

Life Course Indicators Tip Sheet: Racial Residential Segregation (LC-16)

About AMCHP

The Association of Maternal & Child Health Programs (AMCHP) is a national resource, partner and advocate for state public health leaders and others working to improve the health of women, children, youth and families, including those with special health care needs.

Life Course Indicators

AMCHP launched a project designed to identify and promote a set of indicators that can be used to measure progress using the life course approach to improve maternal and child health. This project was funded with support from the [W.K. Kellogg Foundation](#).

To find more tools and resources to help you use the life course indicators, visit:

Life Course Metrics Project:
amchp.org/programsandtopics/d

What are the data elements for the Racial Residential Segregation (RRS) calculation of the Dissimilarity Index (DI)?

Each data element in the DI formula is described in the table below:

RRS Formula Variable	Label	Variable in Table DP05 from American Factfinder
x_i	Census tract minority population estimate	HC01_VC50 (For non-Hispanic black)
X	County minority population estimate	HC01_VC50 (For non-Hispanic black)
y_i	Census tract reference group population estimate	HC01_VC49 (For non-Hispanic white)
Y	County reference group population estimate	HC01_VC49 (For non-Hispanic white)

The variable for census tract and county level population estimates is the same since both estimates are coming from the same data table (ACS DP05). The sample SAS code available for this calculation shows how to deal with this when merging your data sets to perform the RRS calculation.

What exactly is the DI formula telling me to do?

$$D = .5 * \sum_{i=1}^n |x_i / X - y_i / Y|$$

A breakdown of the steps to calculate the dissimilarity index is below:

1. x_i/X → Divide the minority population in census tract i by the minority population in the county
2. y_i/Y → Divide the reference population in census tract i by the reference population in the county
3. $|x_i/X - y_i/Y|$ → Find the absolute value of the quotient from step 1 – the quotient from step 2
4. Perform steps one through three for each census tract in the county and sum these absolute values
5. Multiply the sum from step four by 0.5
6. The product of step five is the dissimilarity index for the county

Where can I download the necessary American Community Survey (ACS) data files and which ones should I use?

You can download the ACS data files for this calculation under the American FactFinder advanced search tab. A step-by-step process for pulling these tables from American FactFinder is shown on the next page. To perform the calculation the way that AMCHP analysts calculated RRS, for which there is sample SAS code available, you must download two of the same data files with different geography. The data file is DP05 – ACS Demographic and Housing Estimates. Download

this file at your larger geographic level and then again at your smaller geographic level. For example, if you are going to calculate RRS on the county level, download DP05 once for county level geography and once for census tract level geography.

This tip sheet uses examples from a county level calculation with census tracts as the smaller geographic components using non-Hispanic black as the minority population and non-Hispanic white as the reference population. It is appropriate to substitute lower or higher geographic levels and other minority populations for which data is available, keeping in mind how those substitutions will change the interpretation of the dissimilarity index.

Pulling tables from American FactFinder

Go to American FactFinder (factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml), select “Advanced Search,” and search for the table name in the “topic or table name” box. When you get to a screen like that of Figure 1 below, select your desired geography from the menu on the left (see notes in Figure 2). Check the box next to the table you want, and then select download to get as zip file for the full set of census tracts. Figure 3 shows what an individual table looks like – you do not want to be copying these down one tract at a time.

Figure 1. List of tables from America FactFinder

The screenshot shows the American FactFinder search results page. The top navigation bar includes 'U.S. Department of Commerce', 'United States Census Bureau', and 'AMERICAN FactFinder'. The search results section is titled 'Search Results: 1-7 of 7 tables and other products match "Your Selections"'. A table lists the following results:

ID	Table, File or Document Title	Dataset	About
<input type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 American Indian and Alaska Native SF	?
<input checked="" type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 Demographic Profile SF	?
<input type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 SF1 100% Data	?
<input type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 SF2 100% Data	?
<input type="checkbox"/>	DP-1 Profile of General Demographic Characteristics: 2000	2000 SF1 100% Data	?
<input type="checkbox"/>	DP-1 Profile of General Demographic Characteristics: 2000	2000 SF2 100% Data	?
<input type="checkbox"/>	DP-1 Profile of General Demographic Characteristics: 2000	2000 SF4 Sample Data	?

