Measuring Subcounty Differences in Population Health Using Hospital and Census-Derived Data Sets: The Missouri ZIP Health Rankings Project

Elna Nagasako, MD, PhD, MPH, **Brian Waterman, MPH**, **Mathew Reidhead, MA**, **Min Lian, MD, PhD**, and **Sarah Gehlert, PhD, MA, MSW**

Division of General Medical Sciences, Washington University School of Medicine in St. Louis, St. Louis, Missouri (Drs Nagasako and Lian); Center for Clinical Excellence, BJC HealthCare, St. Louis, Missouri (Dr Nagasako); Missouri Hospital Association, Jefferson City, Missouri (Messrs Waterman and Reidhead); George Warren Brown School of Social Work, Washington University in St. Louis, St. Louis, Missouri (Dr Gehlert); and Alvin J. Siteman Cancer Center, St. Louis, Missouri (Drs Lian and Gehlert)

Abstract

Context—Measures of population health at the subcounty level are needed to identify areas for focused interventions and to support local health improvement activities.

Objective—To extend the County Health Rankings population health measurement model to the ZIP code level using widely available hospital and census-derived data sources.

Design—Retrospective administrative data study.

Setting—Missouri.

Population—Missouri FY 2012–2014 hospital inpatient, outpatient, and emergency department discharge encounters (N = 36 176 377) and 2015 Nielsen data.

Main Outcome Measures—ZIP code–level health factors and health outcomes indices.

Results—Statistically significant measures of association were observed between the ZIP code–level population health indices and published County Health Rankings indices. Variation within counties was observed in both urban and rural areas. Substantial variation of the derived measures was observed at the ZIP code level with 20 (17.4%) Missouri counties having ZIP codes in both the top and bottom quintiles of health factors and health outcomes. Thirty of the 46 (65.2%) counties in the top 2 county quintiles had ZIP codes in the bottom 2 quintiles.

Correspondence: Elna Nagasako, MD, PhD, MPH, Division of General Medical Sciences, Washington University School of Medicine in St. Louis, Campus Box 8005, 4523 Clayton Ave, St. Louis, MO 63110 (nagasakoe@wustl.edu).

Conclusions—This proof-of-concept analysis suggests that readily available hospital and census-derived data can be used to create measures of population health at the subcounty level. These widely available data sources could be used to identify areas of potential need within counties, engage community stakeholders, and target interventions.

Keywords

health factors; health outcomes; health rankings; population health; public health surveillance; small-area health estimates; subcounty-level health estimates

Effectively engaging communities to address the social, economic, environmental, clinical, and behavioral factors that affect health is critical for improving population health outcomes. Assessments have identified subcounty neighborhood-scale measures of health and related factors as a data need.[1–4] Multiple metrics, indices, and rankings have been developed for assessing community health, including community and neighborhood indicators, well-being indices, deprivation indices, and health indicators and rankings.[5–9] However, few of these are widely available at the subcounty level.

Geographic variation in health outcomes and related risk factors at the small-area level (eg, substate and subcounty) has received increasing attention.[10–15] Local-level primary data collection and aggregation are important for obtaining small-area data.[1,6,16,17] Alternatively, extension of county-level population measures to the subcounty level using small-area estimation techniques can address this need.[15,18–23] However, these approaches can be costly, resource intensive, or impeded by sampling constraints. This exploratory proof of concept study was funded by County Health Rankings (CHR) to (1) design a method to extend the CHR population health measurement framework to the ZIP code level using widely available data; (2) examine the agreement between published CHR indices and ZIP-derived measures reapportioned to the county level; and (3) quantify the subcounty variation at the ZIP code level in Missouri.

Materials and Methods

Data sources

The primary study aim was to evaluate whether hospital and census-based data sets could be used within the CHR framework to create health factors and health outcomes indices at the ZIP code level in Missouri. Candidate model input variables were gathered from Missouri Hospital Association, Hospital Industry Data Institute FY 2012-FY 2014 (October 1, 2011 to September 30, 2014) hospital inpatient, outpatient, and emergency department discharge databases for Missouri residents (N = 36 176 377) and linked by reported residential ZIP code to Nielsen sociodemographic data that use spatially defined census block group-to-ZIP code correspondence.[24] Administrative hospital discharge data are commonly used in public health applications such as disease surveillance programs,[25–33] feature standardized record layouts, and are widely available.[34] Candidate variables for the CHR socioeconomic domain were gathered primarily from the 2015 Nielsen Pop-Facts Premier database,[24] which provides intercensal estimates based on block group-level American Community Survey data. The socioeconomic health factor domain was augmented with a socioeconomic

deprivation index developed by Schootman, Lian, and colleagues.[35–37] Supplemental Digital Content Table 1, available at http://links.lww.com/JPHMP/A317, displays all candidate and retained model variables, external validation variables, and data sources by CHR domain and subdomain.

Employment of the aggregate data utilized as model inputs was governed by Hospital Industry Data Institute master data use agreements. Academic personnel participation was reviewed by the Washington University School of Medicine Human Research Protection Office.

Measures

Conceptual framework—Variables were extracted on the basis of correspondence to the CHR conceptual framework,[38] which includes a health factors domain with 4 subdomains (health behaviors, clinical care, social and economic factors, and physical environment) and a health outcomes domain with 2 subdomains (length of life and quality of life). *International Classification of Diseases, Ninth Revision Clinical Modification* codes for variables drawn from hospital data were identified through literature review, keyword search within diagnosis code descriptions, and expert input. A detailed description of variables evaluated for each CHR domain is included in the Supplemental Digital Content Table 1, available at http://links.lww.com/JPHMP/A317.

