Rates of Cancer Incidence and Mortality

Mandi Yu, PhD
National Cancer Institute

Workshop on 2020 Census Data Products: Data Needs and Privacy Considerations

December 12, 2019

Disclaimer

- *The opinions expressed in this presentation are the author's own and do not reflect the view of the National Cancer Institute, National Institutes of Health, the Department of Health and Human Services, or the United States government.*
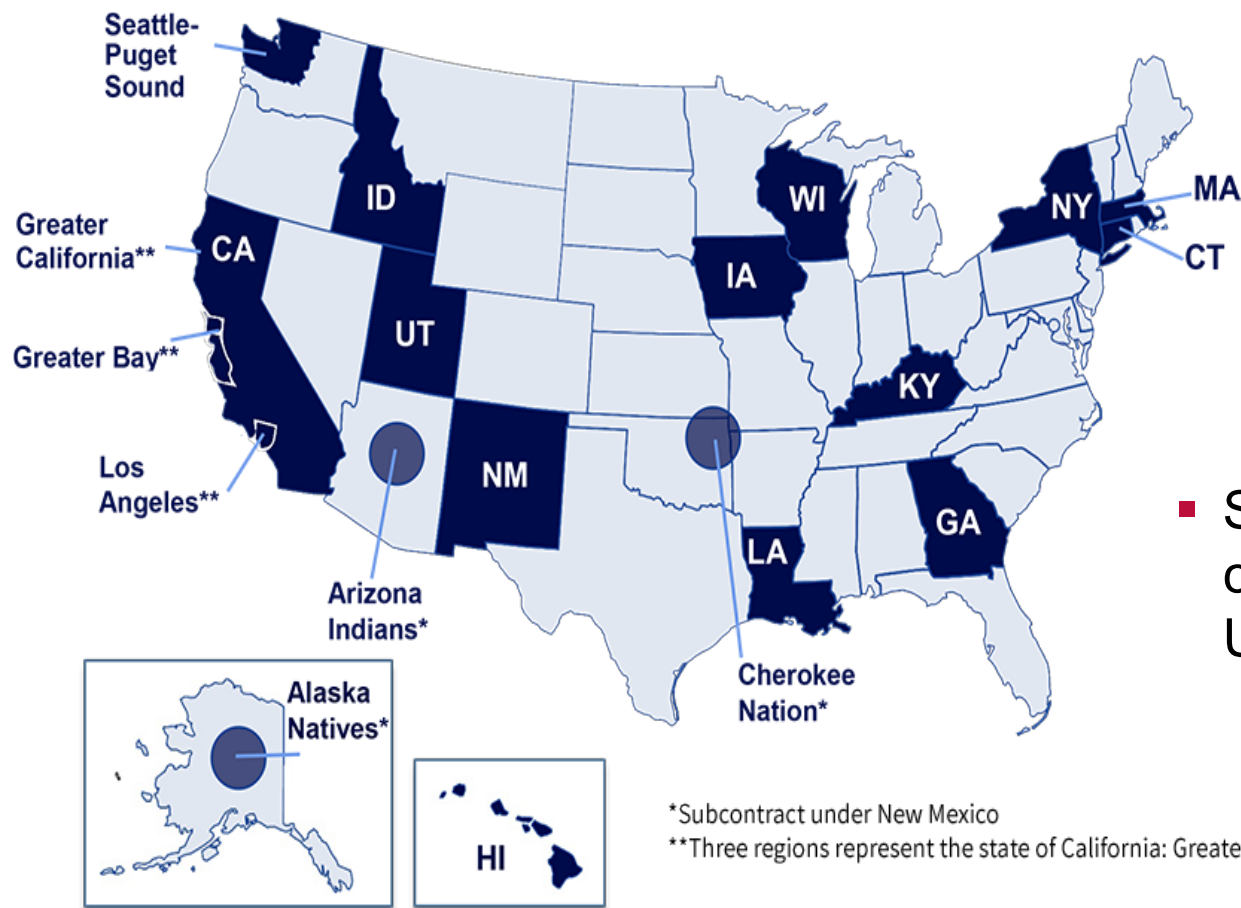
Collaborators

