SEER comparison. Overall, there were no meaningful differences in the population adjusted age, race, and sex distributions between rural SEER and rural non-SEER counties. The only difference on subgroup analysis was that SEER micropolitan areas had a lower proportion of white persons compared to non-SEER micropolitan areas (72.9% vs 78.3%, St. Diff = $-.126$). There were no meaningful

Table 1 Comparison of demographic, socioeconomic, health behavior and health access measures among rural SEER and rural non-SEER counties

Measures ^a	SEER rural (n=584)	Non-SEER rural (n=1898)	SD	SEER micro ^b (n=258)	Non-SEER micro (n=1043)	SD	SEER non- core ^b (n=326)	Non-SEER non- core (n=855)	SD
<i>Demographics</i>									
Age (2016)									
% < 19	23.0	22.7	0.01	23.3	22.8	0.01	22.5	22.4	0.00
% 18–65	56.8	56.6	0.01	57.5	57.1	0.01	55.7	55.7	0.00
% > 65	18.1	18.8	– 0.02	17.3	17.8	– 0.01	19.4	20.1	– 0.02
<i>Race/Ethnicity mean %</i>									
% White	76.2	78.8	– 0.06	72.9	78.3	– 0.13*	81.6	79.6	0.05
% Black	8.4	7.8	0.02	8.1	7.7	0.01	8.9	7.9	0.03
% Asian	1.7	0.9	0.07	2.3	1.1	0.10*	0.8	0.7	0.01
% Hispanic	9.5	8.4	0.04	11.4	9.4	0.06	6.4	6.9	– 0.02
Sex, mean % female	49.9	49.9	0.00	50.1	50.1	0.00	49.6	49.7	0.00
<i>Socioeconomic</i>									
Education									
% < HS (2013–2017)	15.0	14.4	0.02	14.2	13.8	0.01	16.3	15.3	0.03
% at least HS (2013– 2017)	85.0	85.6	– 0.02	85.8	86.2	– 0.01	83.7	84.8	– 0.03
% College (2013–2017)	19.8	19.2	0.02	21.5	20.6	0.02	17.0	17.2	0.00
Unemployment rate, mean %	5.3	4.7	0.03	5.2	4.6	0.03	5.4	4.9	0.02
Per capita income, \$, mean (SD)	39,688 (10,328)	39,614 (9315)	0.01	40,639 (11,028)	40,515 (10,452)	0.01	38,127 (8896)	38,302 (7163)	– 0.02
% Persons in poverty; mean	17.7	16.1	0.04	17.4	15.8	0.04	18.3	16.6	0.04
<i>Health behaviors</i>									
% Tobacco use	18.1	18.7	– 0.02	17.8	18.6	– 0.02	18.6	18.8	– 0.01
% Obesity	30.6	31.9	– 0.03	29.9	31.8	– 0.04	31.9	32.0	0.00
% Excessive drinking	16.3	16.6	– 0.01	16.8	16.8	0.00	15.6	16.2	– 0.02
% Phys inactivity	25.5	26.8	– 0.03	24.4	26.4	– 0.05	27.3	27.5	0.00
<i>Health access</i>									
% < 65 Uninsured	9.6	11.9	– 0.07	9.4	11.4	– 0.06	9.9	12.7	– 0.09
<i>Physician density^c</i>									
Primary care, mean	0.53	0.53	0.01	0.58	0.58	0.02	0.45	0.46	– 0.02
Surgeon, mean	0.23	0.22	0.02	0.30	0.29	0.01	0.12	0.11	0.01
IM subspecialist, mean	0.04	0.04	0.01	0.06	0.06	0.00	0.02	0.01	0.01
Rad/Onc, mean	0.01	0.01	– 0.01	0.01	0.01	0.00	0.00	0.00	0.01
<i>Cancer screening</i>									
Mammography screen- ing (2008–2010), % up to date	67	67	0.00	68	68	0.01	67	67	– 0.01
Colorectal screening (2008– 2010), % up to date	61	61	0.01	61	62	0.02	60	60	0.00

All estimates from 2017 unless otherwise noted in the table

SD standard difference, HS high school, IM internal medicine

*Standard difference 0.1 or greater

^aArea Health Resource File used for all measures except tobacco, obesity, excessive drinking, physical inactivity rates (County Health Rankings). National Institutes of Health State Cancer Profiles used for breast and colorectal cancer screening rates

^bRural areas stratified further by micropolitan and non-core areas

^cPhysician density per 1,000 people

differences in educational attainment, unemployment, per capita income, or percent of persons in poverty between rural SEER and rural non-SEER counties.