Data analysis

Evaluation of candidate model inputs—Descriptive statistics and pairwise correlations were conducted to initially evaluate standardized candidate model inputs from hospital and Nielsen data sources against county-level CHR indices. Candidate measures with highly skewed distributions and nonpositive pairwise correlations were further evaluated and transformed or eliminated. Pairwise correlations with external validation measures and CHR county-level subdomain scores were used to further reduce candidate measures sets; only measures with statistically significant pairwise correlations of 0.20 or greater with assigned CHR subdomain scores were retained. Injury-related mortality was retained despite a low correlation to ensure conceptual domain coverage in principal components analyses. Table 1 contains a list of the candidate variables that were retained for inclusion in our final model.

Model creation and evaluation at the county level—Principal components analysis was applied to subdomain input sets to derive subdomain analog factor scores.[39,40] Each yielded only 1 principal component with an eigenvalue greater than 1, indicating that a component was sufficient to explain the common variation in subdomain indicator groups. Linear regression was then used to model CHR Health Factors and Health Outcomes scores as a function of derived subdomain analog factor scores. Consistent with the CHR conceptual framework, the Health Factors model included derived factor scores for Behavior, Environment, Clinical Access, and Socioeconomic Status subdomains as independent variables. The Health Outcomes model included scores from the Quality of Life and Mortality subdomains supplemented with predicted scores from the regression model for Health Factors, for improved model fit, and predictive accuracy. Predicted values from

the 2 regression models served as derived county-level analog scores for CHR Health Factors and Health Outcomes. Pairwise Pearson correlations and scatterplots were used to assess strength of association between analog scores and CHR scores.

Model creation and evaluation at the ZIP code level—Principal components analysis was applied to each ZIP-level input data set to derive ZIP-level CHR subdomain analog scores, using an identical approach to that used for the county-level analysis. Each ZIP-level subdomain model yielded a single component score with an eigenvalue greater than 1. Summarization of ZIP-level estimates into county-level estimates allowed comparison of ZIP-derived results with the CHR results. This spatial interpolation was facilitated using a ZIP ("source" zone) to county ("target" zone) weighting file derived by allocating ZCTA (the census surrogate of ZIP codes based on whole census blocks) population totals to counties based on ZCTA-county proportional allocations using the MABLE/Geocorr version 12 engine.[41] The process was applied to all Missouri ZIP codes and provided an empirical basis for handling codes that overlap counties.

ZIP-level Health Factors and Health Outcomes domain scores were computed using regression weights derived from county-level analyses. General linear mixed modeling was used to model derived ZIP-level CHR Health Factors and Health Outcomes scores as a function of random county-level intercepts. ZIP-to-County mapping weights produced intraclass correlation estimates of the variation in ZIP-level scores explained by within-county clustering. These weights produced Best Linear Unbiased Predictors of county-level Health Factors and Health Outcomes scores.

Pairwise Pearson correlation coefficients and scatterplots were used to assess the strength of correlation between ZIP-derived analog scores reapportioned to the county-level and corresponding CHR scores. Cross-classification tables, weighted κ statistics, and agreement plots were used to assess pairwise agreement between quintile-ranked CHR scores and corresponding derived analogs. Mapped displays of CHR and ZIP-derived quintile-ranked results were produced for visual comparison.

Assessment of intracounty variation at the ZIP code level—The proportion of variation at subcounty level was assessed to determine within-county variation using model-based intraclass correlations. Concordance between the ZIP code and county-level quintile ranking in health factors and health outcomes was also assessed.

Advisory group

This project was informed by an advisory group comprising local public health, philanthropic, hospital association, hospital community benefit, academic, and community advocate organization members. Quarterly meetings were held to obtain ongoing immediate feedback on this project's aims, methods, and results to ensure that the study's approach would address the differing needs of multiple stakeholders.

Results

Evaluation of candidate model inputs

We found that model inputs drawn from hospital and census-derived data were significantly correlated with CHR indices. Table 1 contains summary statistics and correlation with CHR subdomains for all candidate variables retained as final model inputs. All pairwise correlations were statistically significant at $P < .05$, with the exception of injury-related mortality, which was retained to help ensure Environment subdomain coverage.

Model creation and evaluation at the county level

Principal components analysis produced models for health factors and health outcomes using hospital and commercial data with moderate to substantial correlations with CHR domain and subdomain scores, yielding 1 significant factor for each subdomain. Variable inputs, principal component eigenvalues, and pairwise correlations with CHR subdomain scores for 114 Missouri counties are displayed in Supplemental Digital Content Table 2, available at http://links.lww.com/JPHMP/A317. One Missouri county lacked sufficient data and was excluded. Correlations between CHR Health Outcomes and Health Factors overall domain scores and county-level indices calculated from derived subdomain analog scores using standard CHR weights are displayed in Table 2. All pairwise correlations were statistically significant ($P < .05$).

Quintile agreement between CHR Health Factors and the derived analogs for 114 Missouri counties was 54%, with 92% of counties landing within 1 quintile (weighted $\kappa = 0.66$, $P < .05$). For Health Outcomes, 43% of counties fell within the same quintile and 89% of ZIP-derived analogs fell within 1 quintile difference of the CHR results for 2015 (weighted $\kappa = 0.54$, $P < .05$). Similar agreement was observed when evaluated as octiles and deciles. Scatterplots and correlations between CHR Health Outcomes and Health Factors overall domain scores and county-level indices calculated from derived subdomain analog scores using standard CHR weights are displayed in Supplemental Digital Content Figures 1 and 2, available at http://links.lww.com/JPHMP/A317.

