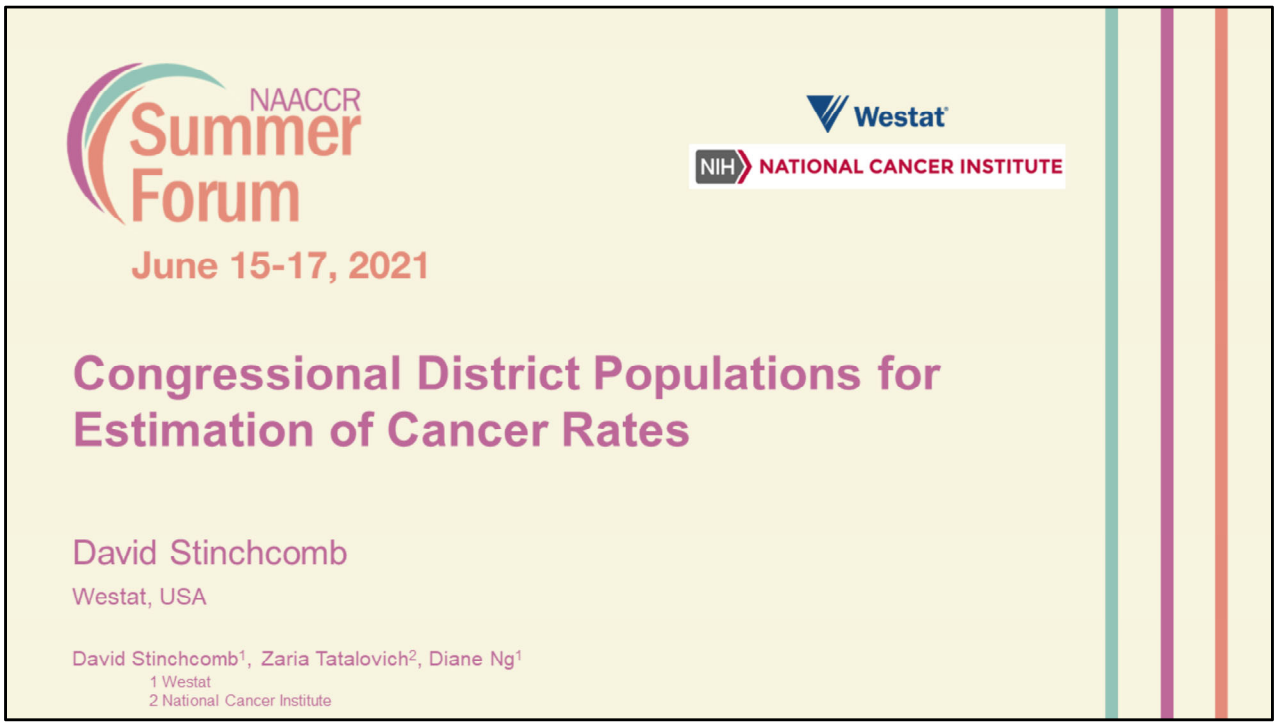
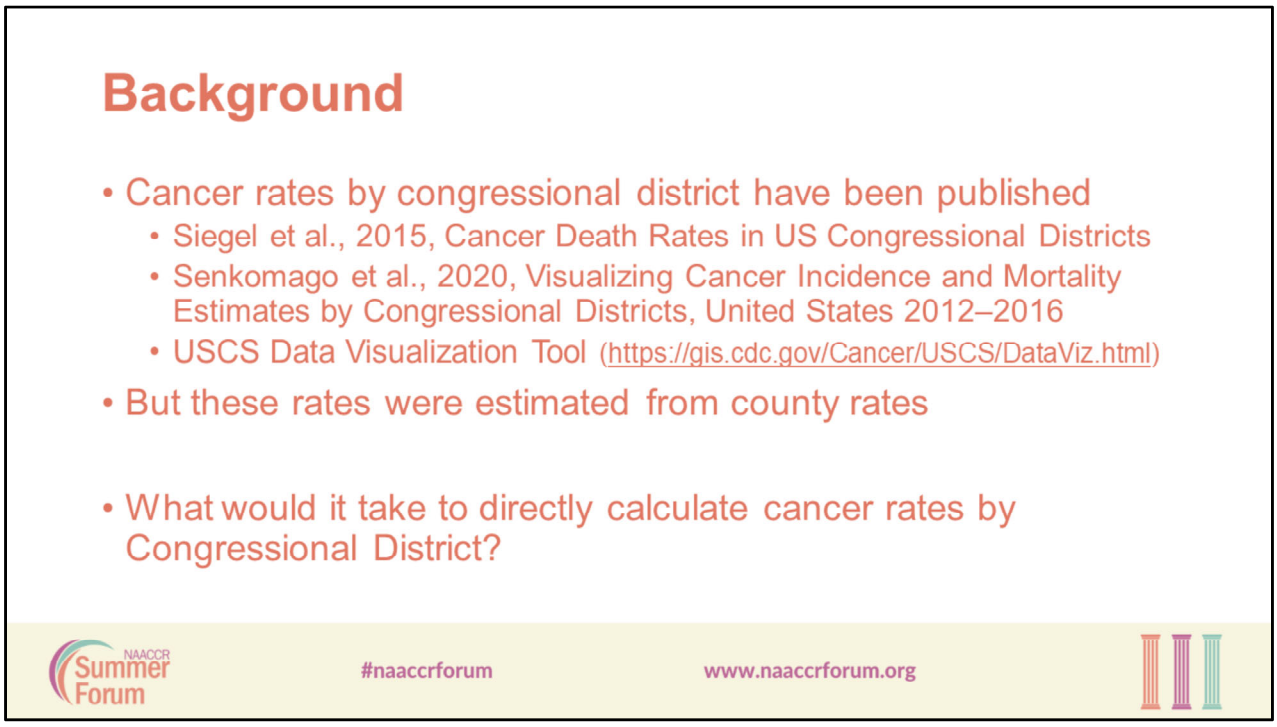


The previous talks in this session have described the activities of a workgroup that has been meeting to discuss population denominators. As part of these activities, we were asked to take a look at possible ways to estimate populations for Congressional Districts in the United States.