Health access and health behaviors

Table 1 shows additional measures of health access and health behaviors. The proportion of uninsured patients under 65 was slightly higher in non-SEER counties (9.6% vs 11.9% in rural SEER counties). There were no differences in any provider density measures, rates of mammography screening (67% for both groups), or rates of colorectal cancer screening (61% for both groups). Rates of tobacco smoking (18.1% vs. 18.7%), obesity (30.6% and 32%), physical inactivity (25.5% vs 26.8%), and excessive drinking (16.3% vs 16.5%) were all also similar between rural SEER and non-SEER areas, respectively (Table 1).

Cancer incidence and mortality

Table 2 shows the all cause cancer incidence in rural SEER and non-SEER counties. The number of included counties for each group are listed as some counties were

suppressed due to low number of cases. Overall incidence was similar between groups (447.1 and 444.4 per 100,000; St. Diff = 0.062 in rural SEER and rural non-SEER, respectively). However, when stratified by race, minority groups (Hispanic ethnicity 376.9 vs 321.8, SD: 0.75; Asian 291.2 vs 255.7, SD 0.78) had higher incidences in rural SEER areas in general (Table 2). Three-year average percent deaths due to cancer were 21% and did not differ between rural SEER and rural non-SEER counties (Table 2).

Travel time analysis

Figure 1 shows a broad overview of travel time from each census tract to nearest CoC facility. Of 72,615 census tracts, all but 125 tracts mapped to the nearest facility. Most of these were remote census tracts in Alaska and Montana. Other tracts in Hawaii did not match as there were no complete road networks to the nearest CoC facility. Overall, 91.4% of the 2010 population lived within 60 min of a CoC facility. In all SEER areas, 93.8% live within 60 min of a CoC facility and 90.0% in non-SEER areas. A prolonged travel time of > 2 h was seen for 2.0% of the SEER population and 2.5%

Table 2 County level cancer incidence and mortality rates in SEER and non-SEER counties

	SEER rural (n=584)	Non-SEER rural (n=1898)	SD	SEER micro (n=258)	Non-SEER micro (n=1043)	SD	SEER non- core (n=326)	Non-SEER non- core (n=855)	SD
<i>All cause cancer incidence rate^a, 2012–2016</i>									
All	447.1	444.4	0.06	475.4	440.8	0.58*	446.1	444.7	0.03
St. Dev	46.7	41.6		60.2	59.2		45.9	39.8	
n	582	1731		256	922		326	809	
White	456.8	443.1	0.32*	476.0	437.7	0.64*	456.2	443.5	0.31*
St. Dev.	44.9	42.0		60.2	59.8		44.1	40.2	
n	582	1720		256	911		326	809	
Black	434.8	438.4	− 0.05	488.2	469.2	0.17*	434.1	437.5	− 0.05
St. Dev.	61.0	71.5		129.0	107.1		59.4	70.0	
n	343	833		80	217		263	616	
Hispanic	376.9	321.8	0.72*	396.8	355.8	0.32*	376.7	321.1	0.74*
St. Dev.	68.5	78.9		123.1	129.1		67.9	77.5	
n	228	615		32	140		196	475	
Asian	291.2	255.7	0.75*	380.8	452.9	NA	291.1	255.6	0.75*
St. Dev.	38.6	50.9			64.5		38.5	50.7	
n	173	362		1	2		172	360	
<i>Cancer mortality^b, 2015–2017</i>									
Mean % deaths due to cancer	21.4%	21.22%	0.00	21.3%	21.1%	0.01	21.4%	21.4%	0.00