Model creation and evaluation at the ZIP code level

Concordance between ZIP code–level health factors and health outcomes scores for Missouri derived from hospital and commercial data sets and the original CHR county scores was evaluated by interpolating the ZIP code results to the county level. We found moderate to substantial, statistically significant agreement between the published CHR indices and the derived indices. Correlations between CHR Health Outcomes and Health Factors overall domain scores and ZIP-level analogs calculated from derived subdomain analog scores using standard CHR weights are displayed in Table 2. All pairwise correlations were statistically significant ($P < .05$).

Quintile agreement between CHR Health Factors and the ZIP-derived analogs for Missouri counties was 52% with 94% of counties landing within 1 quintile (weighted $\kappa = 0.66$, $P < .05$) (see Supplemental Digital Content Figure 3, available at http://links.lww.com/JPHMP/A317). For Health Outcomes, 49% of counties fell within the same quintile and 84% of ZIP-

derived analogs fell within 1 quintile of the 2015 CHR results (weighted $\kappa = 0.56$, $P < .05$) (see Supplemental Digital Content Figure 4, available at http://links.lww.com/JPHMP/ A317). Mapped representations of ZIP-level ranking quintiles for Missouri based on derived Health Factors and Health Outcomes indices versus 2015 CHR indices are displayed in Figure 1 (color version available in Supplemental Digital Content Figure 5, available at http://links.lww.com/JPHMP/A317).

Evaluation of subcounty variation in health factors and outcomes

ZIP code–level health factors and outcomes indices showed substantial variation at the ZIP code level. Mixed models yielded intraclass correlation estimates associated with county-level random intercepts of 0.44 and 0.5 for Health Factors and Health Outcomes, respectively, indicating that 50% to 56% of the variance in derived domain scores is observed at the ZIP code level. Substantial variation was observed at the ZIP code level, with 20 (17.4%) counties having ZIP codes in both the top and bottom quintiles of health factors and health outcomes. Thirty of the 46 (65.2%) counties in the top 2 quintiles had ZIP codes in the bottom 2 quintiles. This was observed in both urban and rural areas (see Figures 2 and 3; color versions available in Supplemental Digital Content Figures 6 and 7, available at http://links.lww.com/JPHMP/A317).

Discussion

Using the conceptual framework of the County Health Rankings & Roadmaps,[38] we identified candidate measures available at the ZIP code level in hospital discharge and commercial census-derived data sets for each CHR health factors and health outcomes domain and subdomain. We derived ZIP code–level health factors and health outcomes scores for Missouri using these measures. We evaluated concordance with the original CHR indices at the county level by apportioning the ZIP code results to the county level. Finally, we assessed the extent of subcounty variation in health factors and outcomes indicated by these indices. The results of this exploratory study show statistically significant agreement between published CHR indices and ZIP code–level indices derived using hospital and census-derived data that are widely available at subcounty levels. Although the degree of agreement was limited, this finding suggests that despite the limitations of hospital data in capturing population health, these data can be combined with other sources to assess variation in health at the subcounty level. These findings serve as a starting point for further work to develop data sources for use by community stakeholders working to improve health in communities. In addition, mixed-models results and graphical displays illustrate how these health factors and health outcomes indices vary at subcounty levels (see Figures 1–3; color versions available in Supplemental Digital Content Figures 5–7, available at http:// links.lww.com/JPHMP/A317). This underscores the potential for subcounty data to identify small-area variation and target scarce community health improvement resources.

It is important to note that the results of this exploratory study are not presented to suggest the model and measures we derived are replications of or alternatives to CHR constructs. Rather, they are offered as "proof-of-concept" evidence that selected data sources can be used in the context of the CHR framework to derive alternative measures that correspond

sufficiently with established county-level rankings to support their plausible use as a basis for assessing subcounty variation demonstrably associated with CHR domains. Although hospital data will not completely capture population health, local decision making around public health of necessity often involves hospitals and hospital data. Given the origin of our approach in the needs of hospital-public health partnerships, the objective of our exploratory study was to apply these data in a way that was appropriate and realistic for local stakeholder groups. We assert that the pattern and magnitude of presented correlations between CHR and our derived analog measures support further exploration of the practical use of these measures as a basis for identifying subcounty areas of higher and lower relative community health need.

Geographic variation in health factors and outcomes at the small-area level, including the substate, subcounty, and census tract levels, has been noted in many contexts.[10–12] Subcounty health data and their drivers can be used by local public health departments to inform policy decisions and engage community partners. Yet, data are often unavailable at the ZIP code or census tract levels, and small area data have been identified as a need at both the local and federal levels.[1–4,6] Primary data collection using surveys at the subcounty level and the application of small-area estimation techniques to existing survey results are an important source of subcounty data.[1,16,17] However, data collection can be costly and resource intensive[1] and small-area estimation may be impeded by the lack of rural data, the potential discordance between geographic areas and the area boundaries of interest, and the area-level context effect on outcomes independent of individual characteristics.[17]

