

April 2015

Measuring community health outcomes: New approaches for public health services research

Public Health agencies are increasingly asked ‘to do more with less.’ Tough economic times have increased competition for financial resources and LHDs are competing for local, state, and federal dollars. At the same time, there is increasing pressure to justify public health investments and the effectiveness of public health programs. New measures and methods are needed to help the field of public health describe morbidity and mortality at the community level and demonstrate effectiveness. Toward this end, a research team at the University of North Carolina at Chapel Hill, with assistance from practice partners, has been working to help understand the association between investments in local public health and community health outcomes in North Carolina (NC). Previous briefs have described the variation of public health spending, staffing and services, and associations between spending, staffing, services and community mortality rates. This report describes the utility of a novel data source to describe community morbidity and mapping approaches to identify areas of highest relative risk for potential interventions.

Study Methods

This retrospective study was conducted to examine the utility of multi-payer insurance claims data combined with geospatial analysis for assessing community morbidity and its association with NC local health department (LHDs) spending. Specifically, we were interested in the effects of spending related to the 2008 economic recession. We examined morbidity outcomes for a number of conditions, using hospitalization data and outpatient data, separately and combined. In this brief, we describe data and methods

used for two morbidity outcomes: sexually transmitted diseases and hospitalizations for heart disease. Results of these analyses, using 2008 data are provided to illustrate our processes.

We demonstrate the utility of using health insurance claims data from multiple payers to quantify community disease rates and the ability of a mapping approach that incorporates model adjusted rates with spatial clustering to identify areas of high relative risk at the population-level.

Morbidity data for sexually transmitted diseases (STDs) and hospitalizations for heart disease were obtained using ICD-9 CM Diagnosis Codes from the Integrated Cancer Information and Surveillance System (ICISS). The ICISS program links the NC Central Cancer Registry data to Medicare, Medicaid, private insurance, and other state-level health and demographic datasets. This novel, linked data resource includes 5.5 million unique individuals representing the diversity of the state population. To assure we captured claims for unique individuals, we examined beneficiary data by three mutually exclusive groups based on age and payer: age 65 and older in the 100% Medicare sample, those younger than age 65 in the 100% Medicaid sample, and those younger than 65

represented in the private payer data. Persons younger than age 65 covered not included in the data because they were covered by other payers were assumed to have similar disease rates as persons included in the database who were covered by private insurance. A synthetic population estimate was created to account for uninsured populations by linking and using county level demographic measures together with individual level health care information from the Census Bureau's Small Area Health Insurance Estimates.ⁱ County level hospitalization rates using this approach were validated against the State Inpatient Data from NC Department of Public Health (DPH).

LHD spending and services data were obtained from the National Association of County and City Health Officials (NACCHO) profile survey data from 2008. Spending was analyzed using a measure of per capita expenditure constructed from the total reported LHD expenditures for a respective county, or service region population. Services performed by LHDs were reported in NACCHO data and grouped into six categories: clinical preventive services, medical treatment, specialty care services, population based services, regulatory and licensing services, and environmental services. Within each category of service, we assessed the proportion of specific services in the category that were provided or contracted for, by the LHD.

County level covariates: We included the following county level variables from the Area Resource File (ARF): number of public clinics per 10,000 population, percent of population female, unemployment rate, percent of population non-white, percent of population age 65+, total population, percent college graduates, percent of population who were non-English speakers, number of physicians per 100,000 population, number of hospital beds per 100,000 population, percent of population living in poverty, percent of population uninsured, and urban/rural indicator.

Analyses: Morbidity data were linked to LHDs using Federal Information Processing Standard (FIPS) codes for each county to create an analytic dataset that contained one record per LHD. Six of the 85 LHDs had jurisdictional areas which crossed multiple counties. We summed the data across counties for each of these LHDs to create an LHD-level measure. Following the customary disease rate mapping approach, we mapped age-adjusted morbidity rates per 1000 population for each LHD using ArcGIS to describe the geographic variation across the state. While these maps give the appearance of clusters, these results often change with different cutoffs and can therefore be misleading.

