Cancer incidence trends using American Community Survey estimates are not consistent with SEER for small populations

Julia Mantey[a,*], Julie Ruterbusch[b], Rafael Meza[c], Kendra Schwartz[d]

[a] Deparment of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, United States
[b] Department of Oncology, Wayne State University School of Medicine and Karmanos Cancer Institute, Detroit, MI 48201, United States
[c] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, United States
[d] Department of Family Medicine and Public Health Sciences, Wayne State University School of Medicine, Detroit, MI 48201, United States

ARTICLE INFO

ABSTRACT

Background: American Community Survey (ACS) estimates are said to be uncertain for small areas and small population groups. The Surveillance, Epidemiology and End Results (SEER) database uses a decennial census extrapolation methodology to yield population estimates used by cancer researchers across the country. We compared metropolitan Detroit cancer incidence estimates calculated using ACS data to those using SEER population estimates, which we considered to be the gold standard.
Methods: We generated age-adjusted cancer incidence rate estimates for 1-year, 3-year and 5-year time periods (2005–2010) using SEER and ACS population estimates for four racial/ethnic groups by sex and cancer type for residents in the tri-county Detroit area. We calculated incidence rate ratios (IRRs) with corresponding 95% confidence intervals (CIs), and compared trends.
Results: While the IRRs were rarely significant, there were significant differences in incidence rate estimates for Hispanic males. Additionally, interpretation of trends varied by the estimate source: the ACS-based lung cancer incidence rate estimate for Hispanic females increased from 70.59 (95% CI 44.85, 110.67) to 86.13 (95% CI 54.83, 132.44) per 100,000 women from 2007 to 2010, while the SEER incidence rate estimate decreased from 80.76 (95% CI 53.36, 119.24) to 73.54 (95% CI 49.24, 106.62).
Conclusions: Inconsistencies were found when comparing incidence rate estimates for small population groups using the two population estimate sources. This finding has potential implications for health disparities research.

1. Introduction

The American Community Survey (ACS), administered by the U.S. Census Bureau, is a rolling survey that replaced the decennial long form census after the year 2000. It was created to eliminate the need for the long form census and to meet data users' demand for more frequent and current survey data [1]. Whereas the long form census gathered detailed information about population and housing characteristics from a subset of the population (approximately one in six households) once every decade, the ACS was implemented annually and reaches approximately one in forty households per year [2]. When cumulated to yield a five-year period estimate, the most reliable estimate available from the ACS [3], the sampling frame is approximately one in eight households. These smaller samples and procedural differences in survey follow-up[3] have

contributed to higher margins of error in ACS population estimates [4]. Furthermore, survey inclusion rates have been found to differ significantly between demographic groups [5]. This raises concern about the quality of the ACS population estimates.

Several studies have compared implementation issues with the ACS and long form census and their resulting population estimates [6–10]. Among the findings of these studies was the recognition that demographic estimates may differ between the two surveys [6]. However, no studies to date have considered the implications of this for the estimation of cancer incidence or other disease rates.

Due to the large population of Arab Americans in metropolitan Detroit and other urban areas in the country, the Detroit Surveillance, Epidemiology, and End Results (SEER) program has tracked the cancer burden among this population subgroup. Since Arab American is not a federally recognized racial/ethnic group, we historically have estimated population numbers based on decennial long form census responses to ancestry, country of origin, and language spoken at home [11–15]. These same questions are part of the ACS and could be used for the same purpose; however, due to

* Corresponding author at: 312 E. Canfield, Suite 310, Detroit, MI 48201, United States.
E-mail address: jmantey@umich.edu (J. Mantey).

the uncertainty surrounding the accuracy of the ACS data for smaller populations and geographies [16], we wanted to first assess the comparability of ACS population estimates to those from SEER using racial/ethnic data based on self-report of race/ethnicity. SEER uses population estimates from extrapolated decennial census data [17] and is considered the gold standard in this observational study.

We compared cancer incidence rate estimates for four different racial/ethnic groups calculated using ACS population estimates as population denominators to estimates of the same rates calculated using Surveillance, Epidemiology, and End Results (SEER) population estimates. If the resulting incidence rate estimates and trends associated with those estimates were consistent between the two methods, we and others could confidently use ACS data to continue to research health disparities among minority population groups such as Arab Americans.

2. Methods

2.1. Databases

This observational study was conducted using public use databases. The SEER Program of the National Cancer Institute collects and publishes data on cancer incidence and survival in the U.S. Through population-based registries, the organization collects data on patient demographics and clinical information by primary tumor site. SEER is an authoritative source of cancer incidence information within the U.S. [18].

SEER provides a statistical software program named SEER*Stat [19] that can be used to produce incidence and survival statistics for user-defined populations [20]. SEER*Stat age-adjusted cancer incidence rate estimates are considered the gold standard in this study. We compared incidence rate estimates calculated using population estimates from the ACS to the same rates generated using SEER*Stat. Intercensal population denominator estimates used by SEER are produced by the U.S. Census Bureau's Population Estimates Program by modifying extrapolations from the 2000 Census to meet 2010 Census results [21]. The formula used produces the intercensal estimates using a method referred to as the Das Gupta method, which "assumes that the ratio of the intercensal estimate to the postcensal estimate should follow a geometric progression over the decade" [21].