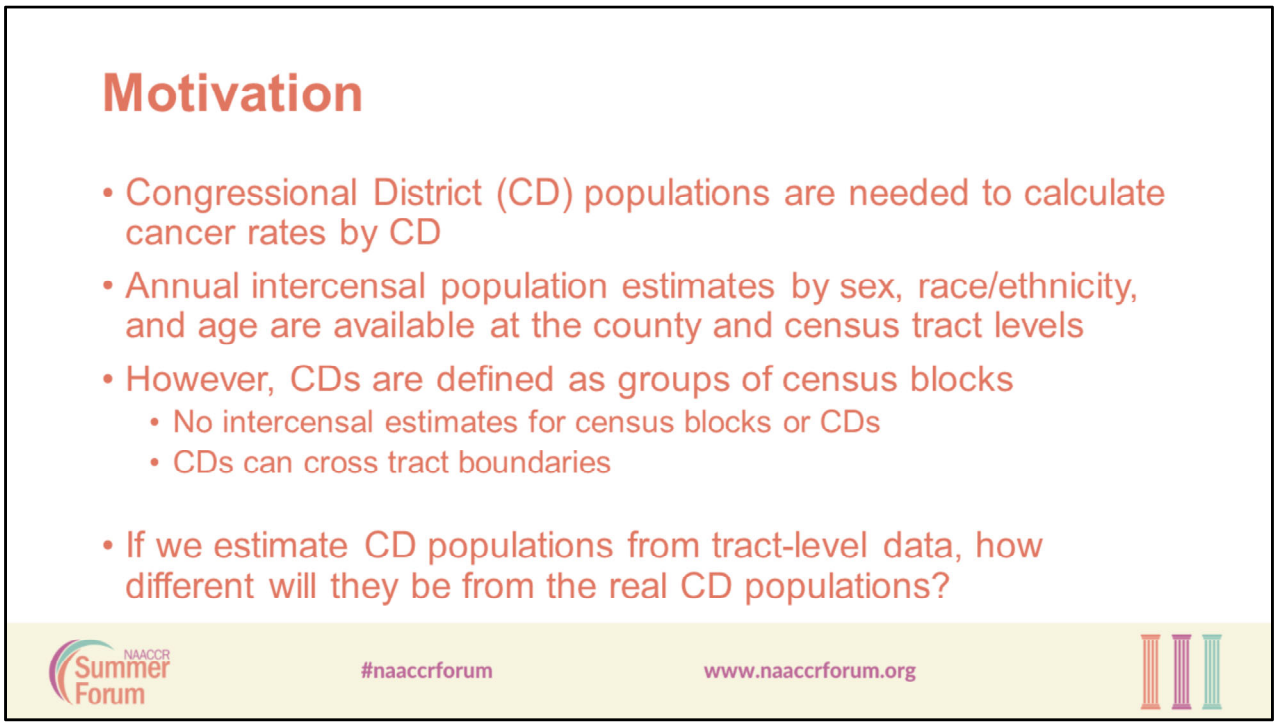
Outline

1. Background and motivation
2. Methods and results
3. Overall findings and next steps

This talk will follow the usual outline with some background information, methods, results, and conclusions.



In 2015, cancer death rates by Congressional District were published by a team from the American Cancer Society. More recently, a team from the CDC published Congressional District incidence and mortality rates and these rates are available on the USCS Data Visualization web site. But the rates for both these studies were estimated from county rates. We wondered what it would take to directly calculate Congressional District cancer rates.



The main challenge is with the population denominators. We have annual intercensal estimates by sex, race/ethnicity and age at the county and census tract levels. But Congressional Districts are collections of census blocks – the lowest level of the census geographies. There are no intercensal population estimates for census blocks or for Congressional Districts. And Congressional Districts can and do cross tract boundaries.

This study sought to answer the question: if we estimate Congressional District populations from the tract-level populations, how different will they be?

Analysis process: three basic steps

1. Identify tracts that are split between CDs and determine the population in each piece
2. Assign each split tract to a single CD
3. Calculate CD populations by sex, race/ethnicity, and age for both the real CDs and the tract-estimated CDs and compare the differences

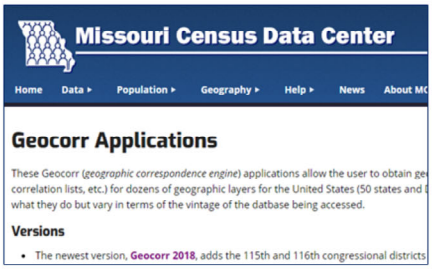
We did this in three steps. First, we identified tracts that are split between Congressional Districts. Next, we assigned each tract to a single Congressional District. And finally, we compared the population differences.

The next set of slides will provide methods and results for each of these steps in turn.