Mapping rates in this way cannot always tell us if the

occurrence of disease is randomly distributed or significantly clustered (Cromely; Goodman).

However, other geospatial approaches are available which can overcome this limitation, such as using a Local Indicators of Spatial Association (LISA) statistic that measures the association of a value in a particular area (i.e., LHD in this case) and the values in the nearby areas by applying weights to directly address Tobler's Law or the fact that observations which are closer together (i.e., neighbors) tend to be more related or correlated (Anselin). A positive LISA statistic identifies areas with similar values--- high values surrounded by high values (High-High clusters or hot spots) or low values surrounded by low values (Low-Low cluster or cold spots). A negative LISA statistic identifies areas with dissimilar values---low values surrounded by high values (Low-High clusters) or high values surrounded by high values (High-Low clusters). To determine if the LISA statistic is statistically significant, Monte Carlo method is used to simulate the distribution under the null hypothesis that there is no spatial autocorrelation.

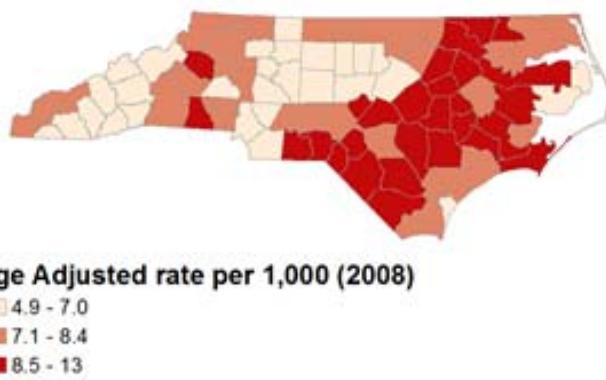
In this study, we then demonstrate mapping approaches using the LISA statistic while adjusting for different factors. Fully adjusted models included: LHD spending, LHD staffing, number of public clinics per 10,000 population, percent of female, unemployment rate, percent of non-white, percent of population age 65+, total population, percent college graduated, percent of non-English speakers, physicians per 100,000 population, hospital beds per 100,000 population, percent of population living in poverty, percent of uninsured, and urban/rural indicator. We first computed a LISA statistic using age-adjusted morbidity rates to see if there are clusters which exceed expectation. Next, we applied LISA to identify significant clusters using model predicted rates, which accounts for the inter-relationships of all the contextual and neighborhood effects in the model such as SES, demographics, and health care access. By mapping the model predicted values, we can also identify how well the covariates explain variation across the state. Using this approach of model-adjusted rates and LISA maps we identify three specific types of areas:

1. Not significant (Areas that have disease rates that are not significantly different from those in surrounding areas, use $p=0.05$ as the cutpoint for significance)
2. High-High (Areas with high disease rates surrounded by areas with high disease rates)
3. Low-Low (Areas with low disease rates surrounded by areas with low disease rates)

Results

Figure 1 shows the map of age-adjusted rate of heart disease hospitalizations in tertiles (three equal size groups). The morbidity rate shows variability across the state--- counties in the highest group have rate values ranging from 4.4 to 9.1 per 1000 population and those in the lowest group have values less than 2 per 1000 population. Areas of NC with high heart disease morbidity, similar to that seen with maps of the “stroke belt” was readily apparent (Musa, Richards, Schieb). The cluster analysis of the age-adjust rates (**Figure 2**) identified two High-High clusters where the LHDs that had high morbidity rates surrounded by neighboring LHDs with high morbidity rates (the cluster map only shows the center areas of the clusters). These two clusters are also represented in **Figure 1**. However the fully adjusted cluster analysis (**Figure 3**) showed that most of the clusters identified in figure 2 are not statistically significant. Only a few and smaller clusters are identified--- one High-High cluster in northeast of NC and two smaller Low-Low clusters in north-central NC.

Figure 1: Rate of Hospitalization for Heart Disease per 1000 in 2008



Much of the variation disappeared in the cluster analysis once the contextual variables such as insurance, poverty, age, race and SES were taken into account. In fact there were several significant predictors in the adjusted analysis which also appeared to be spatially distributed and significant in the LISA analysis. Total public health spending, percent of the population who were college educated, number of public health clinics, and urban status were all significantly associated with lower morbidity; while percent of the population over 65 or non-white were associated with higher morbidity.