Figure 2. Example of selecting census tract geography

The screenshot shows the 'Select Geographies' interface with the following elements:

- Buttons: List, Name, Address, Map
- Section: Select geographies to add to Your Selections
- Radio buttons: most requested geographic types, all geographic types
- Dropdown: Select a geographic type: Census Tract - 140
- Dropdown: Select a state: Arizona
- Dropdown: Select a county: -- select a county --
- Text: Select one or more geographic areas and click Add to Your Selections:
- List box: All Census Tracts within Arizona
- Button: ADD TO YOUR SELECTIONS

Annotations:

- Blue arrow pointing to the state dropdown: Select your state and then select "All census tracts within state."
- Blue arrow pointing to the list box: Click: Add to your selections.
- Green text: Tip: do not select an individual county

Figure 3. Example of a table for a single census tract

Subject	Number	Percent
HOUSEHOLDS BY TYPE		
Total households	447	100.0
Family households (families) [7]	343	76.7
With own children under 18 years	148	33.1
Husband-wife family	175	39.1
With own children under 18 years	81	18.1
Male householder, no wife present	53	11.9
With own children under 18 years	20	4.5
Female householder, no husband present	115	25.7
With own children under 18 years	47	10.5
Nonfamily households [7]	104	23.3
Householder living alone	82	18.3
Male	59	13.2
65 years and over	10	2.2
Female	23	5.1
65 years and over	16	3.6
Households with individuals under 18 years	222	49.7
Households with individuals 65 years and over	138	30.9

Annotations:

- Text: This is what the table looks like for a single census tract.
- Text: Select the table you want (check the box) and then choose [Download](#) to get the final file with all census tracts.
- Green arrow pointing to the table name: Tip: clicking on the table name displays one tract at a time

1 Selected: [View](#) [Download](#) [Compare](#) [Clear All](#) ?

ID	Table, File or Document Title	Dataset	About
<input type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 American Indian and Alaska Native SF	?
<input checked="" type="checkbox"/>	DP-1 Profile of General Population and Housing Characteristics: 2010	2010 Demographic Profile SF	?

Should I use one-year, three-year, or five-year estimates?

AMCHP recommends using five-year estimates when calculating RRS on the county level or below so that the smaller jurisdictions have high enough counts to perform the calculation.

How can I handle small population estimates in some census tracts?

Even using the five-year estimates, some census tracts will have very small population estimates. The formula for the dissimilarity index does not account for normal flow of population in and out of communities, census tracts, counties, and states. Therefore it is important to interpret the dissimilarity index for areas with small populations with caution as the DI can be heavily influenced by small changes in the population. This point is illustrated in the example below:

County A has a population of 1050:

Black population (X) = 50

White population (Y) = 1000

The county has 3 census tracts:

Census tract i:

Black population (x_i) = 20

White population (y_i) = 500

Census tract j:

Black population (x_j) = 15

White population (y_j) = 300

Census tract k:

Black population (x_k) = 15

White population (y_k) = 200

DI Calculation:


$$\left| \left(\frac{20}{50} - \frac{500}{1000} \right) \right| + \left| \left(\frac{15}{50} - \frac{300}{1000} \right) \right| + \left| \left(\frac{15}{50} - \frac{200}{1000} \right) \right|$$
$$\left| 0.4 - 0.5 \right| + \left| 0.3 - 0.3 \right| + \left| 0.3 - 0.2 \right| = 0.2 = \text{well integrated}$$

One family of 5 from the minority population moves from census tract i to census tract k. What happens to the DI? The DI calculation is now:


$$\left| \left(\frac{15}{50} - \frac{500}{1000} \right) \right| + \left| \left(\frac{20}{50} - \frac{300}{1000} \right) \right| + \left| \left(\frac{15}{50} - \frac{200}{1000} \right) \right|$$
$$\left| 0.3 - 0.5 \right| + \left| 0.4 - 0.3 \right| + \left| 0.3 - 0.2 \right| = 0.4 = \text{moderately segregated}$$

The movement of 5 people from one census tract to another census tract within the county caused DI to go from well integrated to moderately segregated, which shows the pitfalls of using the DI for areas with very small populations.

Studies have used different thresholds for a population size cutoff when using the DI. The threshold for defining a “small population size” for the larger geographic area ranges from below 50,000 to below 10,000. For the minority population of interest, a commonly used minimally significant estimate is 1,000. If the county population on a whole is below 10,000 or the minority population of the county is below 1,000, these results should be interpreted with caution. Analysts have grouped lower geographic levels together to handle the small population estimate issue. For example, analysts calculating the DI on the county level have systematically grouped census tracts together in order to use larger geographic areas than census tract to increase population counts. If you do systematically group census tracts to calculate county level RRS, it is important to realize you are now looking at those groups of census tracts as the components of the county. The DI is now measuring the evenness with which the reference population and minority population are distributed across those *groups of census tracts* within a certain county, not across the actual census tracts of that county.

How can I handle zero population estimates in some census tracts?