- Dr. Zaria Tatalovich, National Cancer Institute

- Dr. Kathleen Cronin, National Cancer Institute

What is SEER

- Surveillance, Epidemiology, and End Results (SEER) program is established and supported by the National Cancer Institute since 1973

- SEER collects and publishes cancer incidence and survival data from population-based Central Cancer Registries across the nation

- Data collected include patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status

- SEER is a mainstay of the National Cancer Program - it provides baseline measures of cancer rates essential to public health surveillance and focuses our research on the most important problems

U.S. Map with SEER Registries shown

Seattle-Puget Sound

Greater California**

Greater Bay**

Los Angeles**

Arizona Indians*

Alaska Natives*

Cherokee Nation*

ID · CA · UT · NM · WI · IA · KY · NY · MA · CT · GA · LA · HI

*Subcontract under New Mexico
**Three regions represent the state of California: Greater Bay, Los Angeles, and Greater California

- SEER registries cover ~ 34.6% the U.S. population.

Federal Cancer Surveillance Funding Agencies

- Cancer Incidence & Survival
  - **NCI SEER Program** (1973)

- Cancer Incidence
  - **CDC National Program of Cancer Registries (NPCR)** (1992)

- Mortality
  - **CDC National Center for Health Statistics (NCHS)** (1930s)

- Populations
  - **Bureau of the Census** (1800s)

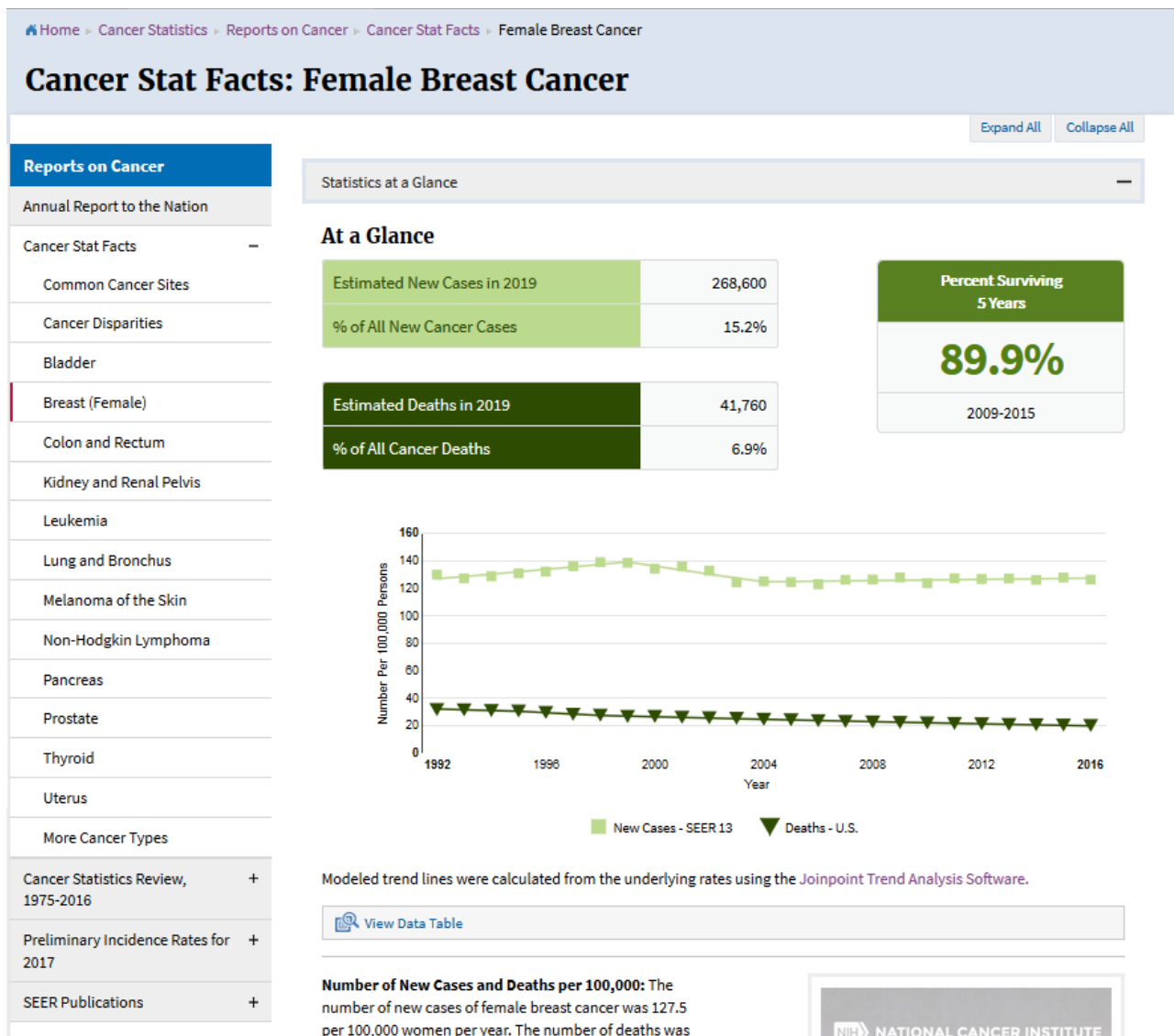# Types of Cancer Statistics Produced and Released by SEER