Step 1 identify split tracts – methods

We compared two ways to identify split tracts and determine the population in each piece:

- The Missouri Geocorr tool



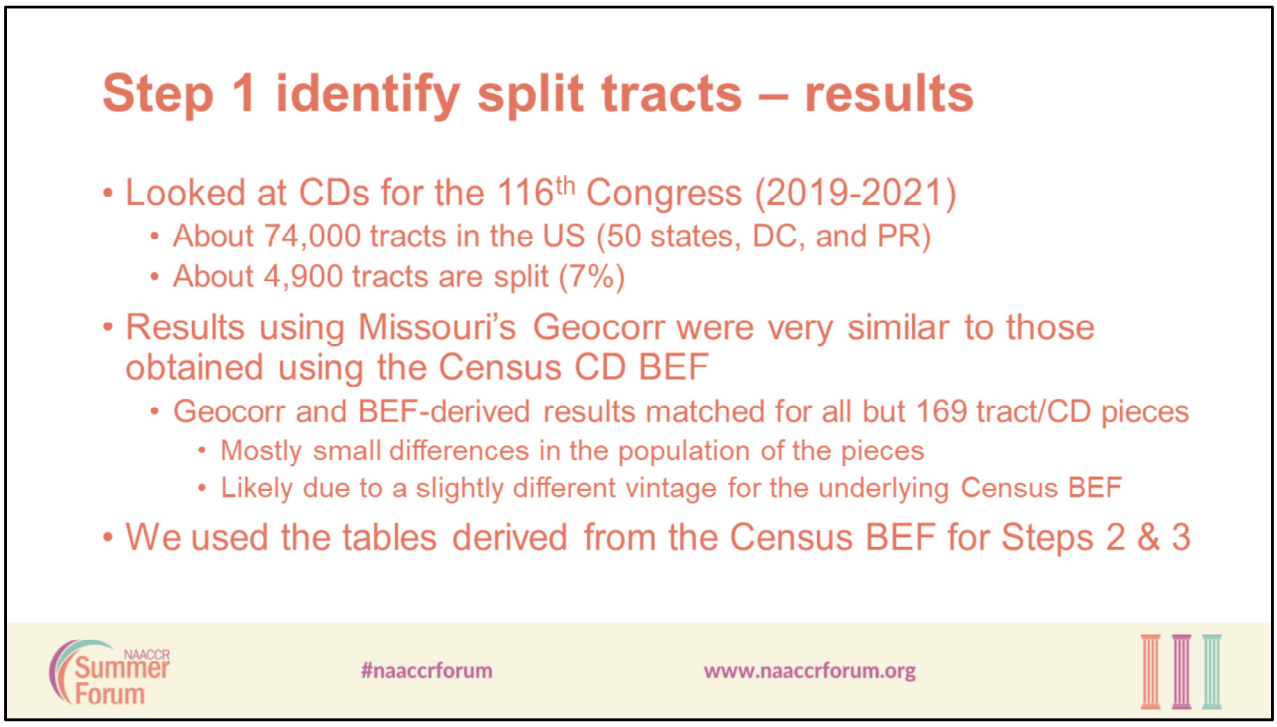
https://mcdc.missouri.edu/applications/geocorr.html

- The Census Block Equivalency File (BEF) and a SAS program

https://www.census.gov/geographies/mapping-files/2019/dec/rdo/116-congressional-district-bef.html

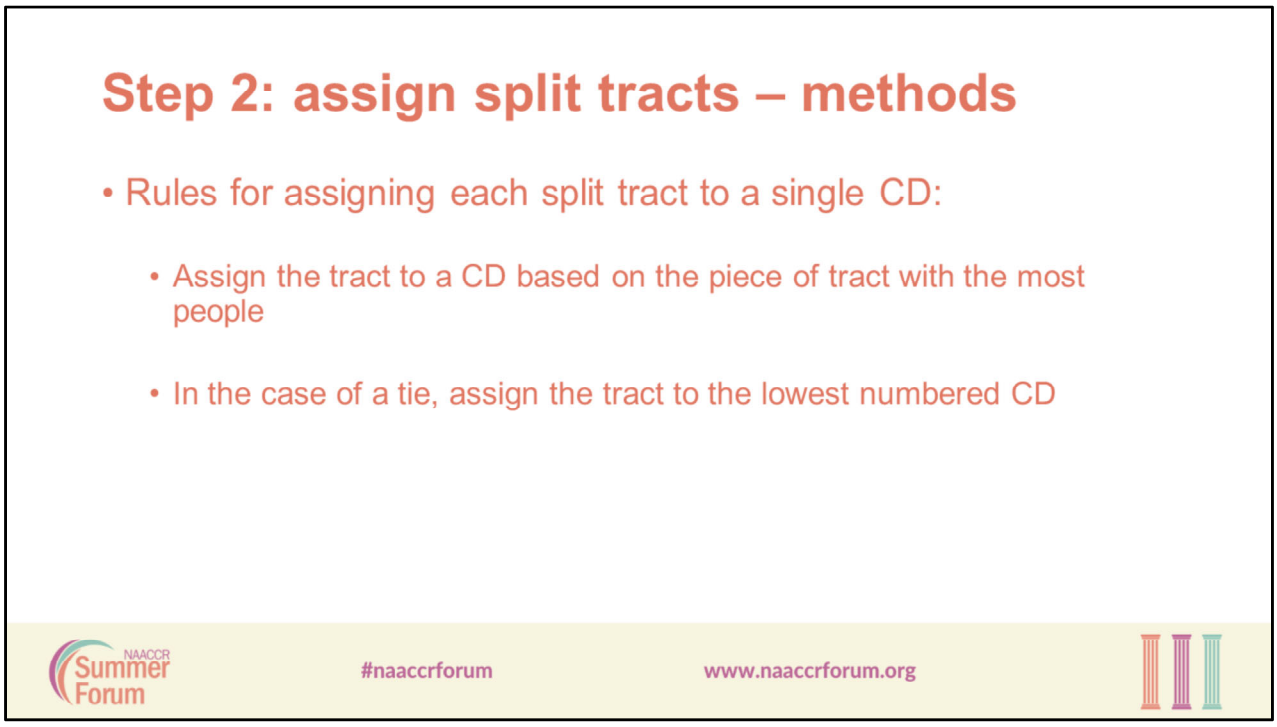
For Step 1, there are a couple of ways to identify the tracts that are split between Congressional Districts. The Missouri Census Data Center has a tool that provides tables showing how two different geographic levels intersect. You can use this tool to get a table with populations of tracts and pieces of tracts that are in different Congressional Districts.