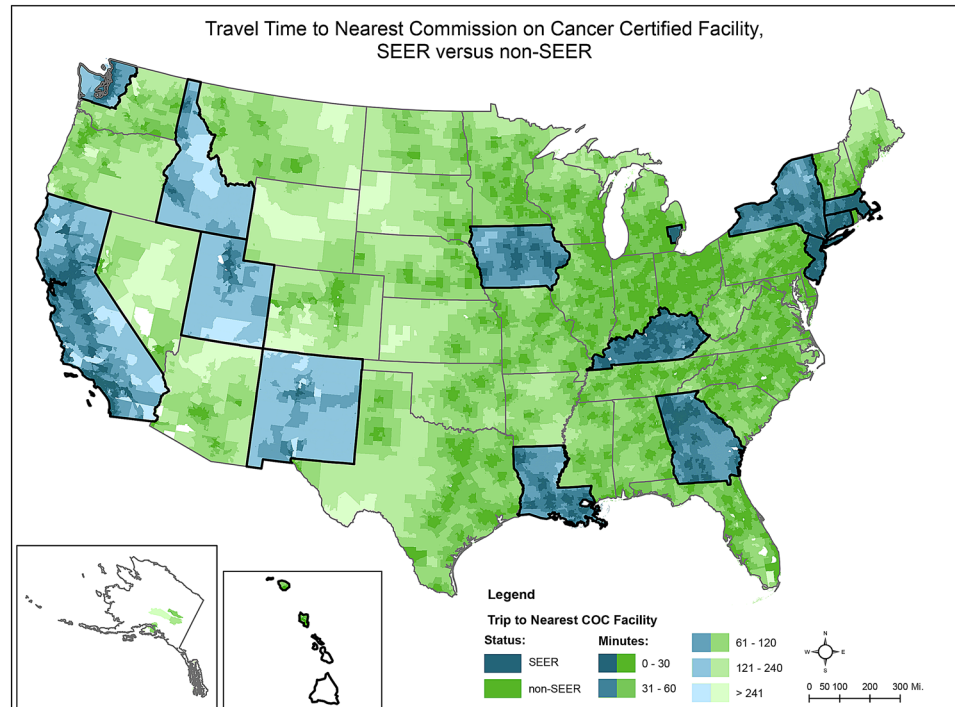
SD standard difference

sStandard difference 0.1 or higher; NA not applicable, unable to calculate as only 1 county was not suppressed for SEER Micro counties for this racial sub-group. n: number of counties

^aCancer incidence rates from National Center for Health Statistics Data

^bCancer mortality data from Area Health Resource File

Fig. 1 Travel time to nearest Commission on Cancer Certified facility by census tract in SEER and non-SEER areas. Blue areas are included in the SEER registry while green areas are not. Darker colors represent shorter travel times. Tracts that are white did not match in the network analysis or were unpopulated



of the non-SEER population. The median number of CoC facilities was similar between SEER and non-SEER areas (15.5 vs 16 in SEER and non-SEER areas, respectively).

Table 3 shows median travel times in SEER and non-SEER areas stratified by census region and level of rurality. When examined generally, median travel time was similar in rural SEER and rural non-SEER counties (52.6 vs 54.3 min, respectively). When stratified by region, median

travel times were shorter in rural SEER regions except for the Midwest, where travel time was longer for rural SEER population compared to rural non-SEER (65.7 vs 47.1 min, respectively). As expected, median travel times increased as tracts increased in rurality.

Table 4 shows the results of the linear regression analysis. When controlling for regional variation, SEER areas had a lower travel time overall and by all levels of rurality

Table 3 Median travel times by region and rurality in SEER vs non-SEER census tracts to nearest commission on cancer certified hospital

	Overall	Northeast	Midwest	South	West
Travel time in minutes, median (IQR)					
All tracts					
Non-SEER	16.2 (9.2–34.2)	15.4 (7.9–30.9)	14.0 (8.1–30.6)	17.5 (10.0–35.8)	17.9 (10.8–48.5)
SEER	11.7 (7.5–21.0)	10.5(6.5–17.2)	13.4 (8.7–29.9)	18.6 (9.4–35.9)	11.3 (7.6–18.5)
All rural					
Non-SEER	54.3 (34.7–82.1)	48.0 (30.7–68.3)	47.1 (29.9–72.5)	55.7 (37.8–77.7)	101.1 (65.8–154.9)
SEER	52.6 (34.1–81.8)	41.9 (29.2–52.9)	65.7 (48.9–89.9)	44.5 (27.8–63.0)	83.5 (46.9–139.4)
Large rural					
Non-SEER	43.0 (23.3–66.6)	33.7 (22.6–49.7)	34.4 (17.2–52.2)	48.2 (28.4–70.6)	85.5 (51.4–121.3)
SEER	45.3 (27.8–72.4)	36.8 (25.1–46.4)	57.3 (36.7–85.2)	38.3 (21.0–57.5)	72.0 (40.0–121.0)
Small rural					
Non-SEER	58.9 (42.4–85.4)	59.6 (42.3–85.3)	50.3 (37.5–74.1)	59.8 (46.1–79.2)	119.0 (77.8–175.6)
SEER	57.3 (38.4–84.0)	50.1 (35.6–66.8)	61.7 (47.6–89.6)	47.9 (33.3–70.2)	87.8 (56.3–148.5)
Isolated rural					
Non-SEER	72.4 (51.1–107.0)	63.4 (49.9–90.3)	68.7 (47.9–100.0)	69.7 (50.0–93.3)	119.2 (81.2–185.2)
SEER	62.9 (44.1–98.5)	45.7 (32.4–59.2)	72.6 (57.3–95.9)	51.8 (40.0–70.2)	116.1 (60.8–166.9)