- **Statistics that Require Population Data as Denominators**

  - Rates of Cancer Incidence (new cases)

  - Rates of Mortality (deaths)

  - Temporal Trends of Cancer Rates of Incidence and Mortality --- measured by Annual Percent Change (APC) in rates, or Average Annual Percent Change (AAPC) in rates over a specified time frame

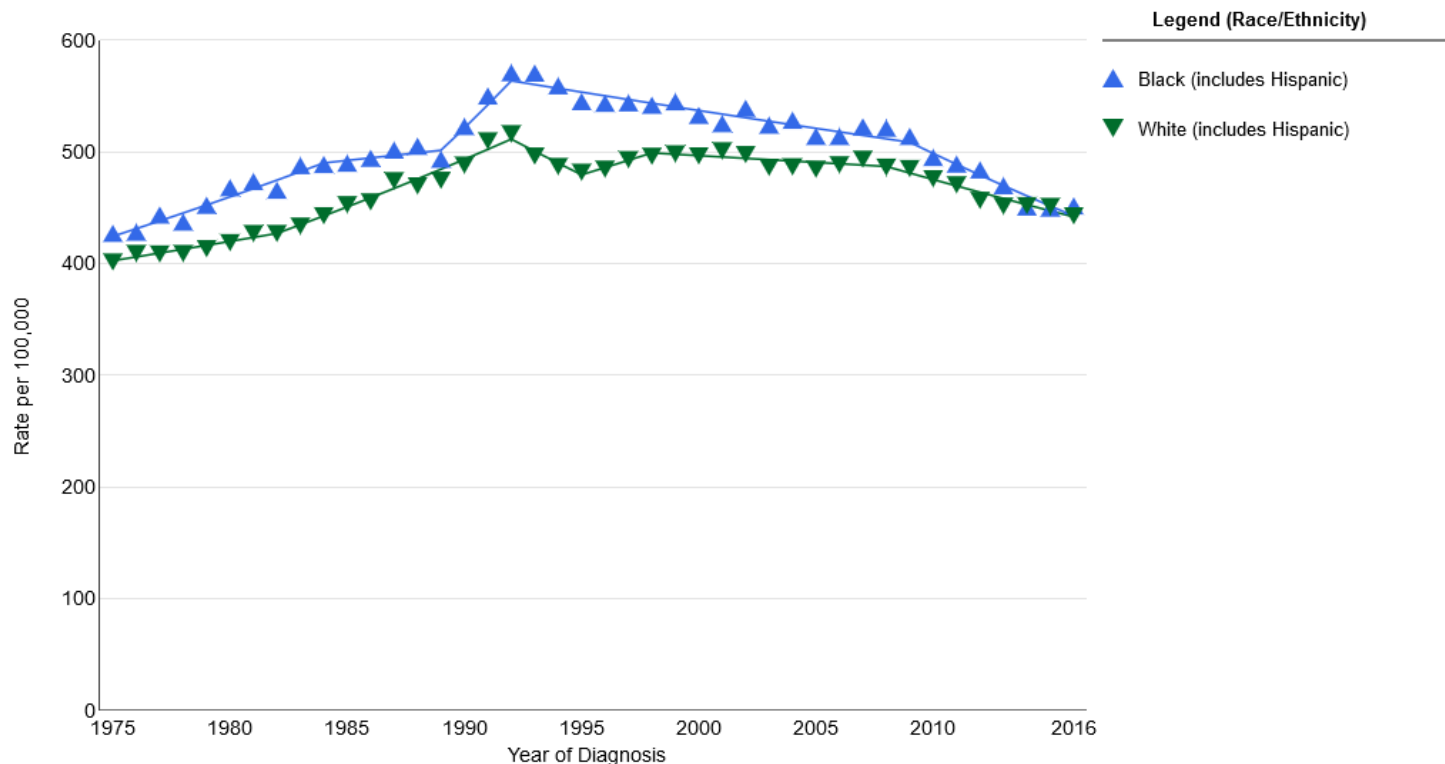- **Statistics that do Not Requires Population Data**