Alternatively, you can create the same type of table by taking the Census-provided Block Equivalency File for Congressional Districts and using a small SAS or R program to summarize by census tract. For this step, we used both methods.

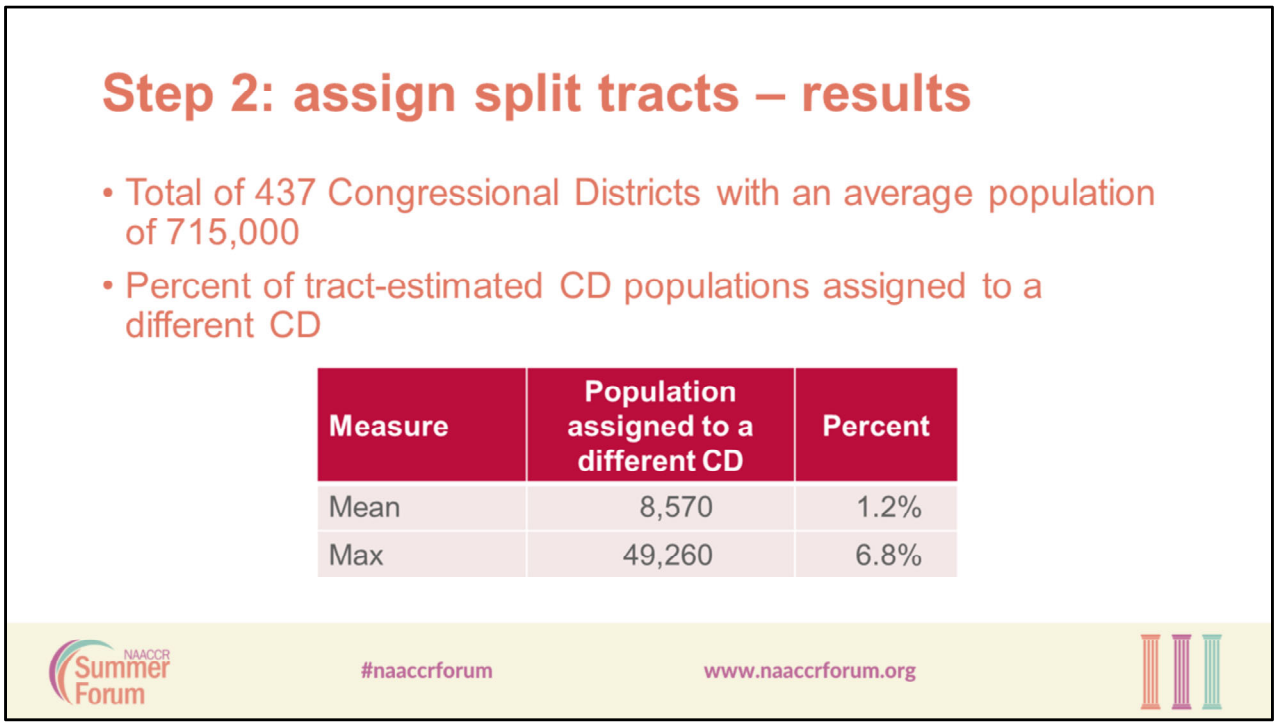


We looked at the Congressional Districts for the 116th Congress which was the current congress at the time of this study. Of the 74,000 tracts in the U.S., only about 4900 or 7% are split.

The two methods yielded very similar results: all but 169 of the tract/Congressional District records were identical and the differences between the 169 records were small. We used the tables derived from the Census Block Equivalency File for the rest of the study.



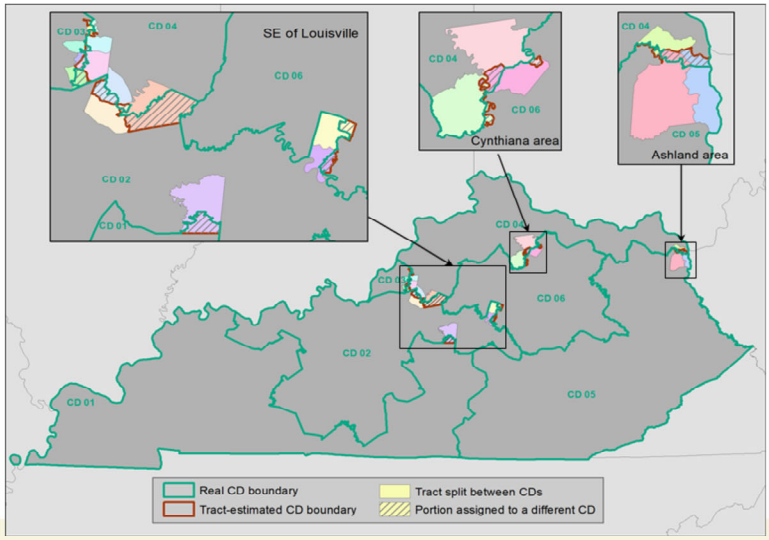
In the second step, we set up some simple rules to assign each split tract to a Congressional District. We just assigned the tract to the Congressional District that had the piece with the most people. In the rare case of a tie, we just picked the lowest numbered Congressional District.



