

Calculation of Health Disparity Indices

By Chakib Battioui

Abstract

Socioeconomic indices are strongly believed to be associated with the risk of disease. The objective of this project is to determine which area-based socioeconomic measures would be most appropriate for US public health surveillance. The specific objective is to determine the relationship between the rate of cancer and the various indices of social and economic (SES) conditions for the study area consisting of Kentucky, Tennessee, North Carolina, Virginia and West Virginia.

Introduction

Socioeconomic indices are strongly believed to be associated with the risk of disease. There is no consensus in the US regarding which area-based measure should be used to measure or monitor socio-economic inequalities in health [1] The objective of this project is to determine which area-based socioeconomic measures would be most appropriate for US public health surveillance. The specific objective is to determine the relationship between the rate of cancer of white female and the various indices of social and economic (SES) conditions for the study area consisting of Kentucky, Tennessee, North Carolina, Virginia and West Virginia.

We used in this project SAS software (SAS Institute; Cary, NC), ArcGis and the SAS bridge for ESRI. SAS is a powerful statistical program with many components. SAS Enterprise Miner is designed for data miners, marketing analysts, database marketers, risk analysts, fraud investigators, engineers and scientists who need to play a major role in solving critical business or research issues. The SAS Bridge for ESRI gives the ability to exchange spatial data between ArcGIS, and SAS. It links spatial, numeric and textual data through a single interface to improve efficiency, produce more intelligent results, and communicate those results more effectively.

Background

Cancer of the cervix begins in the lining of the cervix. The cervix is the lower part of the uterus (womb). The cervix connects the body of the uterus to the vagina (birth canal). The part of the cervix closest to the body of the uterus is called the endocervix. The part next to the vagina is the ectocervix. Most cervical cancers start where these two parts meet [2].

The Socio Economic Index is used to classify people listed in public health databases by the characteristics of their residential neighborhoods. It permits the calculation of population-based rates stratified by location. Many of the obstacles to the use of area-based measures are both technical and conceptual. To date, there exists no consensus in the United States regarding which area-based measures should be used, and at which level of geography to measure or monitor socioeconomic inequalities in health. By contrast, there are several well established, theoretically conceived, and validated area-based deprivation measures in the United Kingdom, such as the Townsend index that permit meaningful comparisons and the monitoring of national, regional, and local socioeconomic gradients in health over time [1].

In this study, we used two websites as our information sources: www.cancer.gov and www.census.gov. We downloaded cervical cancer data for white females from 1970-1994 from the first website. The second site was a good source for geographical data about the five states used in this project: KY, NC, TN, VA and WV. We downloaded files concerning Educational attainment (P037), Occupation (P050), Income (P052), Poverty (P087), Poverty Ratio (P088), Tenure by Person by Room (H022), Tenure by Vehicles (H044), and Value of Housing (H084) for each of the five states. Based on these files we created ten categorical variables (Table 1).

Table 1. Categorical Variables

number	categorical variables
1	Low income: Percentage of households with an income <\$15,000 (LOWINCLEVEL)
2	High income: percentage of population with an income >\$150,000 (HIGHINCLEVEL)
3	Low Education: Percentage of persons aged >=25 with less than High school education (LOWEDLEVEL)
4	High Education: Percentage of persons aged >=25 with at least 4 years of college (HIGHEDLEVEL)
5	High occupation: percentage of persons employed in predominantly working class (HIGHOCCLEV)
6	Low occupation (LOWOCCLEVEL)
7	Poverty: below federally defined line: for example income below \$12,647 for a family of four (POVERTYLEVEL)
8	High value housing: Homes worth >=\$300,000 (HIGHVALLEVEL)
9	Crowding: household with more than one person per room (CROWDLEVEL)
10	No car: percentage of no car ownership (NOCARLEVEL)

Method

SAS software (SAS Institute; Cary, NC) was used to assign levels for the ten categorical variables. Each of the variables was assigned five levels ranked from very bad to very good based on how the level of the socioeconomic factors for the population of each of the five states. A map was created for each variable in order to investigate any relationship with the cancer data. Then, three indices were defined and compared based on the accuracy of predicting the cancer rate in white females.

Method1: Add all 10 categorical variables listed in Table 1 to create a new index level. This new index was used to predict the rate of cancer in white females across the states of study. The general linear model, (The SAS procedure, GLM) was used for this task.

Method2: Factor analysis was used (PROC FACTOR in SAS). The resulting analysis created two factors. The first factor was identified as economic resources, while the second one was identified as education and employment.