Figure 2: Age-Adjusted Cluster Analysis, Heart Disease Hospitalization, 2008

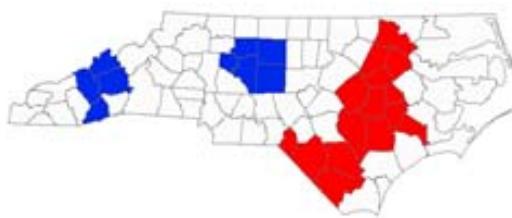


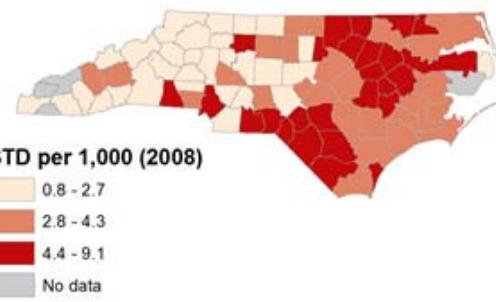
Figure 3: Fully adjusted Cluster Analysis, Heart Disease Hospitalization, 2008



Legend:
■ Significant High Rate surrounded by areas with High Rate
■ Significant Low Rate surrounded by areas with Low Rate

The analysis of Sexually Transmitted Diseases demonstrated a similar pattern, but fewer contextual variables from the model explained the geospatial pattern of disease. The age-adjusted rate of Sexually Transmitted Diseases in NC varied from 0.8 per 1000 population to 9.1 per 1000 population (**Figure 4**). The lowest rates were observed in the western and central region of the state while the highest rates were located in the southern and north-east, or north-central areas. The age-adjusted cluster analysis (**Figures 5 and 6**) confirms two High-High clusters in the southern and north-east areas of the state.

Figure 4: Sexually transmitted disease rates in North Carolina, by county, 2008



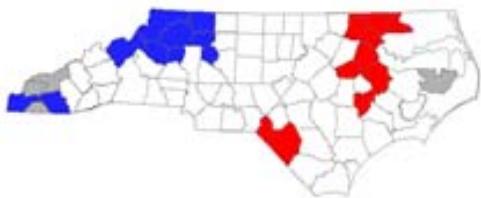
When the analysis is adjusted for other contextual variables, these clusters still exist, suggesting that there are some contextual factors that may be driving the high STD rates that were not represented in the modeling. In the regression model, only the percent of

the population that was nonwhite was significantly associated with STD morbidity.

Figure 5: Age-adjusted Cluster Analysis, Sexually transmitted disease rates Rate, 2008



Figure 6: Fully-adjusted Cluster Analysis, Sexually transmitted disease rates Rate, 2008



Discussion

Our findings demonstrate that aggregated insurance claims data provide a useful tool for assessing the population burden of disease. This may be especially important for diseases which are not consistently or systematically reported to public health agencies. Secondary data collected for other purposes such as administrative (claims) data or Electronic Medical Record (EMR) data may become increasingly useful and available to public health systems and services research (PHSSR) investigators as advances and investments in data infrastructure and standardization continue.

We can identify clusters and significant variation by combining model adjusted rates with geospatial modeling. Applying cluster analysis to contextually adjusted rates helps us understand burden of disease *relative* to what else is happening in the broader environment. This approach can also help visualize the relative impact of covariate adjustment on our outcome measure.

In our example of heart disease, much of the variation in the rates was explained by the contextual variables such as public health spending, SES, age, urban/rural status and access to health care. However there were still several areas which had significant excess morbidity even when compared to their neighbors who also had high rates of disease and a similar contextual environment. Considering STDs, the contextual variables in the model explained much less of the variation observed. This may be a reflection

of a more random spatial distribution of these contextual variables relative to the outcome, unmeasured confounding, or poor covariate measurement. However, by appropriately accounting for the distribution of the disease across the state, we identified areas of excess morbidity, as well as areas which appear to have lower rates relative to their neighbors.