In the U.S., there are 437 Congressional Districts with an average population of about 715,000 people. We looked at the number of people who were assigned to a different Congressional District by this assignment process. The average across the 437 Congressional Districts was about 8600 people or 1.2% of the Congressional District population. In the worst case, 6.8% of the population was assigned to a different Congressional District.

Step 2: assign split tracts – results for Kentucky

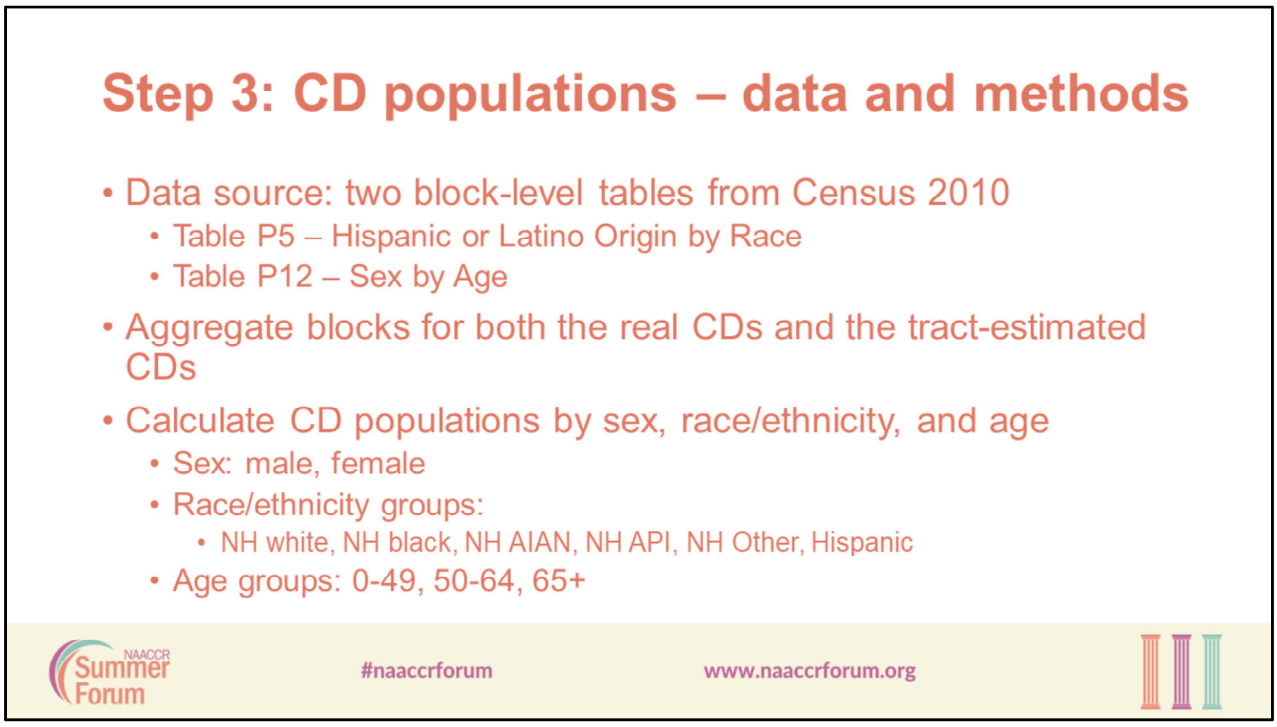
- 6 CDs
- 1,115 tracts
- 19 split tracts
- Population assigned to a different CD
  - Average of 2,631 people per CD
  - 0.4% of the CD population