  - Cancer Survival --- Relative survival

  - Cancer Prevalence --- Prevalence of people alive with a previous diagnosis of cancer

  - Risk of Developing/Dying --- Probability of developing or dying of cancer over a lifetime

Examples of Statistics --- Cancer Stat Facts
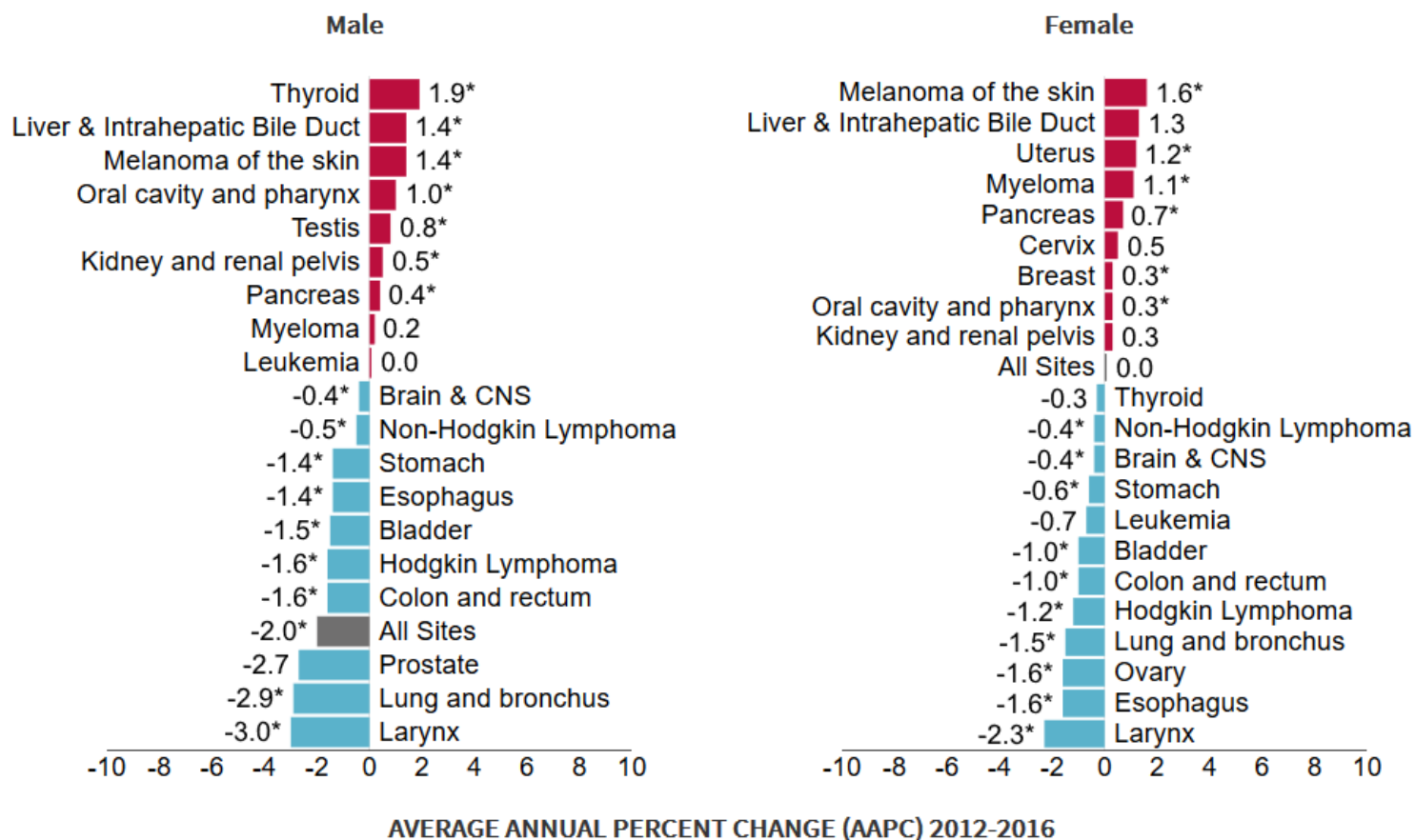
Long-Term Temporal Trends by Race

All Cancer Sites Combined
Long-Term Trends in SEER Incidence Rates, 1975-2016
By Race/Ethnicity
Both Sexes, All Ages, Observed Rates

SEER 9 areas [http://seer.cancer.gov/registries/terms.html] (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, and Atlanta).
Rates are per 100,000 and are age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130).
The Annual Percent Change (APC) estimates were calculated from the underlying rates using the Joinpoint Trend Analysis Software [http://surveillance.cancer.gov/joinpoint], Version 4.6,
February 2018, National Cancer Institute.