Hospital administrative data have been used for public health surveillance and the detection of geographic variation in factors and outcomes.[11,12,25–33] An assessment of chronic disease prevalence using emergency department administrative data found rates comparable with those from survey data and that significant neighborhood variation in diabetes burden was identifiable using this method.[12] Given the importance of sociodemographic determinants,[9] we added Nielsen data to hospital administrative data. Our resulting ZIP code–level indices for the CHR domains and subdomains in Missouri agreed with our predefined gold standard of the published CHR county-level measures and identified significant subcounty-level variation. This supports the continued exploration of hospital administrative data as a subcounty-level source and identifies a method for combining hospital and commercial data sets to produce factors and rankings scalable to other states. Geographically based health rankings and indicators have been used to draw media attention to public health issues, including health disparities, and to engage communities in partnerships to improve health,[7,8,42,43] suggesting that ZIP code–level measures can be used for engagement of communities to address identified subcounty data needs.[1,4] In addition, there is a need for the alignment of hospital community benefits spending with community needs.[44–47] Use of data sets readily available to hospitals to produce public health data can help support these alignment efforts via the engagement of hospital stakeholders.

Limitations

Our approach has limitations. First, because a goal was to use data readily accessible in other states, we did not use data sources arguably more directly related to the CHR domains,

yet not widely available across states. For example, the restriction to hospital and census-derived data limited the measures available in the environmental health factors subdomain, which has been identified as a challenge for the CHR rankings themselves.[48] Future iterations of these models will investigate the incorporation of additional variables in the environmental domain such as particulate matter, land use, and land cover data. Second, the use of hospital administrative data in public health settings has limitations.[27,30,32,33,49,50] The population pool for hospital data sets may not be as representative of the general population as population-based surveys. Although rural data availability is a hospital data strength, low population sometimes meant lack of ZIP code–level data. Also, in both rural and nonrural settings, patients with financial, spatial, and access barriers may seek hospital care at increased or decreased rates, yielding selection bias. The majority of hospital emergency department Medicaid and Medicare patient visits are considered nonurgent—with a diagnosis of acute upper respiratory infection most common. This may present autocorrelation between our measures of hospital utilization and socioeconomic factors. Third, although address is collected as a routine part of billing, patient ZIP code was sometimes unavailable. Some missing ZIP code data may be attributable to errors in data collection; the data generation process also excludes patients who are homeless or decline to provide information on residence. However, this occurred in a relatively small proportion of our discharge records (0.08%) and is unlikely to play a major role in our analysis. In addition, Nielsen data are estimated with American Community Survey block group-level estimates and subject to similar sampling bias. While Nielsen uses enhanced aggregation and distribution techniques to minimize this bias, the potential for artificial spreading of actual variance may arise as a result. Nielsen data are readily available to hospitals as they are widely used for strategic planning purposes, and the use of Nielsen databases can reduce time spent on data extraction; however, the cost of these data could be prohibitive to health departments or other organizations that do not already have access to these data. We anticipate that the substitution of Nielsen data with publicly available American Community Survey data would yield similar results; this is a topic for further investigation. Also, only a single data set encompassing 1 state was used in the study. Finally, our approach is subject to the limitations inherent in producing health rankings for community engagement, including the proliferation of measures encountered by local communities and the need to link the data to community action by ensuring its meaningfulness.[2,5,6,51]

Conclusion

We demonstrated that hospital and census-derived data can be used to extend a commonly used framework for county-level population health to subcounty areas. Although hospital data reflect only a subset of the health of a community, this study suggests that its use in combination with other sources warrants further exploration as a subcounty data source. Our future work will include making these results publicly available using an interactive platform, evaluating their use in meeting data needs of Missouri stakeholders, and evaluating performance of this method on successive years of data, in other states, and with the inclusion of additional data sources for domains not readily captured in hospital and census-derived data sets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Castrucci BC, Rhoades EK, Leider JP, Hearne S. What gets measured gets done: an assessment of local data uses and needs in large urban health departments. J Public Health Manag Pract. 2015; 21(1 suppl):S38–S48. [PubMed: 25423055]

2. National Academies of Sciences, Engineering, and Medicine. Metrics That Matter for Population Health Action: Workshop Summary. Washington, DC: The National Academies Press; 2016.

3. National Committee on Vital and Health Statistics. [Accessed October 6, 2016] Letter to the Secretary of the Department of Health & Human Services. Recommendations on supporting community data engagement by increasing alignment and coordination, technical assistance, and data stewardship education. http://www.ncvhs.hhs.gov/wp-ontent/uploads/2013/12/2015-Ltr-to-Secy-CommunityHealthDataEngagement-redacted.pdf. Published May 28, 2015

4. National Committee on Vital and Health Statistics. [Accessed October 6, 2016] NCVHS measurement framework for community health and well-being, V2. http://www.astho.org/IntegrationForum/Successes-Measures/NCVHS-Population-Health-Framework-July-2016/. Published July 2016

5. National Committee on Vital and Health Statistics. [Accessed October 6, 2016] Population Health Subcommittee. Environmental scan of existing domains and indicators to inform development of a new measurement framework for assessing the health and vitality of communities. http://www.ncvhs.hhs.gov/wp-content/uploads/2016/06/NCVHS-Indicators-Envirn-Scan_2016-06-01-FINAL.pdf. Published June 2016

6. Howell EM, Pettit KLS, Ormond BA, Kingsley GT. Using the national neighborhood indicators partnership to improve public health. J Public Health Manag Pract. 2003; 9(3):235–242. [PubMed: 12747321]

7. Erwin PC, Myers CR, Myers GM, Daugherty LM. State responses to America's health rankings: the search for meaning, utility, and value. J Public Health Manag Pract. 2014; 20(5):472–480. [PubMed: 24253406]