Method3: The interaction effect method was used on two groups identified as Low social Class variables and High social class variables. The Low social Class variables are poverty, low income, low education, low occupation and having no car, while the high social class variable are high education, high income, high occupation and high housing value. SAS code for this method is given below

```
proc glm data=SASUSER.a1a1a1a1 ;
  class HIGHVALLEV LOWEDLEVEL HIGHEDLEVE LOWOCCLEVE
  HIGHOCCLEV LOWINCLEVE
  HIGHINCLEV POVERTYLEV CROWDLEVEL NOCARLEVEL;
  model RATEWFlevel_num=
  HIGHVALlev*HIGHEDleve*HIGHOCClev*HIGHINClev
  LOWOCCleve*LOWINCleve*POVERTYlev*CROWDlevel*NOCARlevel*LOWEDI
  evel /solution;
  output out=sasuser.method3data p=_pred;
run;
data sasuser.allmethod;
set sasuser.method3data;
predvalue3=round(_pred);
run;
```

Results

The maps in Figure 1a and Figure 2a are examples of the ten maps created by the data. The first map shows the distribution of cervical cancer rates in the five state area compared to those with a high education, while the second one shows the distribution of cervical cancer rate of white female versus poverty. The first map shows that the eastern part of Kentucky and the northern part of Tennessee have very high rates of cancer and very low rates of higher education. In

contrast, Virginia has a very high rate of education. The second map also shows that the east of Kentucky and the north of Tennessee have very high rates of poverty, which is not the case in the east of Virginia. Figure 1b and Figure 2b show the histogram distribution of cervical cancer rate of white female versus high education and poverty respectively.

Figure 1a. Distribution of cervical cancer rate of white female versus a high education

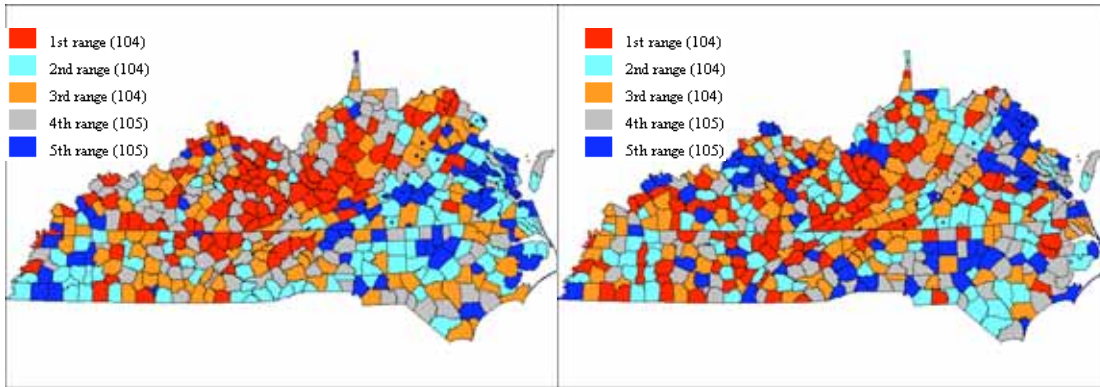


Figure 1b. Distribution of cervical cancer rate of white female versus a high education