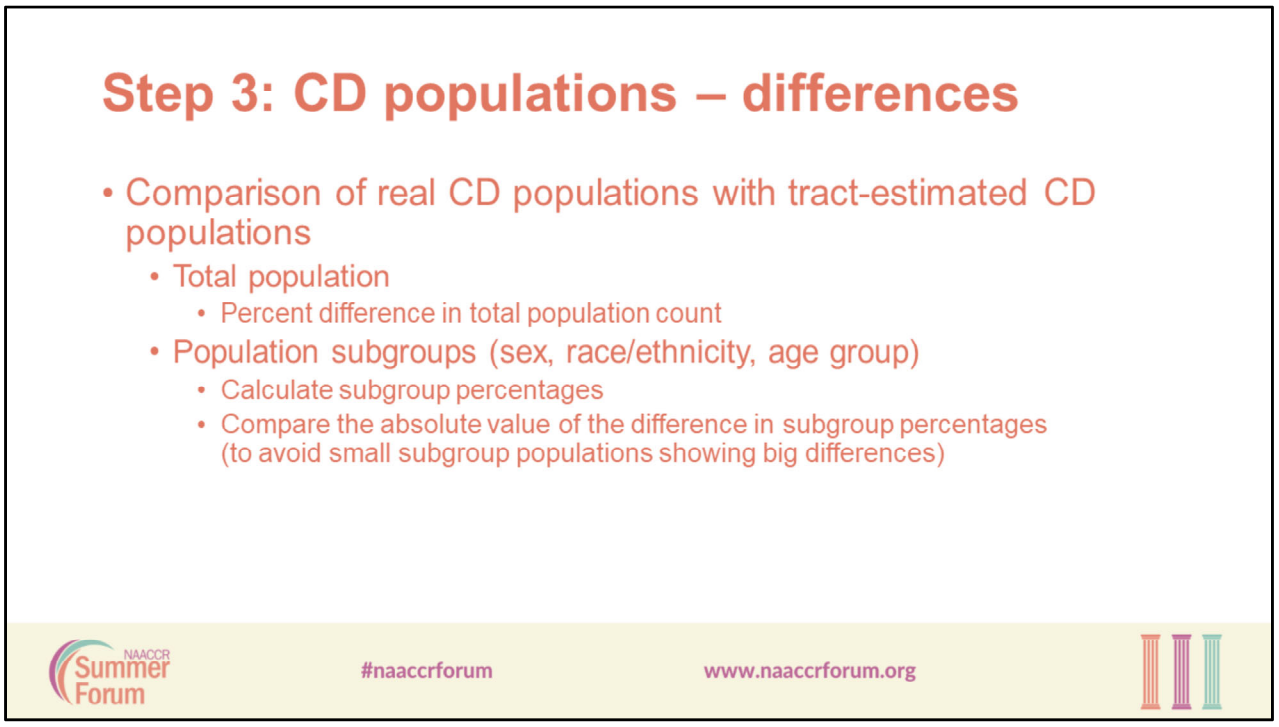
NAACCR Summer Forum #naaccrforum www.naaccrforum.org

Here is an example showing how this worked in Kentucky. Kentucky has 6 Congressional Districts and 1,115 tracts. Of these, only 19 tracts were split between Congressional Districts and these tracts are shown as colored areas in the map. The cross-hatched areas are the portion of the tract assigned to a different Congressional District. As you can see, the hatched areas are generally quite small. In terms of the population, an average of about 2600 people were assigned to a different Congressional District representing less than a half a percent of the population.

But the real question is, how different are the **characteristics** of the populations between the real Congressional Districts and the tract-estimated Congressional Districts?



In the third step, we compared the differences in the characteristics of the populations. We took two block-level tables from the 2010 census, one with counts by race/ethnicity and one with counts by sex and age. We aggregated these counts for both the real Congressional Districts and the tract-estimated Congressional Districts. We then compared the populations by sex, race/ethnicity, and age using the groupings shown here.



We calculated statistics based on the differences. For the total population we looked at the percent difference. For the population subgroups, we looked at the absolute value of the difference in subgroups percentages.

Step 3: CD populations – results

- Total of 437 Congressional Districts, average population 715,000
- Differences in total populations:

Measure	Total Diff	Pct Diff
Mean pop diff	3,585	0.5%
Max pop diff	40,147	5.6%

- Top 5 CD population differences:

State	CD	Real Pop	Tr-level Pop	Pct Pop Diff
Illinois	11	712,813	672,666	5.6
Illinois	14	712,813	738,410	3.6
Texas	02	698,488	721,578	3.3
Texas	07	698,488	675,643	3.3
Texas	09	698,488	716,865	2.6

NAACCR Summer Forum #naaccrforum www.naaccrforum.org

For the 437 Congressional Districts, the average difference in the total population is quite small: about a half of a percent. In the worst case, the difference was 5.6% for a Congressional District in Illinois. As can be seen in the second table, after the 5.6% difference for District 11 in Illinois, the percentages fall off pretty quickly.

Step 3: CD populations – comparisons by sex

- Maximum pop difference by sex: US and top 10 states

US/State	CD Count	PopMale MaxDiff	PopFem MaxDiff
_Total US	437	0.22%	0.22%
Washington	10	0.22%	0.22%
New Jersey	12	0.16%	0.16%
Texas	36	0.16%	0.16%
Illinois	18	0.12%	0.12%
Louisiana	6	0.12%	0.12%
Ohio	16	0.12%	0.12%
Oregon	5	0.11%	0.11%
Wisconsin	8	0.10%	0.10%
Florida	27	0.09%	0.09%

Looking at the differences by population subgroup, they are all quite small. This table shows the maximum differences by sex. The largest difference is well below half a percent.

Step 3: CD populations – comparisons by race/ethnicity

- Maximum pop difference by race/ethnicity: US and top 10 states (for NH-White)

US/State	CD Count	NH-White MaxDiff	NH-Black MaxDiff	NH-AIAN MaxDiff	NH-API MaxDiff	Hispanic MaxDiff
_Total US	437	1.37%	1.30%	0.17%	0.68%	1.13%
Texas	36	1.37%	0.67%	0.01%	0.21%	1.13%
Illinois	18	1.20%	0.43%	0.01%	0.16%	0.92%
South Carolina	7	1.00%	1.01%	0.00%	0.02%	0.02%
Louisiana	6	0.97%	1.02%	0.00%	0.01%	0.08%
Alabama	7	0.91%	0.91%	0.00%	0.02%	0.09%
Florida	27	0.80%	1.30%	0.00%	0.05%	0.49%
Maryland	8	0.79%	0.70%	0.00%	0.26%	0.17%
Tennessee	9	0.76%	0.70%	0.00%	0.04%	0.10%
Ohio	16	0.75%	0.62%	0.00%	0.04%	0.12%