8. Rohan AMK, Booske BC, Remington PL. Using the Wisconsin County Health Rankings to catalyze community health improvement. J Public Health Manag Pract. 2009; 15(1):24–32. [PubMed: 19077591]

9. Kreiger N, Waterman PD, Spasojevic J, Li W, Maduro G, Van Wye G. Public health monitoring of privilege and deprivation with the index of concentration at the extremes. Am J Public Health. 2016; 106:256–263. [PubMed: 26691119]

10. Anderson L, Martin NR, Flynn RT, Knight S. The importance of substate surveillance in detection of geographic oral health inequalities in a small state. J Public Health Manag Pract. 2012; 18(5): 461–468. [PubMed: 22836538]

11. Knudson A, Casey M, Burlew M, Davidson G. Disparities in pediatric asthma hospitalizations. J Public Health Manag Pract. 2009; 15(3):232–237. [PubMed: 19363403]

12. Lee DC, Long JA, Wall SP, et al. Determining chronic disease prevalence in local populations using emergency department surveillance. Am J Public Health. 2015; 105:e67–e74. [PubMed: 26180983]

13. Riva M, Gauvin G, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. J Epidemiol Community Health. 2007; 61:853–861. [PubMed: 17873220]

14. Jia H, Muennig P, Borawski E. Comparison of small-area analysis techniques for estimating county-level outcomes. Am J Prev Med. 2004; 26(5):453–460. [PubMed: 15165663]

15. Zhang X, Onufrak S, Holt J, Croft J. A multilevel approach to estimating small area childhood obesity prevalence at the census block-group level. Prev Chronic Dis. 2013; 10:120252.

16. Land GH. Measuring 2010 national objectives and leading indicators at the state and local level. J Public Health Manag Pract. 2002; 8(4):9–13. [PubMed: 15156632]

17. Wang Y, Ponce NA, Wang P, Opsomer JD, Yu H. Generating health estimates by zip code: a semiparametric small area estimation approach using the California health interview survey. Am J Public Health. 2015; 105:2534–2540. [PubMed: 26544642]

18. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health. 2001; 55(2):111–122. [PubMed: 11154250]

19. Gupta RS, Zhang X, Sharp LK, Shannon JJ, Weiss KB. Geographic variability in childhood asthma prevalence in Chicago. J Allergy Clin Immunol. 2008; 121(3):639–645.e1. [PubMed: 18243285]

20. Li W, Kelsey JL, Zhang Z, et al. Small-area estimation and prioritizing communities for obesity control in Massachusetts. Am J Public Health. 2009; 99(3):511–519. [PubMed: 19150906]

21. Holt JB, Zhang X, Presley-Cantrell L, Croft JB. Geographic disparities in chronic obstructive pulmonary disease (COPD) hospitalization among Medicare beneficiaries in the United States. Int J Chron Obstruct Pulmon Dis. 2011; 6:321–328. [PubMed: 21697996]

22. Eberth JM, Zhang X, Hossain M, Tiro JA, Holt JB, Vernon SW. County-level estimates of human papillomavirus vaccine coverage among young adult women in Texas, 2008. Texas Public Health J. 2013; 65(1):37–40.

23. Zhang X, Holt JB, Lu H, et al. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. Am J Epidemiol. 2014; 179(8):1025–1033. [PubMed: 24598867]

24. The Nielsen Company. [Accessed November 11, 2016] Consumer activation. Nielsen. https://segmentationsolutions.nielsen.com/consumeractivation/. Copyright 2016

25. Elliott AF, Davidson A, Lum F, et al. Use of electronic health records and administrative data for public health surveillance of eye health and vision-related conditions in the United States. Am J Ophthalmol. 2012; 154(6 suppl):S63–S70. [PubMed: 23158225]

26. Centers for Disease Control and Prevention. Analytical challenges for emerging public health surveillance. MMWR Suppl. 2012; 61(3):35–40. [PubMed: 22832996]

27. Huff L, Bogdan G, Burke K, et al. Using hospital discharge data for disease surveillance. Public Health Rep. 1996; 111(1):78–81. [PubMed: 8610197]

28. Nsubuga, P., White, ME., Thacker, SB., et al. Public health surveillance: a tool for targeting and monitoring interventions. In: Jamison, DT.Breman, JG.Measham, AR., et al., editors. Disease Control Priorities in Developing Countries. 2. Washington, DC: The International Bank for Reconstruction and Development/The World Bank; 2006. p. 997-1015.

29. Feldman KA, Trent R, Jay MT. Epidemiology of hospitalizations resulting from dog bites in California, 1991–1998. Am J Public Health. 2004; 94(11):1940–1941. [PubMed: 15514233]

30. Salemi JL, Tanner JP, Sampat D, et al. The accuracy of hospital discharge diagnosis codes for major birth defects: evaluation of a state wide registry with passive case ascertainment. J Public Health Manag Pract. 2016; 22(3):E9–E19.

