

TOOLS AND TECHNIQUES

Choropleth Map Design for Cancer Incidence, Part 1

Thomas B. Richards, MD; Zahava Berkowitz, MSPH; Cheryl C. Thomas, MSPH; Stephanie Lee Foster, MPH; Annette Gardner; Jessica Blythe King, MPH; Karen Ledford, CTR; Janet Royalty, MS

Suggested citation for this article: Richards TB, Berkowitz Z, Thomas CC, Foster SL, Gardner A, King JB, et al. Choropleth map design for cancer incidence, part 1. *Prev Chronic Dis* 2010;7(1):A23. http://www.cdc.gov/pcd/issues/2010/jan/09_0054.htm. Accessed [date].

Abstract

Choropleth maps are commonly used in cancer reports and community discussions about cancer rates. Cancer registries increasingly use geographic information system techniques. The Centers for Disease Control and Prevention's Division of Cancer Prevention and Control convened a Map Work Group to help guide application of geographic information system mapping techniques and to promote choropleth mapping of data from central cancer registries supported by the National Program of Cancer Registries, especially for planning and evaluation of comprehensive cancer control programs. In this 2-part series in this issue of *Preventing Chronic Disease*, we answer frequently asked questions about choropleth map design to display cancer incidence data. We recommend that future initiatives consider more advanced mapping, spatial analysis, and spatial statistics techniques, and include usability testing with representatives of state and local programs and other cancer prevention partners.

Introduction

Maps are an effective tool for cancer control planning and evaluation (1-3). Data displayed on a map allow users to visualize spatial relationships and draw attention to areas of importance. Maps can be used to identify boundaries of

complex geography, display rates for specific areas, reveal geographic patterns, and suggest questions for research (eg, what is the spatial relationship between cancer rates and risk factors such as socioeconomic status?) (1).

The National Program of Cancer Registries (NPCR), Division of Cancer Prevention and Control (DCPC), Centers for Disease Control and Prevention (CDC) supports state central cancer registries (CCR) in the collection of high-quality cancer incidence data (4). An increasing number of these registries assign geocodes (eg, latitude and longitude coordinates) to residential addresses of people with incident cases (5,6). These geocoded cases can be used to develop maps of cancer incidence rates and as part of spatial statistical analyses (7).

Choropleth maps are a common starting point for mapping cancer incidence. DCPC convened a Map Work Group to develop guidance for the design of choropleth maps and to promote mapping of NPCR-supported CCR cancer incidence data. Choropleth maps of cancer incidence rates assign colors to rate categories and then fill the area in the geographic units of interest (eg, states, counties, census tracts) with the color corresponding to that unit's rate (8). The National Cancer Institute (NCI) and CDC state cancer profiles Web site provides good examples of choropleth maps (9). Many more advanced mapping methods exist, but these methods typically require investment in additional software or training for state program staff (7,10,11).

This 2-part series summarizes Map Work Group responses to common questions about choropleth map design. In this article we discuss the purpose of the map, geographic units of analysis, cancer sites, age-adjusted rates, rate

ratios, and reliability. In Part 2 we discuss suppression rules; questions related to mapping cancer stage, rates, and percentages; classes for map display; comparing maps over time; map color schemes, labels, projections, and output media; and limitations in interpretation (12).

Frequently Asked Questions About Choropleth Map Design

1. What is the purpose of the map?

Map design requires consideration of the audiences to which the map will be presented, the purpose that the map serves for each audience, and plans to provide supplemental information to help interpret the map. For example, in the context of comprehensive cancer control, multiple audiences potentially exist, including community members, policy makers, clinicians, geographers, epidemiologists, and state comprehensive cancer control staff. For internal program use by CCR staffers who have signed an agreement to protect privacy and confidentiality of cancer data, maps can be useful to show point locations of cancer cases. However, to protect privacy and confidentiality, this type of map would not be distributed to the public. Similarly, although maps developed with advanced geographic information systems methods (eg, adaptive spatial filtering) can be used to engage community participation (11), such maps may require that the map maker meet with community representatives to explain the methods used and how to interpret the map.

Sharing maps with end users during development ensures that the content, meaning, and audience interpretation are appropriate. More formal usability testing may be helpful, especially when requesting user feedback on Web applications with maps (13). The same map may not be equally suited for all audiences or be able to answer all questions. A single map may lead end users to request additional maps. More than 1 map or different types of maps in addition to tables, graphs, and explanatory text may be needed to answer all of the questions posed by a specific audience.