The ACS is administered by the U.S. Census Bureau. Population estimates from the ACS are available in summary files from the Census Bureau website [22], and as downloadable data from the Integrated Public Use Microdata Series (IPUMS) website [23]. Estimates are available for one-year, three-year, and five-year time periods. We prepared ACS population estimates for comparison to the Detroit SEER area (Wayne, Oakland, and Macomb counties) by obtaining micro data for each period of interest from IPUMS, restricting the datasets to Wayne, Oakland, and Macomb counties, and applying population weights. We then created a new age group variable to group the data using the same 19 age categories used by SEER in calculating age-adjusted rates [13]. We generated summary data for the tri-county area from each time period by sex and for the 4 population sub-groups examined: Non-Hispanic Whites, Non-Hispanic Blacks, Non-Hispanic Asians/Pacific Islanders, and Hispanics.

Before comparing incidence rates calculated using the two different population estimates, we tested if the age distribution of population estimates varied by race, ethnicity, and sex between SEER and ACS using different ACS pooled-period estimates: one-year estimates for 2007 and 2010; three-year estimates for 2005–2007 and 2008–2010; and five-year estimates for 2005–2009 and 2006–2010. We used a 2-sample Kolmogorov-Smirnov test to determine whether the ACS and SEER age and race/ethnicity specific population estimates, as represented by distribution across the age groups, differed significantly. We conducted this test separately for all race/ethnicities in both sexes (48 tests, in total), and found no significant differences (data not shown).

2.2. Incidence rate estimate calculations

Incidence rates were calculated using ACS and SEER population estimates stratified by race/ethnicity for cancers of lung, colon, and all cancers combined for both sexes, as well as prostate cancer in males and breast cancer in females for the metropolitan Detroit SEER area. SEER*Stat age-adjusted cancer incidence rate estimates, cancer incidence counts, and age-group population were obtained. We reproduced the age-adjusted cancer incidence rate estimates (standardized to the U.S. standard million population for 2000) using Fay and Feuer's method based on gamma intervals to compute rate estimates, variance, and confidence intervals [24]. The same calculations were repeated using ACS population

Table 1
Comparison of Incidence Rate Estimates of All Cancers Among Males from Metropolitan Detroit at Different Time Periods Using ACS and SEER Population Estimates.

	Time Period 1						Time Period 2					
	ACS		SEER		ACS:SEER		ACS		SEER		ACS:SEER	
	Rate	95% CI	Rate	95% CI	IRR	95% CI	Rate	95% CI	Rate	95% CI	IRR	95% CI
5 year	2005–2009						2006–2010					
NHW	590.38	(584.58, 596.22)	599.40	(593.51, 605.34)	0.98	(0.97, 1.00)	601.43	(595.55, 607.35)	600.76	(594.89, 606.68)	1.00	(0.99, 1.02)
NHB	710.09	(697.27, 723.12)	699.79	(687.22, 712.56)	1.01	(0.99, 1.04)	705.44	(692.86, 718.23)	702.48	(689.99, 715.16)	1.00	(0.98, 1.03)
NHAPI	240.42	(219.00, 263.77)	263.46	(238.42, 291.26)	0.91	(0.80, 1.04)	233.32	(213.18, 255.24)	257.83	(234.25, 283.91)	0.90	(0.80, 1.03)
HISPANIC	525.28	(490.00, 562.77)	583.93	(544.72, 625.74)	0.90	(0.82, 0.99)	605.82	(564.09, 650.22)	562.37	(524.61, 602.57)	1.08	(0.98, 1.19)
3 year	2005–2007						2008–2010					
NHW	589.95	(582.43, 597.54)	603.64	(595.97, 611.38)	0.98	(0.96, 1.00)	594.51	(587.01, 602.09)	595.05	(587.53, 602.64)	1.00	(0.98, 1.02)
NHB	687.16	(671.08, 703.58)	695.80	(679.51, 712.44)	0.99	(0.96, 1.02)	701.47	(685.44, 717.84)	700.49	(684.51, 716.79)	1.00	(0.97, 1.03)
NHAPI	268.46	(235.96, 305.09)	276.16	(240.84, 316.84)	0.97	(0.81, 1.17)	258.69	(224.38, 300.36)	258.11	(229.27, 290.77)	1.00	(0.84, 1.20)
HISPANIC	537.42	(490.44, 588.63)	601.50	(549.10, 658.58)	0.89	(0.79, 1.01)	657.79	(599.21, 721.43)	556.04	(509.26, 606.60)	1.18	(1.04, 1.34)
1 year	2007						2010					
NHW	610.97	(597.75, 624.42)	621.09	(607.71, 634.70)	0.98	(0.95, 1.01)	600.76	(587.68, 614.08)	597.54	(584.59, 610.74)	1.01	(0.97, 1.04)
NHB	718.63	(689.859, 748.58)	712.25	(684.10, 741.44)	1.01	(0.95, 1.07)	699.92	(672.02, 728.91)	687.24	(660.29, 715.13)	1.02	(0.96, 1.08)
NHAPI	309.94	(241.46, 399.76)	279.09	(220.79, 353.17)	1.11	(0.80, 1.54)	231.61	(191.38, 279.59)	272.57	(223.20, 333.11)	0.85	(0.65, 1.11)
HISPANIC	553.15	(461.05, 662.71)	537.09	(453.00, 634.98)	1.03	(0.81, 1.31)	986.04	(817.20, 1185.12)	552.83	(475.25, 641.06)	1.78	(1.38, 2.31)

ACS: American Community Survey. SEER: Surveillance, Epidemiology and End Results Program. NHW: non-Hispanic white, NHB: non-Hispanic black, NHAPI: non-Hispanic Asian/Pacific Islander. All Rates are per 100,000 population. IRR: Incidence Rate Ratio.