TRACKING PROGRAM SUB-COUNTY METHODOLOGY

This method document will outline the steps the Centers for Disease Control and Prevention's Environmental Public Health Tracking Program (Tracking Program) takes to create its sub-county geographies. The Tracking Program uses the term "sub-county" to refer to geographies at a finer resolution than a county. This methodology aggregates census tracts, weighted by block group populations, to a minimum population threshold of 5,000 persons. These geographies are then aggregated, weighted by census tract populations, to a minimum population threshold of 20,000 persons. The methodology also includes steps to remove census tracts with zero population before aggregation occurs, as well as combining counties that have a population below the desired threshold to meet the specified threshold population.

Software required to perform aggregation work
- Mapping software (e.g., ArcGIS, QGIS)
- Geographic Aggregation Tool (GAT)

Data needed prior to aggregation
- 2010 Census block group shapefiles - Cartographic boundary files
- 2010 Census tract shapefiles - Cartographic boundary files
- 2010 Census county shapefiles - Cartographic boundary files
- 2010 Census block group population data
- 2010 Census tract population data
- 2010 Census county population data

Steps to complete before aggregation
1) Load shapefiles and population data tables into mapping software. Join population data tables to respective shapefiles.
2) In the census tract shapefile, select census tracts with a total population of 0. Save these census tracts as a new shapefile.
3) Select all census tracts with a non-zero population or at least one resident. Save this selection as a new shapefile. This will be the census tract shapefile in the following steps.

Steps to aggregate census tracts weighted by block group populations to a minimum population of 5,000 persons
1) Add a new text field with a length of 11 to the block group shapefile called "CTBG_Join." Calculate the field to concatenate the state, county, and census tract fields. This will set the value to the geography's 11-digit census tract FIPS code.
2) Repeat step 1 on the census tract shapefile with no zero population census tracts.
3) Join the census tract shapefile and block group shapefile using the CTBG_Join field and keep only matching records. This will remove any block groups that are within the zero population census tracts. Save this shapefile (e.g., as (jurisdiction)_CTandBG), as this will be the working shapefile in the following steps.
4) Determine the geographic centroids of the working block group shapefile in decimal degrees.

5) In the county shapefile, select counties with fewer than 5,000 population, if there are any. Save these as a separate shapefile (e.g., (jurisdiction)_CountyBelow5k).
6) Join the county shapefile (e.g., (jurisdiction)_County) to the working block group shapefile (jurisdiction)_CTandBGJoin based on county FIPS.
7) If no counties are below 5,000 population, skip to the next step. In (jurisdiction)_CTandBG, select the block groups within counties with at least 5,000 population. Save this selection as a shapefile that will now only have block groups within counties with at least a population of 5,000 (e.g., (jurisdiction)_CTandBG5k). This layer is now ready for aggregation.
8) Using the SAS or R version of the Geographic Aggregation Tool (GAT), specify the shapefile location and name. The ID Variable is CTBG_Join, and merging should occur within COUNTY. Specify the other variables according to your file.
9) The SAS GAT tool creates a .dbf output with a field named "newregion." These are unique identifiers for each new aggregation.
10) Add the .dbf output to mapping software and join it to the shapefile used as input to the GAT, then dissolve the layer based on the "newregion" field. If possible, in software, sum the block group population during the dissolve process. This will create the polygons of the 5,000-person minimum geographies with the summed populations.