Maps are especially useful to help users visualize the answers to “where” questions and questions about spatial relationships. Such questions are commonly asked as part of state comprehensive cancer control planning and evaluation (1), for example:

- Where are high-priority populations for cancer prevention interventions?
- Where are cancer screening services provided?
- Where do preventable cancers occur, especially advanced-stage cases?
- Are there gaps between the locations of high-priority populations and locations where cancer prevention services are provided?
- What is the cancer incidence rate for a specific area?
- Where are areas with unusually high or low rates?
- Are the geographic patterns on a map caused by normal random variation?
- How do spatial patterns in cancer incidence rates change over time?
- What is the spatial correlation between geographic patterns for cancer incidence rates and those for cancer risk factors?

2. Are some geographic units of analysis more advantageous than others for choropleth cancer incidence maps?

In 2003, Boscoe and Pickle (14) reviewed 12 geographic units that can be used for choropleth maps of cancer incidence data and identified the following characteristics as desirable:

- high degree of resolution
- homogeneity of population size
- homogeneity of land area
- observation of minimum population thresholds and land area thresholds
- temporal stability and currency
- compactness of shape
- audience familiarity
- data availability
- the functional relevance of the unit to the phenomena mapped

They concluded that 1) each of the 12 geographic units had some advantages and disadvantages; 2) depending on the specific study question, some units may be preferable to others; and 3) none of the units was optimal for all purposes (14). For national maps of the continental United States, they assigned highest ratings to states, counties, and the Health Service Areas used in CDC’s *Atlas of United States Mortality* (14,15).

In addition to the units reviewed by Boscoe and Pickle (14), Hao et al (16) suggest that presentation of cancer

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the US Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors’ affiliated institutions. Use of trade names is for identification only and does not imply endorsement by any of the groups named above.

data using congressional district boundaries may be useful in communicating with legislators and persuading them to enact new cancer control programs and to strengthen existing ones. Because many congressional districts do not follow state or county boundaries, Hao et al (16) describe a method to estimate age-adjusted death rates for congressional districts by using county-level data.

Other investigators have concluded that analyses using geographic units at the subcounty level would be advantageous. For example, Goodman et al (17) define primary care service areas based on US zip codes where Medicare beneficiaries prefer to receive primary care. California has mapped advanced-stage colon cancer cases by using medical service study areas, based on aggregations of census tracts that local communities considered “rational service areas” for primary health care (18). Gregorio et al (19) suggest that, except for investigations focused on a specific cancer cluster in a limited geographic area, spatial analysis at the census tract level might be a sufficient resolution for surveillance of cancer spatial patterns in a single state.

An additional consideration in choice of geographic unit may be the ability to use geography to accurately link cancer incidence data with census demographics, risk factors, and other data. In 2002, Krieger et al (20) concluded that census tract or block group units were better than zip codes for analyses of US socioeconomic inequalities in health.

3. What cancer sites would be good starting points for illustrating how cancer registry data may be used to help answer cancer prevention and control questions?

The Map Work Group recommended breast, colorectal, and cervical cancer as reasonable starting points for the development of cancer incidence maps for comprehensive cancer control. These cancers can be prevented by implementation of the US Preventive Services Task Force (USPSTF) recommendations for community preventive services and clinical interventions (21). The USPSTF recommends screening men and women aged 50 years or older for colorectal cancer; biennial screening mammography for women aged 50 to 74 years; and screening for cervical cancer in women who have been sexually active and have a cervix.

Other cancer sites and types of data also may be of interest. For example:

- For a specific state, any high-priority cancer identified in

the state comprehensive cancer control plan (22).

- For lung cancer, maps and geographic analyses of trends in tobacco use by high school students (23).

4. What types of questions are best addressed by maps showing incident cancer case counts, unadjusted (crude) rates, direct age-adjusted rates, or indirect age-adjusted rates?

Presenting cancer case counts in a table with the geographic unit (eg, county) as the row can be a useful starting point for cancer prevention and control discussions. On the basis of information in the table, a choropleth map of case counts can be developed. However, because case counts are often proportional to population size, decision makers may ask questions that require tables showing the case-to-population ratio or rate by geographic unit and choropleth maps designed on the basis of that information.

Rates for many cancers increase with age, and differences in the population age distribution in different areas can influence the observed crude cancer rates in each area. To control for such differences, direct and indirect methods can be used for age adjustment (sometimes referred to as age standardization) (11,24,25).