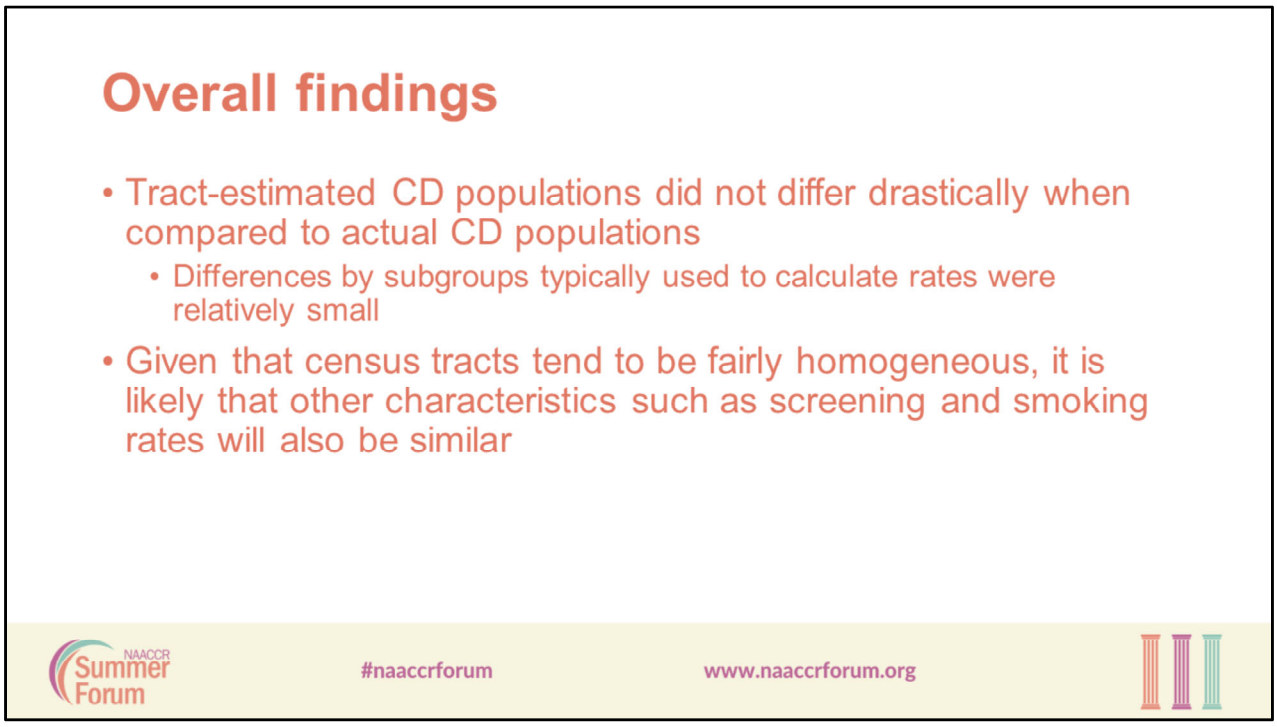
By race/ethnicity, some of the differences are a bit bigger but they are all still small. This table shows maximum differences for the top 10 states ranked by the differences for non-Hispanic Whites. In the worst case, the difference is about 1.4% and most differences are less than 1%. The differences for non-Hispanic blacks are similar with a worst-case difference of about 1.3%. Differences for the other race/ethnic groups are even smaller.

Step 3: CD populations – comparisons by age group

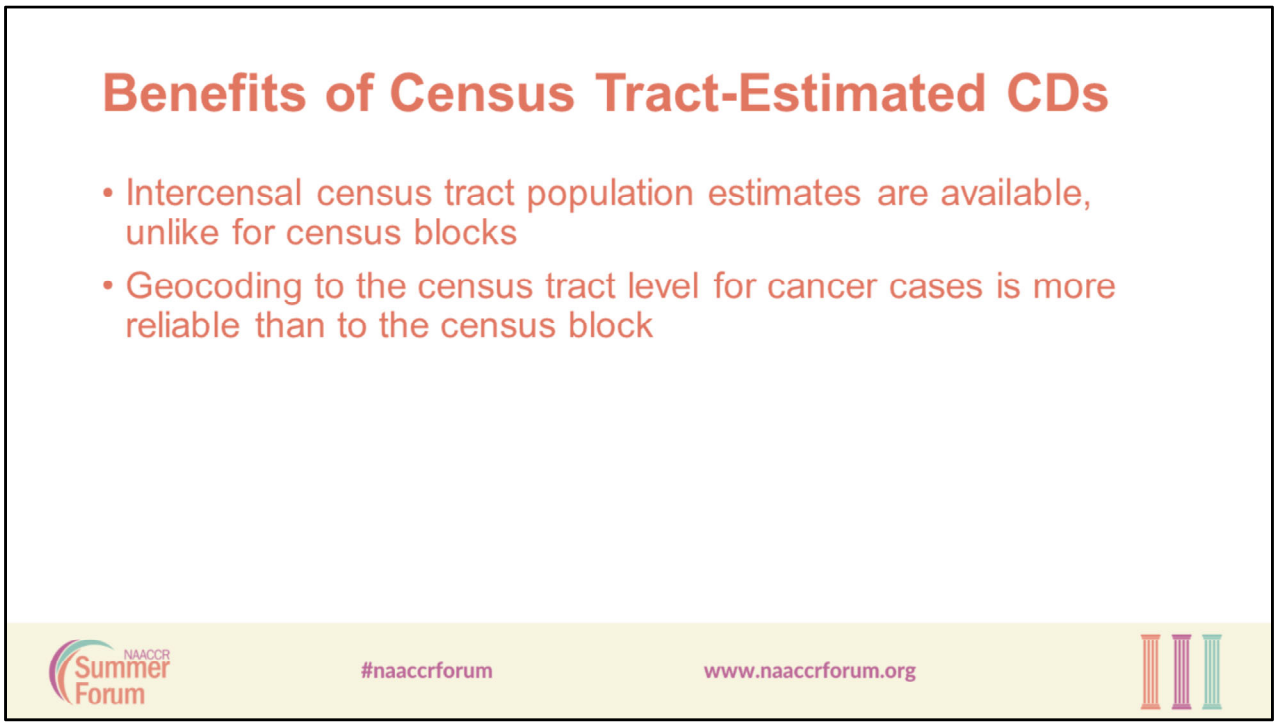
- Maximum pop difference by age group: US and top 10 states (for Age 0 to 49)