If both the reference and minority population are equal to zero for a census tract, we recommend removing these census tracts from the calculation. It is possible that there could be census tracts with airports, open fields, etc that actually have zero population or it is possible that values have been suppressed or are truly missing. If a census tract with a reference population of zero and a minority population of zero is factored into the RRS formula it will falsely drive down the dissimilarity index for that county even though there is no population data available for that census tract. If a census tract has zero minority population, but still has a reference population estimate greater than zero, these should be kept in the calculation as there are people living in that census tract. If you drop any census tracts from the calculation, it is important to note this somewhere in your interpretation, because the DI is now only measuring the evenness with which the reference population and minority population are distributed across the remaining, non-zero census tracts.

Can I calculate RRS for other minority populations?

Yes, you can also analyze the evenness of other minority populations. Population estimates available in the ACS table DP05 include black, American Indian and Alaska Native, Asian, Native Hawaiian and other Pacific Islander, and Hispanic

or Latino populations. However, given the example above on small populations, depending on the state, county, and census tract, the DI for some minority groups may need to be interpreted with caution due to small population estimates. Grouping census tracts as discussed above is one option to handle small minority populations. Another option is to sum all minority population estimates and calculate a DI for all minorities, still using white population estimates as the reference group.

What is the Dissimilarity Index (DI) telling me?

The DI is a demographic measure of the evenness with which two groups are distributed across the smaller geographic areas that make up a larger area. For example, you could take a county and measure the evenness with which white residents and black residents are distributed across census tracts in that county. Since the measure is an index, the highest value you can calculate is 1.0, while the lowest value is zero. A DI of 1.0 calculated for white residents and black residents for a county would be telling you that white residents live in all-white census tracts and black residents live in all-black census tracts. The lowest value you could calculate is zero, which would be telling you that in that particular county, there was a completely even distribution of black residents and white residents across census tracts. In reality, we find dissimilarity indices somewhere in between.

How do I interpret my results?

Using the DI you can characterize jurisdictions as “well integrated” ($DI < 0.3$), “moderately segregated” ($0.3 \leq DI \leq 0.6$), and “highly segregated” ($DI > 0.6$). If county x has a DI of 0.65 calculated from census tracts with white residents as the reference population and black residents as the minority population, the interpretation would be: Based on the distribution of white residents and black residents across census tracts in county x, county x is highly segregated.

Limitations of the Dissimilarity Index

There are limitations for this measure and checking to make sure your measures make sense with your data is good practice. Limitations include:

- DI can only measure the evenness of two groups at a time. To investigate segregation among multiple populations in an area, you have to calculate multiple DIs and can only calculate the evenness of the distribution for a minority population with one reference population.
- DI does not give information about spatial patterns of segregation. A census tract may be clustered into racially segregated neighborhoods but if there is evenness of minority population and reference population among census tracts within the county, the DI will indicate the county is well integrated. Mapping techniques that allow you to examine clusters of neighborhoods might be better tools for examining spatial patterns.
- The strength and reliability of the measure is affected by small population sizes.

Do you have any SAS code examples for this calculation?

Yes, AMCHP staff developed SAS code for an example calculation of RRS for Alabama counties is below. You can also download the code [here](#).

```
**County Level calculation for Alabama;
```

```
libname census " N:\jfarfalla\US Census\ACS\County level calculation\County Level Data ";
```

```
**Download 2 DP05 data tables from American FactFinder: 1 with census tract level geography and 1 with county level geography
```

```
**Import both data sets;
```

```
PROC IMPORT OUT= CENSUS.County  
    DATAFILE= "N:\jfarfalla\US Census\ACS\County level calculati  
on\County Level Data\ACS_13_5YR_DP05_County.csv"  
    DBMS=CSV REPLACE;  
    GETNAMES=YES;  
    DATAROW=2;
```

```
RUN;
```

```
PROC IMPORT OUT= CENSUS.Tract  
    DATAFILE= "N:\jfarfalla\US Census\ACS\County level calculati  
on\County Level Data\ACS_13_5YR_DP05_tract.csv"  
    DBMS=CSV REPLACE;
```

```

GETNAMES=YES;
DATAROW=2;
RUN;

**Check the contents of each data file;
Proc contents data=census.Tract;
proc contents data=census.County;
run;

**Create a census tract level data set and county level data set with only necessary
variables
& rename black and white population estimate variables so I can eventually merge the data
sets;
data census.ALTract;
Set census.Tract (Rename=(HC01_VC49=totWtract HC01_VC50=totBtract));
Keep GEO_ID GEO_ID2 GEO_Display_Label totWtract totBtract;
Run;

**Do the same step for county level data;
data census.ALCounty;
Set census.county (Rename=(HC01_VC49=totWcnty HC01_VC50=totBcnty));
Keep GEO_ID GEO_ID2 GEO_Display_Label totWcnty totBcnty;
Run;

**Check the contents of the new shorter data sets with renamed population estimate
variables;
Proc contents data=census.ALTract;
proc contents data=census.ALCounty;
run;

**Create a county ID variable to merge by in each data set - ensures county level
population estimates will populate for each census tract;
**Convert the population estimates to numeric variables;
Data census.Counties;
Set census.ALcounty;
Length County_ID $5; **limits the county ID variable to 5 characters;
blackcounty=totBcnty*1;**converts character variable to numeric;
whitecounty=totWcnty*1;
county_ID= SUBSTR(Geo_ID2, 1, 5); **Create county ID;
Keep blackcounty whitecounty county_ID; /*dropped Geo_id and Geo_id2*/
Run;

Data census.Tracts;
Set census.ALTract;
Length County_ID $5;
whitetract=totWtract*1;
blacktract=totBtract*1;
county_ID= SUBSTR(Geo_ID2, 1, 5);
Keep GEO_Display_Label whitetract blacktract county_ID;
Run;

proc format; **Format county_ID with county names;
value $County_ID 01001="Autauga"
                 01003="Baldwin"
                 01005="Barbour"
                 01007="Bibb"
                 01009="Blount"
                 01011="Bullock"
                 01013="Butler"
                 01015="Calhoun"
                 01017="Chambers"
                 01019="Cherokee"