The APC's direction is "rising" when the entire 95% confidence interval (C.I.) is above 0, "falling" when the entire 95% C.I. is lower than 0, otherwise, the trend is considered stable.
For years prior to 1990, the Census Bureau has only provided county-level population estimates for White, Black, and "Other" races.
See SEER*Explorer Cancer Site Definitions [https://seer.cancer.gov/explorer/cancer-sites.html] for details about the coding used for SEER Incidence data.
Created by https://seer.cancer.gov/explorer on Wed Dec 11 2019.

Incidence Trends Measured by AAPC (Average Annual Percent Change) in Rates --- Recent Trends (5 year)

NATIONAL INCIDENCE TRENDS

Male

Site	AAPC
Thyroid	1.9*
Liver & Intrahepatic Bile Duct	1.4*
Melanoma of the skin	1.4*
Oral cavity and pharynx	1.0*
Testis	0.8*
Kidney and renal pelvis	0.5*
Pancreas	0.4*
Myeloma	0.2
Leukemia	0.0
Brain & CNS	-0.4*
Non-Hodgkin Lymphoma	-0.5*
Stomach	-1.4*
Esophagus	-1.4*
Bladder	-1.5*
Hodgkin Lymphoma	-1.6*
Colon and rectum	-1.6*
All Sites	-2.0*
Prostate	-2.7
Lung and bronchus	-2.9*
Larynx	-3.0*