To aggregate 5,000-person minimum geographies weighted by census tract population to a minimum population of 20,000 persons
1) Join the output of the SAS GAT tool from the previous steps to the census tract shapefile with no zero population census tracts on the CTBG_Join field and keep only matching records. Then join the final shapefile created in the previous steps to the census tract shapefile. This shapefile should now have all census tracts excluding those with zero population and all the relevant information for aggregation. Save as a new shapefile (e.g., as (jurisdiction)_5kandCT).
2) Calculate the geographic centroids of the (jurisdiction)_5kandCT shapefile in decimal degrees.
3) In the (jurisdiction)_County shapefile, select counties with fewer than 20,000 population, if any. Save these as a separate shapefile (e.g., (jurisdiction)_CountyBelow20k).
4) Join the county shapefile to the working census tract shapefile based on county FIPS.
5) If no counties are below 20,000 population, skip to the next step. Select the block groups within counties with at least 20,000 population. Save this selection as a shapefile that will now only have census tracts within counties with at least a population of 20,000 (e.g., (jurisdiction)_5kandCT20k). This layer is now ready for aggregation.
6) Using the SAS or R version of the Geographic Aggregation Tool (GAT), specify the shapefile location and name. The ID Variable is the "newregion" field from the previous aggregation, and merging should occur within COUNTY. Specify the other variables according to your file.
7) The SAS GAT tool creates a .dbf output with a field named "newregion." These are unique identifiers for each new aggregation.
8) Add the .dbf output to mapping software and join it to the shapefile used as input to the GAT, then dissolve the layer based on the "newregion" field. If possible, in software, sum the block group population during the dissolve process. This will create the polygons of the 20,000-person minimum geographies with the summed populations.

To aggregate counties below 5,000 population to a minimum of 5,000 population

1) From previous steps, shapefile should already exist of counties within jurisdictions below 5,000 population. Using this shapefile, load it to mapping software and determine which counties are eligible for being combined. The Tracking Program considers counties eligible if they neighbor a county also under 5,000 population with the summed population of the counties that could be combined being equal to or above 5,000. The Tracking Program considers counties "neighbors" if a person could conceivably cross from one county directly to the other. Save the counties eligible for aggregation in the jurisdiction in a new shapefile.

2) Load census tracts from the jurisdiction to the mapping software. In the census tract and eligible counties shapefile, add a new text field with length 5 named "St_cofips." Calculate the field to concatenate the state and county fields. This will set the value to the geography's 5-digit county FIPS code.

3) Join the eligible counties shapefile to the census tract shapefile based on the "st_cofips" field and keep only matching records. This will now only have census tracts within the eligible counties. Save this as a new shapefile.

4) Calculate the geographic centroids of the working eligible counties shapefile in decimal degrees.

5) Using the SAS or R version of the Geographic Aggregation Tool (GAT), specify the shapefile location and name. The ID Variable is st_cofips, and merging should occur within STATE. Specify the other variables according to your file.

6) The SAS GAT tool creates a .dbf output with a field named "newregion." These are unique identifiers for each new aggregation.

7) Add the .dbf output to mapping software and join it to the shapefile used as input to the GAT, then dissolve the layer based on the "newregion" field. If possible, in software, sum the block group population during the dissolve process. This will create the polygons of the 5,000-person minimum combined county geographies with the summed populations.

Following software aggregation of the combined counties, the Tracking Program then performs a manual review for potential adjustments. Areas reviewed for adjustment are geographies with more than double the minimum population. The goal is to create more polygons and ultimately provide more data. In this next section, combined county areas with a population of 10,000 or more are reviewed.

8) Load the original eligible counties shapefile to mapping software along with the automated aggregation of counties shapefile.

9) Create a "manualID" text field with length 6 in the shapefile of eligible counties.

10) All counties in the eligible counties shapefile will need a manualID assigned. The first one done will be 00001, then 00002, and so forth.

11) Begin with the highest populated areas, specifically with areas more than twice the minimum population (i.e., more than 10,000 residents for the 5,000-population aggregations). These should be the best candidates to be broken up into more shapes. When adjusting boundaries, aim to keep areas compact, contiguous, and with no empty spaces. Examples of software-generated counties that may be candidates for breaking up into more shapes can be seen in Appendix A.

12) After all double-the-minimum-population aggregations have been reviewed, move to the remaining aggregations. For areas below twice the minimum population (i.e., 9,999 or less),

        maintain their groupings using the GAT-created aggregations for reference and assign their manualID as necessary.