IQR interquartile range, SEER surveillance, epidemiology and end results

Table 4 Regression model showing effect of SEER region on mean travel time

	Mean difference SEER vs non-SEER tracts (minutes)	95% CI	Mean % difference SEER vs non-SEER tracts	95% CI
	Linear model*		Log-linear model*	
Overall	- 8.8	- 9.5,- 8.1	- 21.1%	- 22.4,- 19.8
All rural	- 14.8	- 18.0,- 11.6	- 11.2%	- 14.3,- 8.1
Large rural	- 13.4	- 17.6,- 9.1	- 10.0%	- 14.8,- 4.9
Small rural	- 16.3	- 23.6,- 9.1	- 14.6%	- 18.7,- 10.3
Isolated rural	- 15.7	- 21.6,- 9.9	- 14.1%	- 18.0,- 10.1
	Linear model without Alaska*		Log-linear model without Alaska*	
Overall	- 7.4	- 7.9,- 6.8	- 20.9%	- 22.1,- 19.5
All rural	- 8.4	- 10.4,- 6.5	- 10.0%	- 13.1,- 6.8
Large rural	- 7.0	- 9.6,- 4.4	- 8.8%	- 13.7,- 3.6
Small rural	- 9.1	- 12.9,- 5.4	- 13.3%	- 17.4,- 9.0
Isolated rural	- 11.3	- 15.2,- 7.3	- 13.2%	- 17.1,- 9.1

Models control for Census Region (Northeast, Midwest, South, West)

assessed. Rural SEER tracts on average had a 14.8 min (95% CI - 11.6,- 18.0) shorter travel time than rural non-SEER tracts. The largest reduction in travel time was in small rural SEER towns compared to the small rural non-SEER towns (16.3 min shorter; 95% CI - 9.1,- 23.6). In the sensitivity analysis using a log-linear model, similar trends were observed. Rural SEER tracts had an 11.2% reduction in travel time compared to rural non-SEER tracts (95% CI - 8.1%,- 14.3%). In our additional sensitivity models without Alaska, effect sizes were smaller, but still significantly lower, in both our linear and log-linear models (Table 4). Rural SEER tracts had an average of 8.4 min shorter travel time compared rural non-SEER tracts (95% CI - 6.5,- 10.4) in our linear model and an estimated 10% shorter travel time in rural SEER compared to rural non-SEER tracts (95% CI - 6.8%,- 13.1%) in the log-linear model.

Discussion

Travel time to cancer care is increasing over time: a recent analysis showed a doubling in the number of patients who travel more than an hour to the nearest cancer hospital from 2005 to 2015 [22]. Regionalization to high volume centers, particularly for surgical care, is likely a major factor underlying this trend [22, 23]. Given the well-documented relationship between higher volume and improved complex cancer surgery outcomes, shifting to care to high volume facilities will improve outcomes on a population level [24-26]. But this comes with a cost of increasing travel for some patients, particularly rural patients, and prolonged travel may be an insurmountable barrier for vulnerable populations at many points in the cancer care continuum [27-29].