Female

Site	AAPC
Melanoma of the skin	1.6*
Liver & Intrahepatic Bile Duct	1.3
Uterus	1.2*
Myeloma	1.1*
Pancreas	0.7*
Cervix	0.5
Breast	0.3*
Oral cavity and pharynx	0.3*
Kidney and renal pelvis	0.3
All Sites	0.0
Thyroid	-0.3
Non-Hodgkin Lymphoma	-0.4*
Brain & CNS	-0.4*
Stomach	-0.6*
Leukemia	-0.7
Bladder	-1.0*
Colon and rectum	-1.0*
Hodgkin Lymphoma	-1.2*
Lung and bronchus	-1.5*
Ovary	-1.6*
Esophagus	-1.6*
Larynx	-2.3*

AVERAGE ANNUAL PERCENT CHANGE (AAPC) 2012-2016
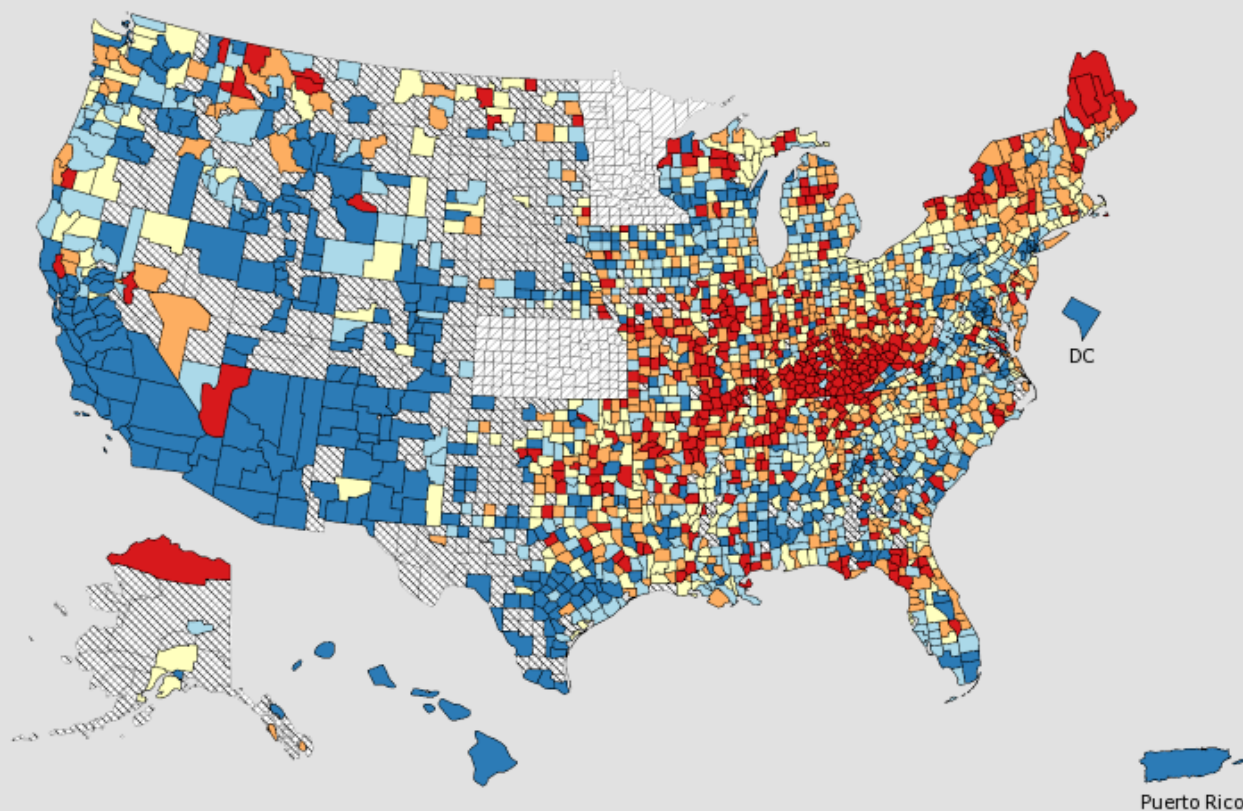
*AAPC is significantly different from zero (p<.05)
Rates were adjusted for reporting delay in the registry