    13) Once all counties have been assigned a manualID, then dissolve the shapefile.

Repeat these steps for counties below 20,000 population.

After performing the above-described steps, the following types of geographies should exist for the jurisdiction:

- Census tracts with zero population
- 5,000-person minimum population geographies
- Counties below 5,000 population
- Combined counties with a 5,000-person minimum population
- 20,000-person minimum population geographies
- Counties below 20,000 population
- Combined counties with a 20,000-person minimum population

These types of geographies can be merged with the other geographies of matching population thresholds to form a layer of full coverage for the jurisdiction. In other words, by merging these together, all census tracts of the jurisdiction should be accounted for in the final layer.

Limitations: Geographies are sorted high to low by their population, then beginning with the first geography are aggregated with their nearest neighbor until the population threshold is met. There are potential alternative methods and tools to create standard, small-area geographies in support of visualizing health and exposure data. The Tracking Program's sub-county geographies are only one example of such methodology.

APPENDIX A

Examples of software-aggregated combined county areas that could be potential candidates for manual division. Geographies outlined in grey are the county boundaries. Geographies outlined in red are the boundaries generated during software aggregation. Green lines represent the division after manual adjustment. Numbers shown on the maps are county populations, with the individual counties shown in blue. Where present, black outlines are state boundaries.