```

```
01021="Chilton"  
01023="Choctaw"  
01025="Clarke"  
01027="Clay"  
01029="Cleburne"  
01031="Coffee"  
01033="Colbert"  
01035="Conecuh"  
01037="Coosa"  
01039="Covington"  
01041="Crenshaw"  
01043="Cullman"  
01045="Dale"  
01047="Dallas"  
01049="DeKalb"  
01051="Elmore"  
01053="Escambia"  
01055="Etowah"  
01057="Fayette"  
01059="Franklin"  
01061="Geneva"  
01063="Greene"  
01065="Hale"  
01067="Henry"  
01069="Houston"  
01071="Jackson"  
01073="Jefferson"  
01075="Lamar"  
01077="Lauderdale"  
01079="Lawrence"  
01081="Lee"  
01083="Limestone"  
01085="Lowndes"  
01087="Macon"  
01089="Madison"  
01091="Marengo"  
01093="Marion"  
01095="Marshall"  
01097="Mobile"  
01099="Monroe"  
01101="Montgomery"  
01103="Morgan"  
01105="Perry"  
01107="Pickens"  
01109="Pike"  
01111="Randolph"  
01113="Russell"  
01115="St. Clair"  
01117="Shelby"  
01119="Sumter"  
01121="Talladega"  
01123="Tallapoosa"  
01125="Tuscaloosa"  
01127="Walker"  
01129="Washington"  
01131="Wilcox"  
01133="Winston";
```

Run;

```
**Merge county and state data sets by County ID;  
proc sort data=census.Tracts; by county_ID; run;
```

```

proc sort data=census.Counties; by county_ID; run;

Data census.R77;
Merge census.tracts (in=a)
      census.counties (in=b) ;
By county_ID;
if a;
If whitetract EQ 0 and blacktract EQ 0 Then Ditract=.; *set Ditract to missing if no
population in tract;
Else Ditract=abs((blacktract / blackcounty)-(whitetract/whitecounty)); **absolute value
of (xi/X -yi/Y) for each census tract;
If Geo_ID2="Id2" Then delete; **removes header row from census file which is no longer
needed;
Run;

proc freq data=census.R77; **proof of concept - check census tracts are matched to
counties appropriately;
table Geo_display_label*county_ID/norow nocol nopercnt;
Run;

**calculate the sum of dissimilarity index across all tracts;
Proc sort data=census.R77; By Geo_ID2; Run;

proc means data=census.R77 Sum;
var Ditract;
By county_ID;
output out=census.segbycounty sum=dissimindex ;
run;

**Final step in the formula (multiply by 0.5)
D=.5*tract sum of (abs val Xi/X - yi/Y);
data census.R77final;
Set census.segbycounty;
dissimindex=0.5*dissimindex;
Length RRSlevel $25;
If dissimindex LT 0.3 Then RRSlevel="well integrated";
Else If dissimindex GE 0.3 and dissimindex LE 0.6 Then RRSlevel="moderately segregated";
Else If dissimindex GT 0.6 then RRSlevel="highly segregated";
run;

**Print the dissimilarity index for each county;
proc print data=census.R77Final noobs;
var county_ID dissimindex RRSlevel;
format County_ID $County_ID.;
run;

**Dissimilarity index range for all counties in AL;
PROC MEANS data=census.R77Final;
Where dissimindex NE 0;
VAR dissimindex;
RUN;

```

To learn more, please contact Caroline Stampfel, associate director, epidemiology and evaluation, at cstampfel@amchp.org or (202) 775-0436.

Association of Maternal & Child Health Programs

2030 M Street, NW, Suite 350
Washington, DC 20036
(202) 775-0436 • www.amchp.org