Lung & Bronchus Cancer Incidence Rates for US Counties

Incidence Rates[†] for United States by County
Lung & Bronchus, 2012 - 2016
All Races (includes Hispanic), Female, All Ages

Populations Used in Routine Reports of Cancer Rates

- Intercensal and Vintage annual time series of July 1st county estimates produced by the **U.S. Census Bureau's Population Estimates Program,** in collaboration with **NCHS,** with support from the **NCI** through an interagency agreement

- These are Vintage Populations (predictions based on the previous census) and/or Intercensal Populations (benchmarked using two censuses) estimated based on demographic models

- These estimates are stratified by 19 Age Grp (with <1 yr, and 85+), Sex, Single Race (White, Black, AIAN, API), and Hispanic Origin

- For comparing rates, each rate is associated with an uncertainty measure, i.e. a standard error, which is estimated by assuming population is measured without error

Several Important Features of Census's Vintage and Intercensal Population Estimates that May Be Affected by Differential Privacy Procedure

- Population estimates are at the County level

  - Cancer statistics are released for counties, although not all are interpretable

  - Registries would producing rates for their own geographic areas

  - Some registries are comprised of one county or several counties, e.g. Los Angeles Registry --- 1 county with pop ~10 million; Seattle-Puget Sound Registry --- 13 counties with a total pop ~5 million

- 19 age group --- rates are routinely age-adjusted to allow (international or subgroup) comparison of rates across populations with different age distributions.

- Single-race categories (at least for the next 5 years) to match race data in the incidence data.
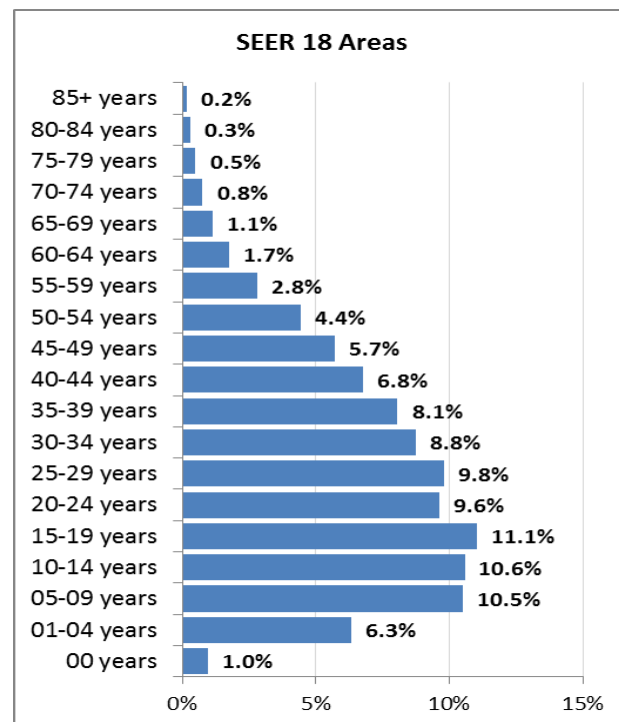
In addition, would the new DP procedures affect Census's Time Series Population Estimates?