To study issues of travel distance or travel time in cancer care, the SEER-Medicare database provides a large cohort where travel distance/time can be calculated. However, no study has examined whether travel times are comparable for SEER and non-SEER areas. In general, our results are in line with other studies showing increasing travel with increasing rurality, with significant variation by region of the country [21, 30]. While our estimated 15-min reduction in average travel time for rural SEER and non-SEER areas is modest (and even lower when Alaska is excluded), the implications of these findings are that to the extent that travel time is a barrier to care for rural patients, the rural areas covered by the SEER-Medicare database may underrepresent that burden in comparison to the non-SEER rural areas. This is important because increased travel burden may be an important mediator for rural-urban cancer disparities [31].

There are only a few previously published analyses comparing sociodemographic factors between the entire (rural and urban) SEER and non-SEER populations. Two of these differ significantly from our study due to the use of older population data and the inclusion of fewer registries in the SEER database at the time of the study [2, 32]. A more recent 2016 study by Kuo et al. [7] used the SEER-18 database in conjunction with the AHRF and 2010 census data and found the SEER population was younger, more educated, and more urban than non-SEER population. SEER areas had a higher representation of racial minorities, which is expected as registries are included in SEER based on their ability to represent these populations, among other factors. The study also found that non-SEER counties were more affluent as defined by lower levels of unemployment and fewer persons living in poverty than SEER counties. In our study, we specifically

focused on the rural population. Having data that compare rural population measures is important because the social determinants of health may partly explain worse cancer outcomes in the rural population [33]. We found the rural SEER areas to be largely representative of the rural non-SEER population. We found similar age and racial distributions, as well as similar socioeconomic and health access measures, even with stratification of rural areas by micropolitan and non-core areas.

The county level incidence rates in our study come from the US Center for Cancer Statistics (USCS), which draws data from both the SEER database and the National Program of Cancer Registries (NPCR)—a population-based cancer surveillance program funded by the Centers for Disease Control and Prevention consisting of state cancer registries included and not included in the SEER database [34]. The USCS data covers the entire US population [35]. The previous analysis by Kuo et al. used 2004–2009 USCS incidence data to compare breast and colorectal cancer incidence in SEER and non-SEER counties. They found similar cancer incidence rates in SEER and non-SEER counties across all racial/ethnic sub-groups [7]. In our analysis restricted to rural cancer incidence rates for rural counties from 2012 to 2016, we found overall similarities in cancer incidence between rural SEER and non-SEER counties. When stratified by race, we found *lower* incidences for White, Hispanic, and Asian patients in rural non-SEER areas, in contrast to what was found in the study including both rural and urban counties. Our findings may differ for several reasons. These studies used overall population incidence where we analyzed publicly available county level incidence data. Therefore, in years with low numbers of cases for a particular racial group, certain counties were suppressed and not included in our analysis. Additionally, since SEER registries are selected based on quality and the ability to represent minority populations, we may expect a lower number of cases captured in the non-SEER registries. Finally, our analysis uses more recent data (2012–2016 as compared to 1999).

Our analyses have several limitations. The cancer incidence comparisons may be limited due to suppressed counties, particularly for Asian patients in rural counties (78% of rural counties suppressed for this racial group). Other racial minorities had better representation (53% and 66% of rural counties suppressed for Black and Hispanic patients, respectively) and the number of suppressed counties was similar for both SEER and non-SEER groups. Travel time was measured to nearest CoC hospital. CoC hospitals report cancer cases to the National Cancer Database (NCDB) and the NCDB has been previously shown to capture approximately 70% of all cancer cases in the US [36]. However, there may be closer hospitals offering cancer services for some census tracts, particularly rural census tracts, that are not CoC certified and would not be

accounted for in our analysis. We also could not account for patient insurance or preferences for certain providers/hospitals. Therefore, the hospitals that each tract mapped to may not be where patients actually receive their care. However, our objective was to measure geographic availability of services.

The SEER database is a vital and high-quality resource for the study of cancer in the United States. The rural SEER areas are representative of the broader rural population in the US across many demographic, socioeconomic, health behavior, and health access measures. Still, there are some differences that remain, particularly with respect to travel time. The SEER-Medicare database may underestimate rural travel time to the nearest specialized cancer center compared to rural non-SEER areas. Researchers should be aware of this potential bias when generalizing studies of travel distance using the SEER-Medicare database.

Availability of data and material

All data sources (U.S. Census data, County Health Rankings, Area Health Resource File, Centers for Disease Control State Cancer Profiles Tool, Commission on Cancer certified facilities, and National Program of Cancer Registries Cancer Surveillance System) are publicly available.