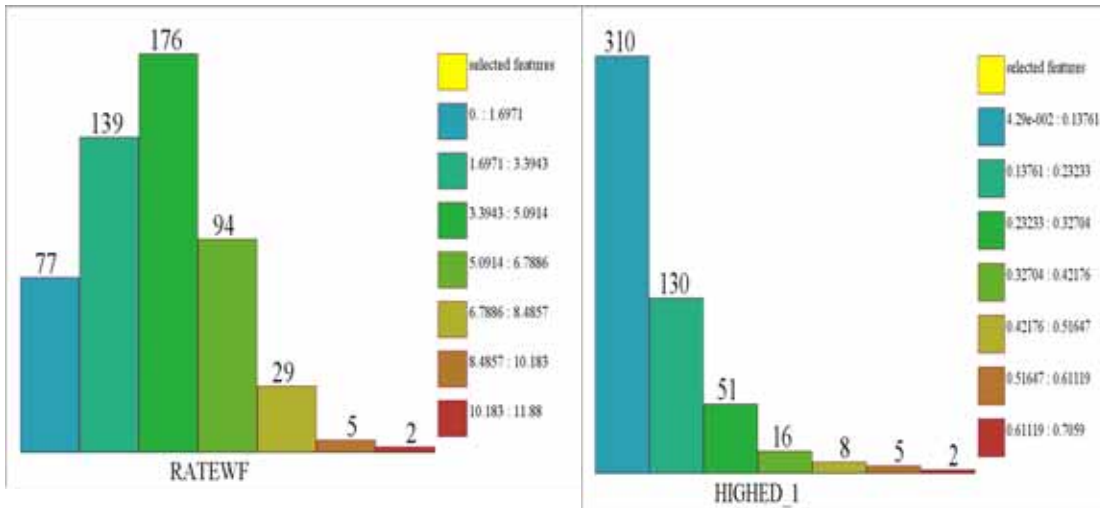


Figure 2a Distribution of cervical cancer rate of white female versus poverty

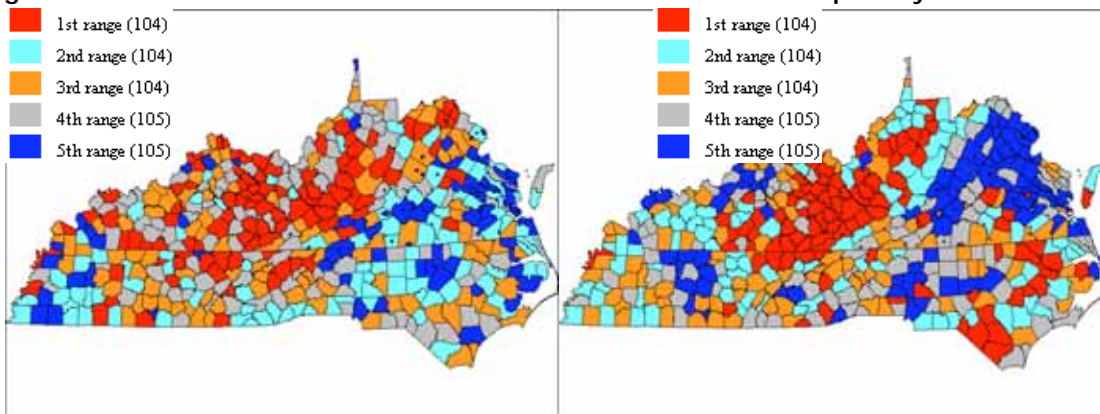
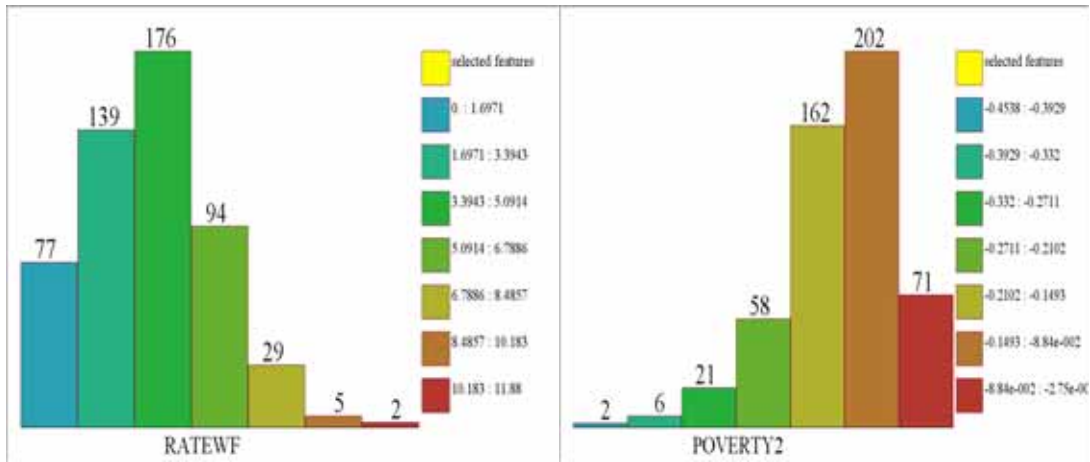


Figure 2b Distribution of cervical cancer rate of white female versus poverty



The first two methods used produced indices that don't provide a satisfactory prediction. We included model 3 that gives a good prediction. Results are given in table 2

Table 2: Prediction table for the interaction effect method.

Frequency Col Pct	1	2	3	4	5	Total
1	104 100.00	0 0.00	1 0.91	0 0.00	0 0.00	105
2	0 0.00	102 99.03	2 1.82	0 0.00	0 0.00	104
3	0 0.00	1 0.97	102 92.73	2 1.94	0 0.00	105
4	0 0.00	0 0.00	5 4.55	97 94.17	2 1.96	104
5	0 0.00	0 0.00	0 0.00	4 3.55	100 98.04	104
Total	104	103	110	103	102	522

Table 2 shows the predicted cervical cancer classes. For example 102 counties from class 2 were correctly predicted with a rate of 99.03%, while only 2 counties were misclassified.

Discussion:

Using the interaction effect method we can get a good prediction of the rate of cancer of white females. We can determine some kind of relationship between the rate of cancer of white females in the five states of study and the various indices of social and economic conditions. We would like to investigate other methods in the future study.

On the other hand, the integration of ArcGis and SAS was very helpful in this study. The SAS Bridge for ESRI gives the ability to exchange spatial data between ArcGIS, and SAS. It links spatial, numeric and textual data through a single interface to improve efficiency, produce more intelligent results, and communicate those results more effectively.

References

1. Nancy Krieger, Jarvis T. Chen, Pamela D. Waterman, Mah-Jabeen Soobader, S.V. Subramanian, and Rosa Carson. "Geocoding of US socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-based Measure and Geographic Level Matter?". *American Journal of Epidemiology*. May 7, 2002.

2. American Cancer Society

http://www.cancer.org/docroot/CRI/content/CRI_2_2_1X_What_is_cancer_of_the_cervix_8.asp?sitearea=

Chakib Battioui
Calculation of health disparity indices
University Of Louisville
Natural sciences building
Louisville, KY 40292
Tel: (502) 852-6240
C0batt01@louisville.edu