1. **A combined county area of many counties.** This is an example in Minnesota of a 20,000-person minimum population combined county area. This is made up of eight counties with a combined population of 59,184. This was divided into two parts.
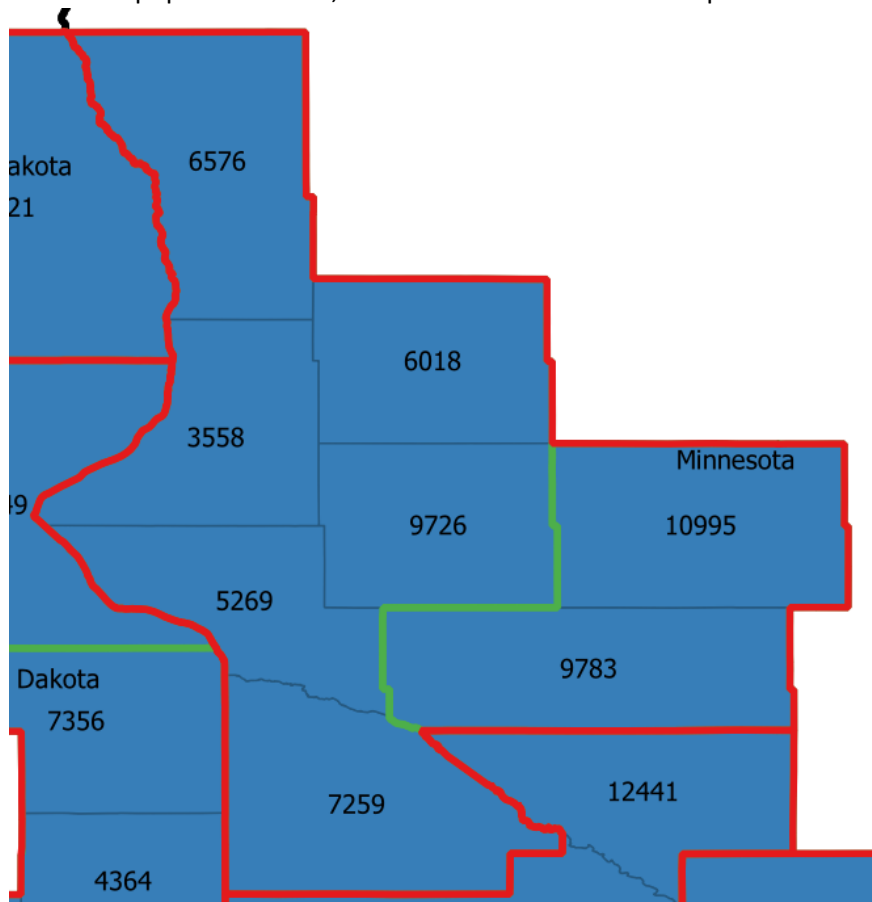
2. **A combined county area in a "circle/donut" shape.** This is an example of an area generated by the GAT tool in South Dakota during the 20,000 minimum population aggregation of counties below 20,000 population. The geography is a continuous circle, with counties far apart from each other grouped together, which may experience different environmental exposures. This area was divided into two new, more compact geographies.