The direct age-adjusted rate is calculated by multiplying the age-specific crude rates for the local study population (eg, for a county) by the corresponding age-specific proportion weight for the standard population (eg, for a state) and then summing these products. Direct age-adjusted rates are reported in the NCI/CDC *State Cancer Profiles* and in *United States Cancer Statistics* reports, using the national population as the standard (9,26).

In contrast, the indirect age-adjusted method estimates the expected cases in the local study area (eg, a county) by multiplying the number of people in an age category for the local study area population by the corresponding age-specific rates of the standard population (eg, the state). Expected cases then are summed across age groups and compared with the actual or observed number of cases in the local study population. The ratio of observed to expected cases takes into account age distribution because both the observed and expected cases are based on the age distribution of the local study population.

If the goal is to compare cancer rates in different local study populations, direct age-adjusted rates are needed.

Indirect age adjustment does not allow rates in different local study populations to be compared because indirect age-adjusted rates are not based on a common age distribution. However, the indirect age-adjustment method can be advantageous when the local study population age groups are too small to calculate stable, local age-specific rates, as in sparsely populated rural counties (11). As summarized by Beyer and Rushton, indirect age-adjustment “applies the stable statewide rate to local populations, instead of applying local disease rates, which for small areas are unstable, to standard population weights” (11).

For questions about allocation of resources, tables of case counts and maps of age-specific rates may be more useful than age-adjusted rates because the case counts and age-specific rates are actual measures of risk within the specific area of interest. In contrast, direct age-adjusted rates are relative indexes, and hypothetical rates reflect the age distribution of the selected standard population rather than the actual number of people in each age category in a specific community. A potential limitation of indirect age-adjusted rates for purposes of resource allocation is that each local area applies a different set of weights reflecting the age distribution of its population. On the other hand, as Beyer and Rushton point out, local decision makers may find indirect age-adjusted rates useful because “the difference between actual and expected numbers of late-stage cancer cases is a measure of the need for additional resources such as screening services” (11).

5. When calculating and evaluating county-to-state rate ratios to identify specific counties with higher or lower rates than the state rate, how should the denominator for the rate ratio be defined when the index county of interest has a relatively large population compared with other counties in that state?

In 2008, 16 counties (in 15 states) accounted for 25% or more of the total population in their state. Examples of such counties are Clark County, Nevada (71.8% of the state population); Maricopa County, Arizona (60.8%); Cook County, Illinois (41.0%); Salt Lake County, Utah (37.4%); King County, Washington (28.6%); and Los Angeles County, California (26.8%) (27).

County-to-state rate ratios are calculated as the ratio of the rate for an index county of interest to the state rate. The state rate can potentially be calculated with the index

county excluded (State Rate 1) or with the index county included (State Rate 2).

State Rate 1 (excluding the index county) is advantageous from a statistical perspective because the numerator rate (the index county rate) and denominator rate (the state rate excluding the index county) are independent. The statistical assumption of nonoverlapping groups is not violated.

State Rate 2 (including the index county) allows overlap between the numerator rate (the index county rate) and the denominator rate (the state rate including the index county). However, if county-to-state rate ratios are needed for every county in a state, the State Rate 2 approach is easier to calculate than the State Rate 1 approach. Using the State Rate 2 approach, the rate for each index county is compared with the same denominator (the state rate including the index county). In contrast, using the State Rate 1 (excluding the index county) approach, a different state rate needs to be calculated with the selection of each index county.

The Map Work Group concluded that the following rule of thumb may be helpful in deciding which approach would be appropriate. The state rate can be calculated with the index county included (State Rate 2) if the population of the index county accounts for less than 25% of total state population. However, if the population of the index county accounts for 25% or more of the total state population, then the state rate should exclude the index county (State Rate 1). When an index county accounts for 25% or more of the total state population, inclusion of the index county in the state rate can result in a reported county-to-state rate ratio that is less than the true county-to-state rate ratio by at least 10%.

6. How should information about reliability (eg, unstable rates) be displayed on a map?

Several methods exist to display information about reliability of rates on a map. One option uses different shades of gray to indicate map areas with small numbers, unstable rates, or missing data. If colors are used to indicate areas with stable rates, areas shaded gray tend to remain in the background.

A second option employs hatched lines to convey rate variance. The hatched lines allow the underlying spatial

patterns to be seen. The *Atlas of United States Mortality* (14) illustrates how double-hatching with parallel white and black lines can be used over light and dark colors.