- We don't know until comparisons are made

- In general, there are concerns over possible impacts of new/different statistical procedures on continuity of time series estimates

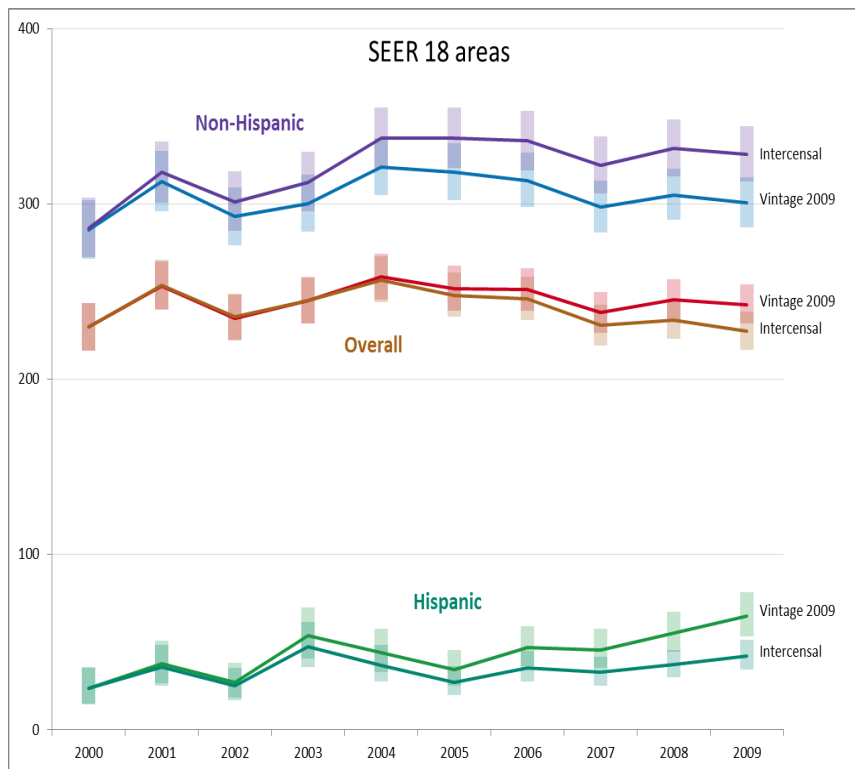- An example with 2010 Census race data editing ('imputation')

Example: Impact of 2010 race data editing on continuity of time series estimates

- Census 2010 implemented an improved race data editing procedure and the result of it is larger numbers of the Hispanic "Some other race" responses to AIAN using internal data

- Intercensal 2001-2009 estimates are based on 2000 and 2010 data

- Vintage 2009 estimates are predictions based solely on 2000 data

- Differences between Intercensal and Vintage are most notable in 2009 estimates

Figure shows Percent of the total increase in the 2009 AIAN Hispanic populations by age group

SEER 18 Areas

Age group	Percent
85+ years	0.2%
80-84 years	0.3%
75-79 years	0.5%
70-74 years	0.8%
65-69 years	1.1%
60-64 years	1.7%
55-59 years	2.8%
50-54 years	4.4%
45-49 years	5.7%
40-44 years	6.8%
35-39 years	8.1%
30-34 years	8.8%
25-29 years	9.8%
20-24 years	9.6%
15-19 years	11.1%
10-14 years	10.6%
05-09 years	10.5%
01-04 years	6.3%
00 years	1.0%

Impact on SEER Rates for AIAN

SEER 18 areas

- Solution --- Census Bureau Estimation Program have worked with NCI/CDC to update 2000 data using 2010 race editing procedure and modified the intercensal 2001-2009 estimates.

Acknowledge the findings from: Espey DK, Jim MA, Richards TB, Begay C, Haverkamp D, Roberts D. Methods for improving the quality and completeness of mortality data for American Indians and Alaska Natives. Am J Public Health. 2014 Jun;104 Suppl 3:S286-94.

Several Observations/Considerations

- Consider assign epsilon to certain large sub-state population, e.g. large county like Los Angeles county.

- Make available the uncertainty in DP modified population data. Uncertainty in population data not only affects estimating standard errors for rates, but also affects estimating face values of the rates (presence of uncertainty in denominator generally lead to attenuation of rates --- results from a NCI study)

- Temporal consistency checks on Census's Estimation Products --- modify previous census data and subsequently intercensal and vintage populations --- additional efforts