Acknowledgments The authors would like to acknowledge May Kuo, PhD for her assistance with the study design. Dr. Herb is partially supported by a National Service Research Award Pre-Doctoral/Post-Doctoral Traineeship from the Agency for Healthcare Research and Quality sponsored by the Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Grant No. 5T32 HS000032

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by JH, RW and PMcD. The first draft of the manuscript was written by JH and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Dr. Herb is partially supported by a National Service Research Award Pre-Doctoral/Post-Doctoral Traineeship from the Agency for Healthcare Research and Quality sponsored by the Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Grant No. 5T32 HS000032.

Code availability STATA v16.0 was used for statistical analyses. ArcGIS was used for travel time calculations. Analysis code is available on reasonable request.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to disclose. Dr. Herb has disclosed his funding source above.

Ethical approval This is an observational study and the data used in this study are deidentified and publicly available. The University of North Carolina Institutional Review Board deemed the study exempt as this is a secondary data analysis of deidentified data.

References

- (2019) Overview of the SEER Program. In: National Cancer Institute Surveillance, Epidemiology and End Results Program. <https://seer.cancer.gov/about/overview.html>. Accessed 2 July 2020
- Nattinger AB, McAuliffe TL, Schapira MM (1997) Generalizability of the surveillance, epidemiology, and end results registry population: factors relevant to epidemiologic and health care research. *J Clin Epidemiol* 50:939–945
- Harper S, Lynch J (2010) Selected comparisons of measures of health disparities health disparities. A review using databases relevant to Healthy People 2010 Cancer-related objectives. *NCI Cancer Surveill Monogr Ser* 92:1–80
- Zeng C, Wen W, Morgans AK et al (2015) Disparities by race, age, and sex in the improvement of survival for major cancers: results from the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) program in the United States, 1990 to 2010. *JAMA Oncol* 1:88–96. <https://doi.org/10.1001/jamaoncol.2014.161>
- Baldwin L-M, Cai Y, Larson E et al (2008) Access to cancer services for rural colorectal cancer patients. *J Rural Health* 24:390–399. <https://doi.org/10.1111/j.1748-0361.2008.00186.x>
- Stitzenberg KB, Chang YK, Smith AB, Nielsen ME (2015) Exploring the burden of inpatient readmissions after major cancer surgery. *J Clin Oncol* 33:455–464. <https://doi.org/10.1200/JCO.2014.55.5938>
- Kuo T, Mobley LR (2016) How generalizable are the SEER registries to the cancer populations of the USA? *Cancer Causes Control* 27:1117–1126. <https://doi.org/10.1007/s10552-016-0790-x>
- Probst JC, Moore CG, Glover SH, Samuels ME (2004) Person and place: the compounding effects of race/ethnicity and rurality on health. *Am J Public Health* 94:1695–1703. <https://doi.org/10.2105/AJPH.94.10.1695>
- Matthews K, Croft J, Liu Y et al (2017) Health-related behaviors by urban-rural county classification—United States, 2013. *Morb Mortal Wkly Rep Surveill Summ* 66:1–12
- (2019) Registry Groupings in SEER Data and Statistics. In: National Cancer Institute Surveillance, Epidemiology and End Results Program. <https://seer.cancer.gov/registries/terms.html>. Accessed 2 Jan 2020
- (2017) OMB Bulletin 17-01: revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas. Washington DC
- Hall SA, Kaufman JS, Ricketts TC (2006) Defining urban and rural areas in U.S. epidemiologic studies. *J Urban Health* 83:162–175. <https://doi.org/10.1007/s11524-005-9016-3>
- (2019) Rural-urban commuting area codes. In: United States Department of Agriculture, Economic Research Service. <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>. Accessed 12 Apr 2019
- (2019) RUCA data: using RUCA data. In: Rural Health Research Center. <https://depts.washington.edu/uwrucaruca-uses.php>. Accessed 1 Sep 2019