A third method, proposed by Carr and colleagues (28,29), provides confidence intervals in addition to mapped rates. The mapped rates are ranked, confidence intervals are calculated around each rate, and a graph of the ranked rates with their confidence intervals is then displayed adjacent to micromaps of the rates. This approach is used in the Comparative Data Display section of the *State Cancer Profiles* (9).

A fourth option is the use of funnel plots. Funnel plots show increasing population size on the x-axis, and higher and lower bounds for predicted limits for rates on the y-axis around a horizontal line corresponding to the overall rate (30,31). The predicted limits decrease as population size increases, resulting in a graph with a shape similar to that of a funnel. Outliers for geographic units of different population sizes are identifiable as the rates located outside the predicted limits.

Conclusion

Design of high-quality, effective choropleth maps of cancer incidence may appear simple but in fact can involve complex issues.

Acknowledgment

We thank Harland Austin, DSc, professor of epidemiology, Rollins School of Public Health, Emory University, for his help with questions about county-to-state rate ratios.

Author Information

Corresponding Author: Thomas B. Richards, MD, Centers for Disease Control and Prevention, 4770 Buford Hwy NE; Mailstop K-55, Atlanta, GA 30341-3717. Telephone: 770-488-3220. E-mail: TRichards@cdc.gov.

Author Affiliations: Zahava Berkowitz, Cheryll C. Thomas, Stephanie Lee Foster, Annette Gardner, Jessica Blythe King, Karen Ledford, Janet Royalty, Centers for Disease Control and Prevention, Atlanta, Georgia. Stephanie Lee Foster is also affiliated with the Agency for

Toxic Substances and Disease Registry, Atlanta, Georgia.

References

1. Bell BS, Hoskins RE, Pickle LW, Wartenberg D. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *Int J Health Geogr* 2006;5:49.
2. Ghetian CB, Parrott R, Volkman JE, Lengerich EJ. Cancer registry policies in the United States and geographic information systems applications in comprehensive cancer control. *Health Policy* 2008;87(2):185-93.
3. Parrott R, Hopfer S, Ghetian C, Lengerich E. Mapping as a visual health communication tool: promises and dilemmas. *Health Commun* 2007;22(1):13-24.
4. National Program of Cancer Registries (NPCR). Centers for Disease Control and Prevention. <http://www.cdc.gov/cancer/npcr>. Accessed April 1, 2009.
5. Boscoe FP, McLaughlin CC, O'Brien DK. Geographic information systems. In: Menck HR, Deapen D, Phillips JL, Tucker TC, editors. *Central cancer registries: design, management, and use*. 2nd edition. Dubuque (IA): Kendall/Hunt Publishing Company; 2007. p. 169-91.
6. Abe T, Stinchcomb D. Geocoding practices in cancer registries. In: Rushton G, Armstrong MP, Gittler J, Greene BR, Pavlik CE, West MM, et al. *Geocoding health data. The use of geographic codes in cancer prevention and control, research, and practice*. Boca Raton (FL): CRC Press; 2008. p. 111-26.
7. Bhowmick T, Griffin AL, MacEachren AM, Kluhsman BC, Lengerich EJ. Informing geospatial toolset design: understanding the process of cancer data exploration and analysis. *Health Place* 2008;14(3):576-607.
8. Brewer CA. Basic mapping principles for visualizing cancer data using geographic information systems (GIS). *Am J Prev Med* 2006;30(2S):S25-S46.
9. State cancer profiles. National Cancer Institute and Centers for Disease Control and Prevention; 2009. <http://statecancerprofiles.cancer.gov/>. Accessed September 26, 2009.
10. Hopfer S, Chadwick AE, Parrott RL, Ghetian CB, Lengerich EJ. Assessment of training needs and preferences for geographic information systems (GIS) mapping in state comprehensive cancer-control programs. *Health Promot Pract OnlineFirst*, published Apr 1, 2008 as doi:10.1177/1524839907309047 [PMID: 18381971].