31. Kluger MD, Sofair AN, Heye CJ, Meek JI, Sodhi RK, Hadler JL. Retrospective validation of a surveillance system for unexplained illness and death: New Haven County, Connecticut. Am J Public Health. 2001; 91:1214–1219. [PubMed: 11499106]

32. Love D, Rudolph B, Shah GH. Lessons learned in using hospital discharge data for state and national public health surveillance: implications for Centers for Disease Control and Prevention Tracking Program. J Public Health Manag Pract. 2008; 14(6):533–542. [PubMed: 18849773]

33. Wang Y, Cross PK, Druschel CM. Hospital discharge data: can it serve as the sole source of case ascertainment for population-based birth defects surveillance programs? J Public Health Manag Pract. 2010; 16(3):245–251. [PubMed: 20357611]

34. National Association of Health Data Organizations—Center for Disease Control Cooperative Agreement Project—CDC Assessment Initiative. [Accessed November 11, 2016] Management and institutional controls for reducing disclosure risk in web-based data dissemination of public health data. Guidelines and resources for health data organizations. https://www.nahdo.org/sites/nahdo.org/files/Resources/Other_Resources/Reducing%20Disclosure%20Risk.pdf. Published May 11, 2011

35. Schootman M, Jeffe D, Lian M, Gillanders WE, Aft R. The role of poverty rate and racial distribution in the geographic clustering of breast cancer survival among older women: a geographic and multilevel analysis. Am J Epidemiol. 2009; 169:554–561. [PubMed: 19103608]

36. Lian M, Schootman M, Doubeni CA, et al. Geographic variation in colorectal cancer survival and the role of small-area socioeconomic deprivation: a multilevel survival analysis of the NIH-AARP Diet and Health Study cohort. Am J Epidemiol. 2011; 174(7):828–838. [PubMed: 21836166]

37. Lian M, Perez M, Liu Y, et al. Neighborhood socioeconomic deprivation, tumor subtypes and causes of death after non-metastatic invasive breast cancer diagnosis: a multilevel competing-risk analysis. Breast Cancer Res Treat. 2014; 147(3):661–670. [PubMed: 25234843]

38. County Health Rankings & Roadmaps. [Accessed November 11, 2016] Our approach. County Health Rankings & Roadmaps. http://www.countyhealthrankings.org/our-approach. Copyrighted 2016

39. Roberts S, Martin MA. Using supervised principal components analysis to assess multiple pollutant effects. Environ Health Perspect. 2006; 114(12):1877–1882. [PubMed: 17185279]

40. Krishnan, V. [Accessed October 6, 2016] Constructing an area-based socioeconomic index: a principal components analysis approach. Early Child Development Mapping Project (ECMap), Community-University Partnership (CUP), University of Alberta. http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf. Published May, 2010

41. Missouri Census Data Center. [Accessed November 3, 2016] MABLE/Geocorr 12: Geographic Correspondence Engine. Missouri Census Data Center. http://mcdc.missouri.edu/websas/geocorr12.html. Published 2012. Updated January 27, 2016

42. Peppard PE, Kindig DA, Dranger E, Jovaag A, Remington PL. Ranking community health status to stimulate discussion of local public health issues: the Wisconsin County Health Rankings. Am J Public Health. 2008; 98:209–212. [PubMed: 18172156]

43. Kanarek N, Tsai H-L, Stanley J. Health ranking of the largest US counties using the Community Health Status Indicators peer strata and database. J Public Health Manag Pract. 2011; 17(5):401–405. [PubMed: 21788776]

44. Rabarison KM, Timsina L, Mays GP. Community health assessment and improved public health decision-making: a propensity score matching approach. Am J Public Health. 2015; 105:2526–2533. [PubMed: 26469657]

45. Singh SR, Young GJ, Lee SD, Song PH, Alexander JA. Analysis of hospital community benefit expenditures' alignment with community health needs: evidence from a national investigation of tax-exempt hospitals. Am J Public Health. 2015; 105:914–921. [PubMed: 25790412]

46. Rubin DB, Singh SR, Jacobson PD. Evaluating hospitals' provision of community benefit: an argument for an outcome-based approach to nonprofit hospital tax exemption. Am J Public Health. 2013; 103:612–616. [PubMed: 23409909]

47. Singh SR, Bakken E, Kindig DA, Young GJ. Hospital community benefit in the context of the larger public health system: a state-level analysis of hospital and governmental public health spending across the United States. J Public Health Manag Pract. 2016; 22(2):164–174. [PubMed: 25783004]

48. Hendryx M, Ahern MM, Zulig KJ. Improving the environmental quality component of the County Health Rankings model. Am J Public Health. 2013; 103:727–732. [PubMed: 23409874]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

49. Townes JM, Kohn MA, Southwick KL, et al. Investigation of an electronic emergency department information system as a data source for respiratory syndrome surveillance. J Public Health Manag Pract. 2004; 10(4):299–307. [PubMed: 15235376]

50. Ngo DL, Marshall LM, Howard RN, Woodward JA, Southwick K, Hedberg K. Agreement between self-reported information and medical claims data on diagnosed diabetes in Oregon's Medicaid population. J Public Health Manag Pract. 2003; 9(6):542–544. [PubMed: 14606195]

51. Winterbauer NL, Rafferty AP, Tucker A, Jones K, Tucker-McLaughlin M. Use and perceived impact of the County Health Rankings report in Florida and North Carolina. J Public Health Manag Pract. 2016; 22(6):1–7. [PubMed: 26594933]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Implications for Policy & Practice

- The use of widely available hospital discharge data supplemented with census-derived area-level data could allow potential application of this method of producing subcounty data in other states.

- Estimates of health factors and outcomes at the ZIP code level can target resources to areas most in need and engage community members using more geographically focused data.