- (2019) 2019 Measures. In: County Health Rankings. <https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/2019-measures>. Accessed 1 Aug 2019
- Liu B, Parsons V, Feuer EJ et al (2019) Small area estimation of cancer risk factors and screening behaviors in US counties by combining two large national health surveys. *Prev Chronic Dis* 16:1–10. <https://doi.org/10.5888/pcd16.190013>
- (2019) Incidence rates table. In: State Cancer Profiles. www.statecancerprofiles.cancer.gov/incidencerates/. Accessed 19 Jan 2020
- (2019) Suppression of rates and counts. In: United States Cancer Stat. https://www.cdc.gov/cancer/uscs/technical_notes/stat_methods/suppression.htm. Accessed 1 Sep 2019
- Flury BK, Riedwyl H, Flury BK, Riedwyl H (2016) Standard distance in univariate and multivariate analysis. *Am Stat* 40:249–251
- Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 46:399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Onega T, Duell EJ, Shi X et al (2008) Geographic access to cancer care in the U.S. *Cancer* 112:909–918. <https://doi.org/10.1002/cncr.23229>
- Diaz A, Schoenbrunner A, Pawlik TM (2019) Trends in the geospatial distribution of adult inpatient surgical cancer care across the United States. *J Gastrointest Surg*. <https://doi.org/10.1097/sla.0000000000003366>
- Stitzenberg KB, Sigurdson ER, Egleston BL et al (2009) Centralization of cancer surgery: implications for patient access to optimal care. *J Clin Oncol* 27:4671–4678. <https://doi.org/10.1200/JCO.2008.20.1715>
- Al-Sahaf M, Lim E (2015) The association between surgical volume, survival and quality of care. *J Thorac Dis* 7:S152–S155. <https://doi.org/10.3978/j.issn.2072-1439.2015.04.08>
- Birkmeyer JD, Sun Y, Wong SL, Stukel TA (2007) Hospital volume and late survival after cancer surgery. *Ann Surg* 245:777–783. <https://doi.org/10.1097/01.sla.0000252402.33814.dd>
- Learn PA, Bach PB (2010) A decade of mortality reductions in major oncologic surgery: the impact of centralization and quality improvement. *Med Care* 48:1041–1049. <https://doi.org/10.1097/MLR.0b013e3181f37d5f>
- Ambroggi M, Biasini C, Del Giovane C et al (2015) Distance as a barrier to cancer diagnosis and treatment: review of the literature. *Oncologist* 20:1378–1385. <https://doi.org/10.1634/theoncologist.2015-0110>
- Herb JN, Dunham LN, Mody G et al (2020) Lung cancer surgical regionalization disproportionately worsens travel distance for rural patients. *J Rural Health*. <https://doi.org/10.1111/jrh.12440>
- Wolfe MK, McDonald NC, Holmes GM (2020) Transportation barriers to health care in the United States: findings from the national health interview survey, 1997–2017. *Am J Public Health* 110:e1–e8. <https://doi.org/10.2105/ajph.2020.305579>
- Chan L, Hart LG, Goodman DC (2006) Geographic access to health care for rural Medicare beneficiaries. *J Rural Health* 22:140–146. <https://doi.org/10.1111/j.1748-0361.2006.00022.x>
- Meilleur A, Subramanian S, Plascak JJ et al (2013) Rural residence and cancer outcomes in the US: issues and challenges. *Cancer Epidemiol Biomark Prev* 22:1657–1667. <https://doi.org/10.1158/1055-9965.EPI-13-0404.Rural>
- Warren JL, Klabunde CN, Schrag D et al (2002) Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 40:IV3–IV18
- Blake KD, Moss JL, Gaysynsky A et al (2017) Making the case for investment in rural cancer control: an analysis of rural

- cancer incidence, mortality, and funding trends. *Cancer Epidemiol Biomark Prev* 26:992–997
34. Wingo PA, Jamison PM, Hiatt RA et al (2003) Building the infrastructure for nationwide cancer surveillance and control—A comparison between The National Program of Cancer Registries (NPCR) and The Surveillance, Epidemiology, and End Results (SEER) Program (United States). *Cancer Causes Control* 14:175–193. <https://doi.org/10.1023/A:1023002322935>
 35. (2020) About US Cancer Statistics. In: Centers for Disease Control and Prevention. <https://www.cdc.gov/cancer/uscs/about/index.htm>. Accessed 21 July 2020
 36. (2019) National Cancer Database. In: American College of Surgeon Cancer Programs. <https://www.facs.org/quality-programs/cancer/ncdb>. Accessed 1 Mar 2020

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