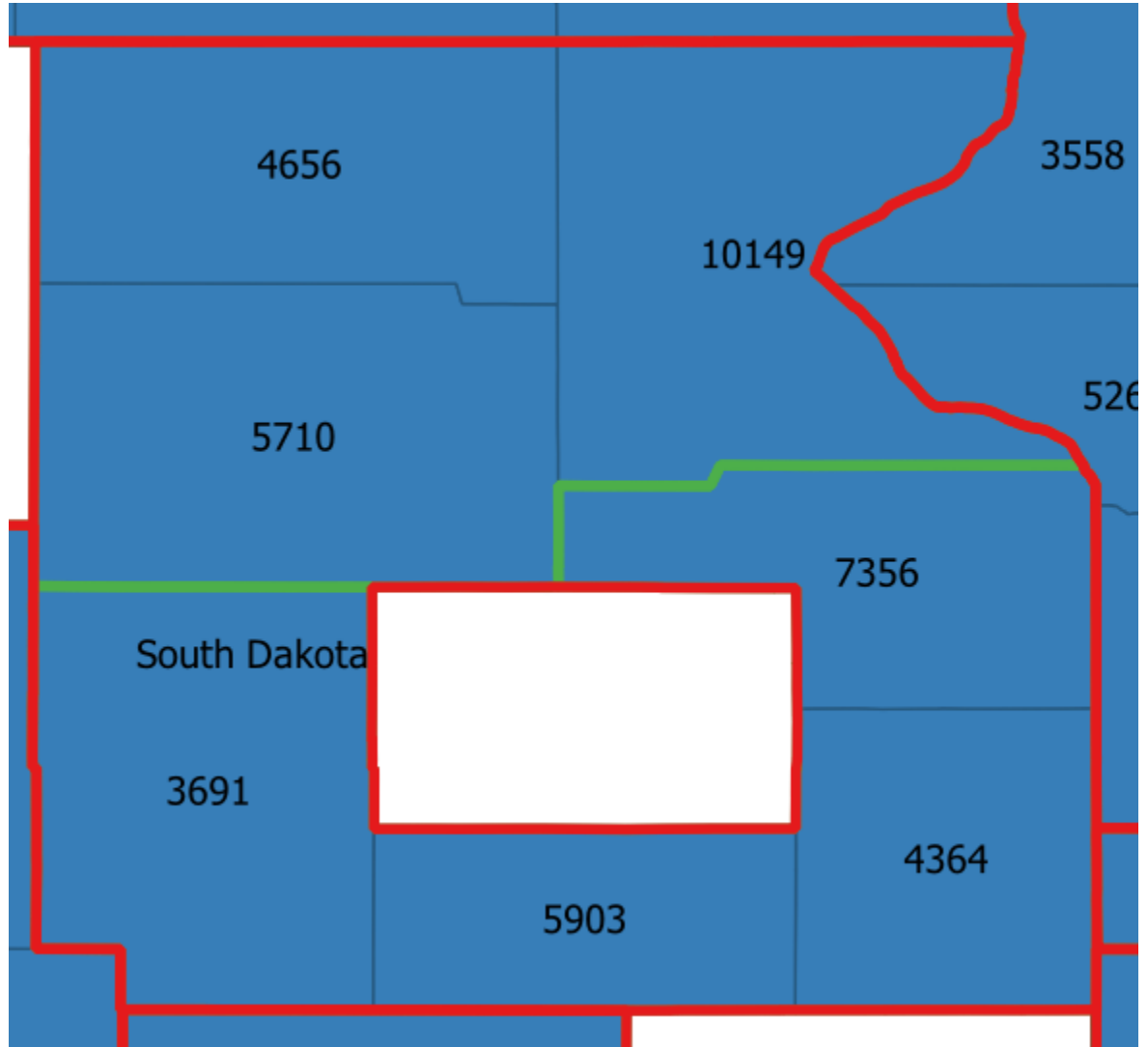
3. **A combined county area in an "L" shape.** This is an example of an area generated by the GAT tool in Iowa during the 20,000 person minimum population aggregation of counties below 20,000 population. This was divided into two new geographies made up of two counties each.
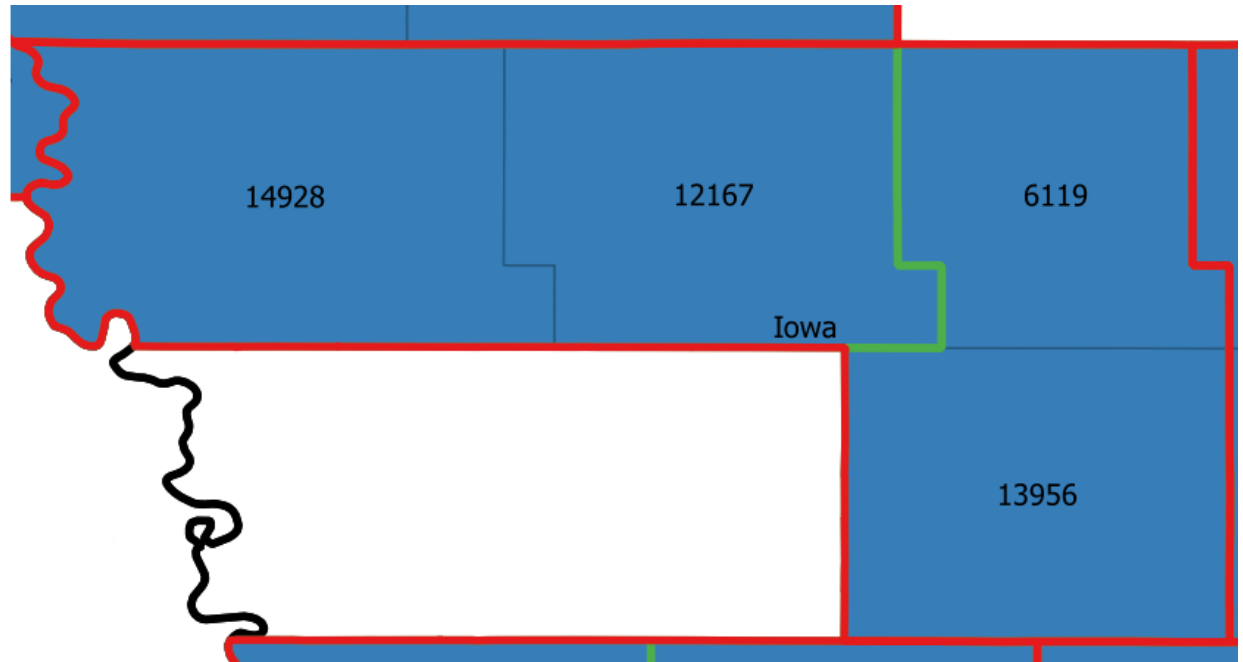
CDC'S ENVIRONMENTAL PUBLIC HEALTH TRACKING PROGRAM

Contact us: trackingsupport@cdc.gov
Visit the Tracking Network today: www.cdc.gov/ephtracking
Follow us on social media:
- Twitter (@CDC_EPHTracking)
- Facebook (facebook.com/CDCEPHTracking)

POWERED BY
TRACKING