11. Beyer KMM, Rushton G. Mapping cancer for community engagement. *Prev Chronic Dis* 2009;6(1). http://www.cdc.gov/pcd/issues/2009/jan/08_0029.htm. Accessed April 1, 2009.
12. Richards TB, Berkowitz Z, Thomas CC, Foster SL, Gardner A, King JB, et al. Choropleth map design for cancer incidence, part 2. *Prev Chronic Dis* 2010;7(1). http://www.cdc.gov/pcd/issues/2010/jan/09_0073.htm. Accessed October 26, 2009.
13. Bhowmick T, Robinson AC, Gruver A, MacEachren AM, Lengerich EJ. Distributed usability evaluation of the Pennsylvania Cancer Atlas. *Int J Health Geogr* 2008;11(7):36.
14. Boscoe FP, Pickle LW. Choosing geographic units for choropleth rate maps, with an emphasis on public health applications. *Cartogr Geogr Inf Sci* 2003;30(3):237-48.
15. Pickle LW, Mungiole M, Jones GK. Atlas of United States mortality. DHHS Pub No (PHS) 97-1015. National Center for Health Statistics, Centers for Disease Control and Prevention; 1996. <http://www.cdc.gov/nchs/products/other/atlas/atlas.htm>. Accessed April 1, 2009.
16. Hao Y, Ward EM, Jemal A, Pickle LW, Thun MJ. US congressional district cancer death rates. *Int J Health Geogr* 2006;5:28.
17. Goodman DC, Mick SS, Bott D, Stukel T, Chang CH, Marth N, et al. Primary care service areas: a new tool for the evaluation of primary care services. *Health Serv Res* 2003;38(1 Pt 1):287-309.
18. Medical service study areas. California Environmental Information Clearinghouse; 2005. <http://gis.ca.gov/catalog/BrowseRecord.epl?id=23784>. Accessed April 1, 2009.
19. Gregorio DI, DeChello LM, Samociuk H, Kulldorff M. Lumping or splitting: seeking the preferred areal unit for health geography studies. *Int J Health Geogr* 2005;4:6.
20. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. *Am J Epidemiol* 2002;156(5):471-82.
21. US Preventive Services Task Force recommendations. Rockville (MD): Agency for Healthcare Research and Quality; 2008. <http://www.ahrq.gov/clinic/prevenix.htm>. Accessed April 1, 2009.
22. Cancer Control PLANET: links to comprehensive cancer control resources for public health professionals. American Cancer Society, the Agency for Health Care Research and Quality, Centers for Disease Control and Prevention, National Cancer Institute, and Substance Abuse and Mental Health Services Administration; 2008. <http://cancercontrolplanet.cancer.gov/>. Accessed April 1, 2009.
23. Healthy Youth! Tobacco use fact sheets. Centers for Disease Control and Prevention. <http://www.cdc.gov/HealthyYouth/tobacco/state-facts.htm>. Accessed April 1, 2009.
24. Pickle LW, White AA. Effects of the choice of age-adjustment method on maps of death rates. *Stat Med* 1995;14(5-7):615-27.
25. Mather FJ, Chen VW, Morgan LH, Correa CN, Shaffer JG, Srivastav SK, et al. Hierarchical modeling and other spatial analyses in prostate cancer incidence data. *Am J Prev Med* 2006;30(2 Suppl):S88-100.
26. US Cancer Statistics Working Group. United States cancer statistics: 1999–2005 incidence and mortality Web-based report. Atlanta (GA): US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2009. <http://apps.cdc.gov/uscs>. Accessed September 26, 2009.
27. Population Division, US Census Bureau. County population, population change, and estimated components of resident population change: April 1, 2000 to July 1, 2008. <http://www.census.gov/popest/counties/files/CO-EST2008-alldata.pdf>. Accessed September 26, 2009.
28. Carr DB. Designing linked micromap plots for states with many counties. *Stat Med* 2001;20(9-10):1331-9.
29. Carr DB, Wallin JF, Carr DA. Two new templates for epidemiology applications: linked micromap plots and conditioned choropleth maps. *Stat Med* 2000;19(17-18):1331-9.
30. Davies E, Mak V, Ferguson J, Conaty S, Møller H. Using funnel plots to explore variation in cancer mortality across primary care trusts in South-East England. *J Public Health (Oxf)* 2008;30(3):305-12.
31. Eastern Region Public Health Observatory, National Health Service Tools — calculating public health measures. Funnel plots. Cambridge (GB): Institute of Public Health; 2008. <http://www.erpho.org.uk/topics/tools/>. Accessed April 1, 2009.

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the US Department of Health and Human Services, the Public Health Service, the Centers for Disease Control and Prevention, or the authors' affiliated institutions. Use of trade names is for identification only and does not imply endorsement by any of the groups named above.