- If applied in an ongoing way, this approach could facilitate evaluations of the effectiveness of population health improvement efforts at the subcounty level.
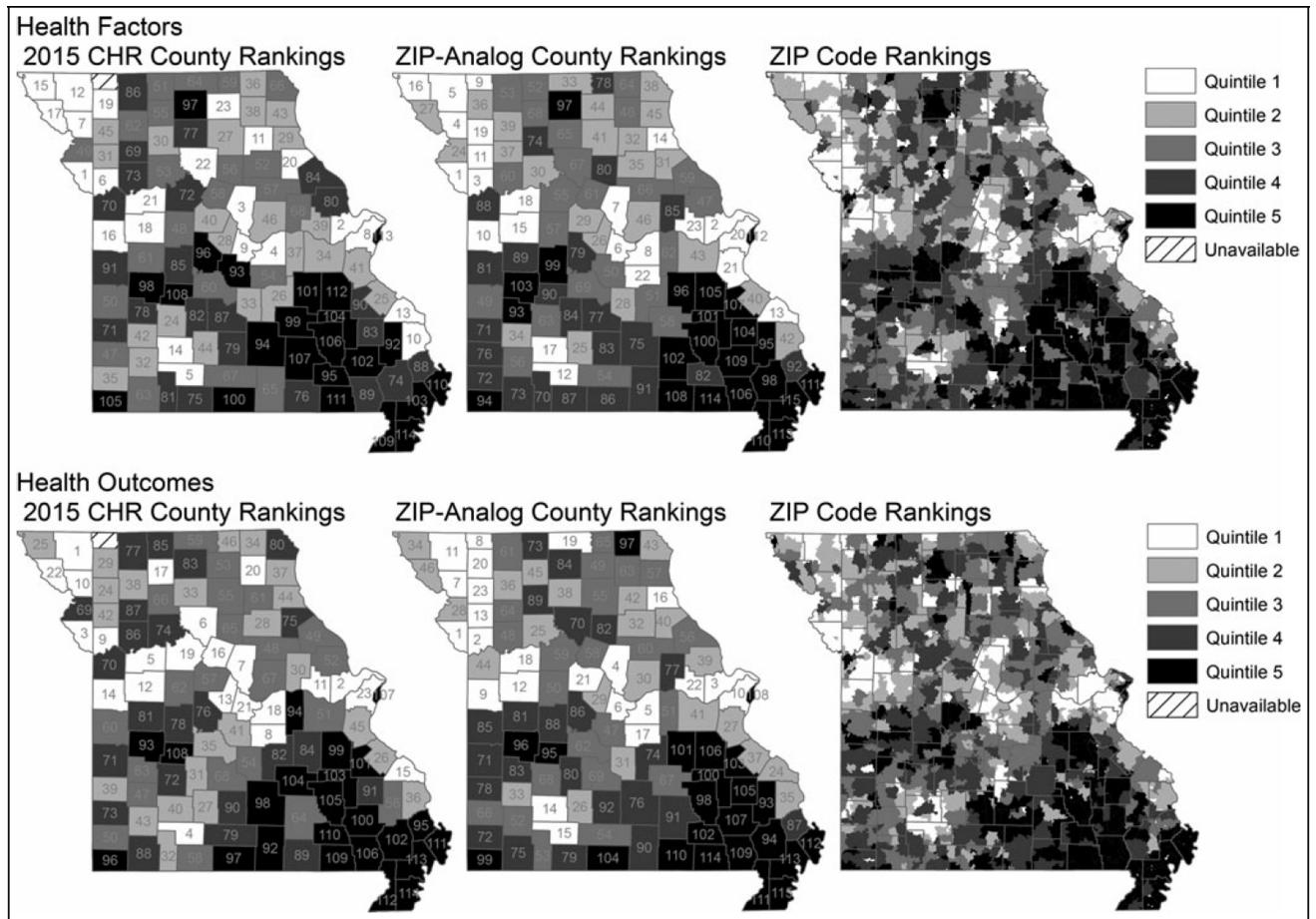
FIGURE 1.
County and ZIP-level rankings derived from hospital and census-derived data sets vs. 2015 county health rankings results
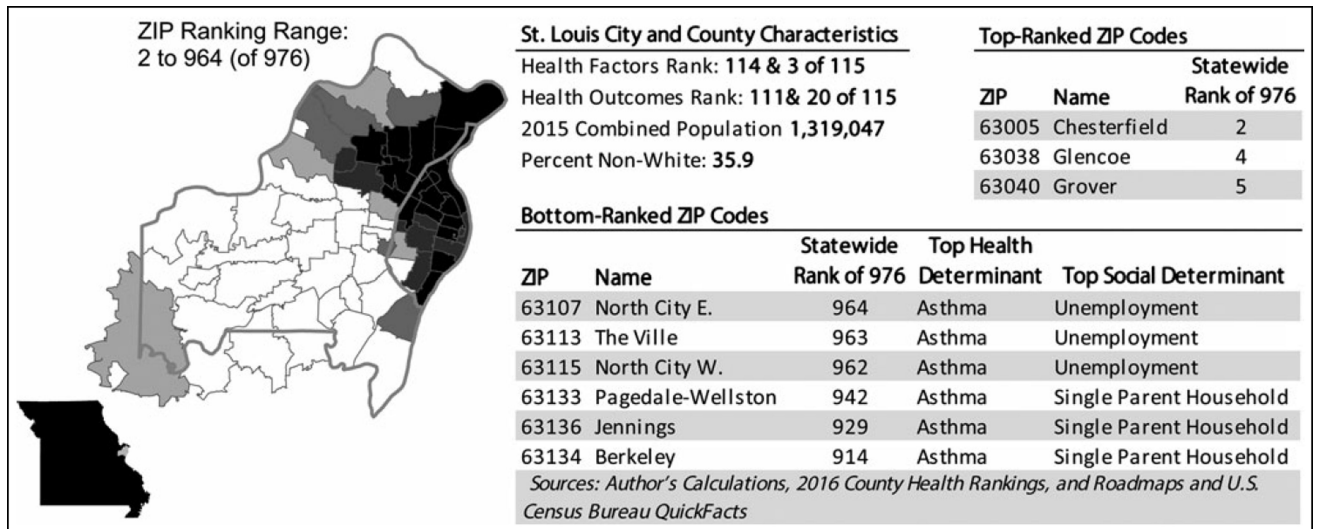
FIGURE 2.
Subcounty variation in health factors and outcomes in urban St. Louis City & County,
Missouri
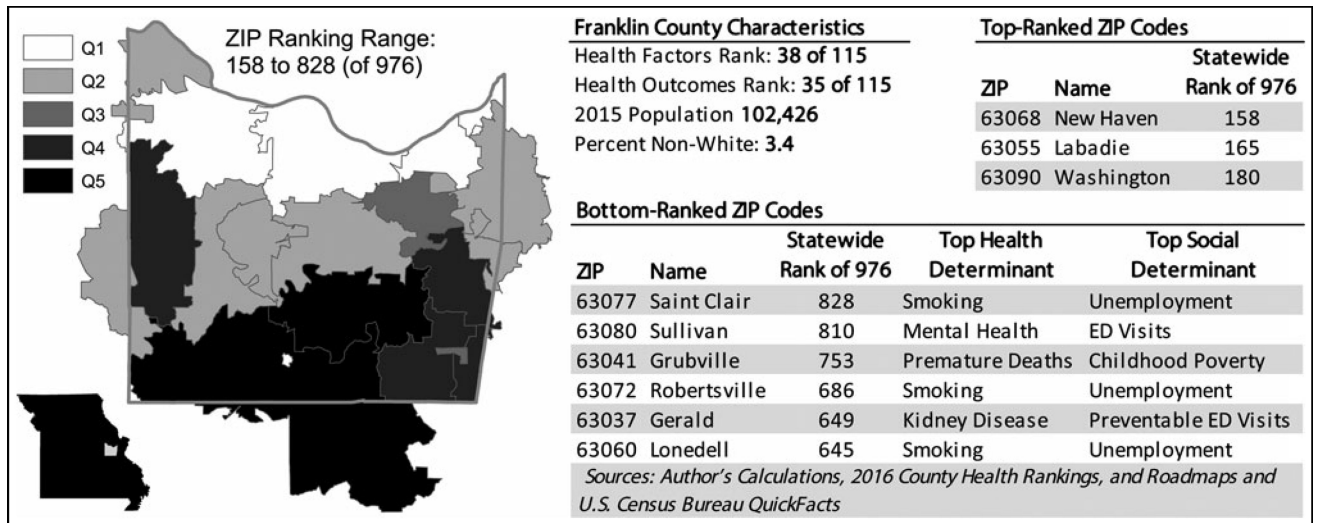
FIGURE 3.
Subcounty variation in health factors and outcomes in rural Franklin County, Missouri