This methodology can also be applied to contrast behavioral and environmental risk factors identify areas of significantly high or low risk. As we demonstrated, simply mapping age-adjusted incidence rates may obscure important underlying correlations, or areas of high relative burden of disease. A better understanding of the broader environmental context can help inform and tailor interventions for a specific area.

To understand how this approach might be helpful, we offer the following illustration. An LHD wanting to inform development of a cancer screening intervention might use conventional public health mapping to identify areas of high cancer rates and low screening rates as their target area. Using the mapping approach we demonstrated in this project, LHDs would be better able to match the most effective type of intervention for a community based on the influence of contextual factors. For example, if the map identifies a cluster that has high disease rates and low screening rates compared to other surrounding areas, and all surrounding areas have insurance, then an educational intervention for cancer screening may be most efficacious because cost of services is less likely to be an issue. However, if the map identifies a cluster that has high risk and low screening rates compared to other surrounding areas, and individuals in these areas lack insurance to pay for screening or have poor access, then an educational intervention is unlikely to succeed. Using mapping to targeting areas of low screening or high risk of disease without assessing the underlying mediating factors of insurance, poverty, or access to care, might imply inappropriate or ineffective interventions.

For the field of PHSSR, geospatial information studies may help us better identify and study areas of 'low' relative burden of disease compared to neighbors while also accounting for all the other environmental influences. In other words, we may be able to identify 'where things are working', or not working. In a time of decreasing resources, these tools can help us better target interventions and more objectively demonstrate effectiveness and return on investments.

Authors & Acknowledgements

Anna P. Schenck, PhD, MSPH

Principal Investigator

Director, North Carolina Institute for Public Health and Professor of the Practice, Public Health Leadership Program, UNC Gillings School of Global Public Health

Anne Marie Meyer, PhD

Co-Principal Investigator

Research Assistant Professor, Epidemiology, UNC Gillings School of Global Public Health

Dorothy Cilenti, DrPH, MPH, MSW

Co-Investigator

Senior Investigator, North Carolina Institute for Public Health, UNC Gillings School of Global Public Health

Tzy-Mey Kuo, PhD

Lead Analyst

UNC Lineberger Comprehensive Cancer Center

Carol Gunther-Mohr, MA

Project Manager

North Carolina Institute for Public Health, UNC Gillings School of Global Public Health

The authors wish to acknowledge the contributions to this research from the North Carolina Public Health Association Academic/Practice-Based Research Section.

The authors also express appreciation to the National Association of County and City Health Officials for the use of survey data of local health departments.

Support for this research was provided by a grant from the Robert Wood Johnson Foundation.

Work on this study was supported by the Integrated Cancer Information and Surveillance System (ICISS), UNC Lineberger Comprehensive Cancer Center with funding provided by the University Cancer Research Fund (UCRF) via the State of North Carolina.

Footnotes

ⁱ <https://www.census.gov/did/www/sahie/>

References

Anselin, L. "Under the hood issues in the specification and interpretation of spatial regression models." *Agricultural economics* 27(3): 247-267.

Cromley, E. and McLafferty, S. GIS and Public Health. 2nd Edition. The Guilford Press, February 2012.

Goodman, D. C. and J. E. Wennberg. "Maps and health: the challenges of interpretation." *J Public Health Manag Pract* 1999:5(4): xiii-xvii.

Musa, G., Chiang, P., Sylk, T., Bavley, R., Keating, W., Lakew, B., Tsou, H. and Hoven, C. "Use of GIS Mapping as a Public Health Tool—From Cholera to Cancer." *Health Services Insights* 2013:6 111–116.

Richards, T., Croner, C., Rushton, G., Brown, C. and Fowler, L. "Geographic Information Systems and Public Health: Mapping the Future." *Public Health Rep.* 1999 Jul-Aug; 114(4): 359–373.

Schieb, L. J., Mobley, L. R., George, M. and Casper, M. "Tracking stroke hospitalization clusters over time and associations with county-level socioeconomic and healthcare characteristics." *Stroke* 44(1): 146-152.