State	CD Count	Age 0 to 49 MaxDiff	Age 50 to 64 MaxDiff	Age 65-plus MaxDiff
_Total US	437	0.56%	0.28%	0.33%
Texas	36	0.56%	0.26%	0.31%
Ohio	16	0.53%	0.28%	0.24%
Illinois	18	0.48%	0.21%	0.33%
Florida	27	0.37%	0.21%	0.22%
California	53	0.36%	0.16%	0.24%
Louisiana	6	0.27%	0.12%	0.15%
New Mexico	3	0.26%	0.13%	0.13%
Virginia	11	0.22%	0.17%	0.26%
Maryland	8	0.21%	0.09%	0.14%

The maximum differences by age group are all less than 1%.



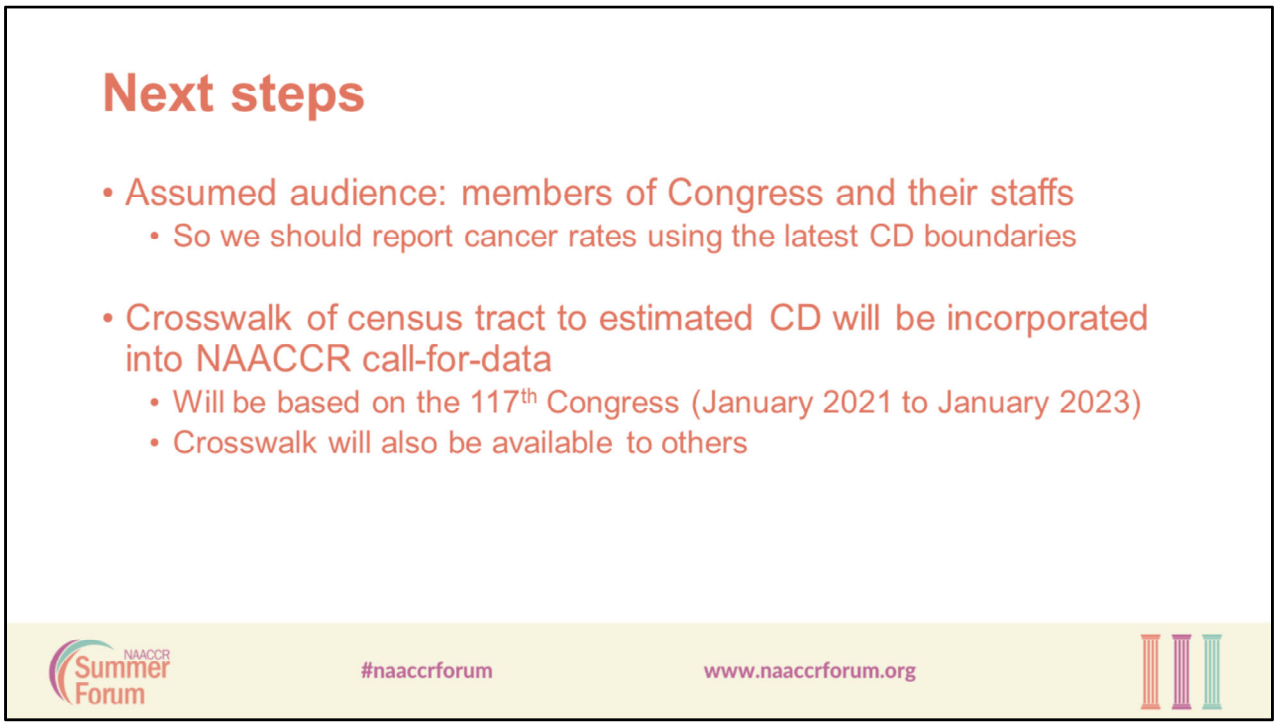
Overall, we found that the populations for the real Congressional Districts and the tract-estimated Congressional District were very similar, both for the total population and the for the population subgroups. Since census tracts are intended to be fairly homogeneous, it is likely that other population characteristics will also be similar.



Using census tracts as the basis for calculating Congressional District cancer rates has advantages for both the numerator and the denominator.

For the denominator, we have the annual intercensal tract population estimates that the other talks in this session have described.

For the numerator, the geocoding of cancer cases to the tract level is fairly well established and reviewed whereas block-level geocoding is not likely to be as reliable.



For reporting Congressional District cancer rates, we assume that the main audience consists of the members of Congress and their staff. So, when we disseminate cancer rates by Congressional District, it would be best to use the geographic boundaries for the current Congress (even though the cancer statistics might be for several years earlier).

The next NAACCR call-for-data will include a crosswalk between tracts and Congressional Districts. This will enable registries to add a Congressional District identifier to cancer records that have tract-level geocodes. The crosswalk will be based on the current 117th Congress. We can also provide the crosswalk to anyone else who is interested.

I look forward to answering any questions you may have. Thank you.