TABLE 1

Summary Statistics and Correlation With CHR Subdomains for Hospital and Census-Derived Data Set Measures Retained as Final Model Inputs[a]

Domain	Subdomain	Measure	Mean (SD)	CHR Correlation
Health outcomes	Mortality	Premature death	5.7 (1.5)	0.66
		Years productive life lost	101.2 (30.2)	0.69
	Quality of Life	ED visits	1371 (356.7)	0.34
		IP visits	416.4 (74.7)	0.42
		Low birth weight	55.1 (16.3)	0.21
		Psychiatric diagnoses	30.6 (11.2)	0.45
Health factors	Behavior	Teen pregnancy	35 (13.1)	0.63
		Sexually transmitted infections	57.7 (24.5)	0.27
	Clinical Access	Off-hours ED visits	0.4 (0.1)	0.33
		Health care worker density	23.7 (7.1)	0.51
		AHRQ PQI total	49.8 (15.6)	0.56
	Environment	Assault diagnoses	12.0 (5.8)	0.39
		Injury-related mortality	2.6 (0.7)	0.12
	Socioeconomic Status	Education less than high school	0.16 (0.1)	0.68
		Unemployment	0.09 (0.03)	0.62
		Childhood poverty rate	0.21 (0.07)	0.71
		Median HH income	42 667 (7995)	0.69
		Socioeconomic deprivation	0 (1.1)	0.72

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; CHR, County Health Rankings; ED, emergency department; HH, household; IP, inpatient; PQI, Prevention Quality Indicators.

[a]One Missouri county was excluded because of insufficient data.

TABLE 2

Correlation Between CHR Domain Scores and Domain Scores Derived From Hospital and Census-Derived Data Sets for 114 Missouri Counties[a]

	CHR Health Factors	CHR Health Outcomes
Model results derived from county-level data		
Health Factors Analog	0.88	0.79
Health Outcomes Analog	0.85	0.83
Model results derived from ZIP code–level data[b]		
Health Factors Analog	0.87	0.82
Health Outcomes Analog	0.84	0.83

Abbreviation: CHR, County Health Rankings.

[a] One Missouri county was excluded due to insufficient data.

[b] ZIP code–level model results proportionally allocated and summarized to county level.