

Generating Subcounty Health Data Products: Methods and Recommendations From a Multistate Pilot Initiative

Trang Q. Nguyen, MD, DrPH; Isaac H. Michaels, MPH; Dulce Bustamante-Zamora, PhD; Brian Waterman, PhD, MPH; Elna Nagasako, MD, PhD, MPH; Yunshu Li, MS; Marjory L. Givens, PhD, MSPH; Keith Gennuso, PhD

ABSTRACT

Background: County Health Rankings & Roadmaps (CHR&R) makes data on health determinants and outcomes available at the county level, but health data at subcounty levels are needed. Three pilot projects in California, Missouri, and New York explored multiple approaches for defining measures and producing data at subcounty geographic and demographic levels based on the CHR&R model. This article summarizes the collective technical and implementation considerations from the projects, challenges inherent in analyzing subcounty health data, and lessons learned to inform future subcounty health data projects.

Methods: The research teams used 12 data sources to produce 40 subcounty measures that replicate or approximate county-level measures from the CHR&R model. Using varying technical methods, the pilot projects followed similar stages: (1) conceptual development of data sources and measures; (2) analysis and presentation of small-area and subpopulation measures for public health, health care, and lay audiences; and (3) positioning the subcounty data initiatives for growth and sustainability. Unique technical considerations, such as degree of data suppression or data stability, arose during the project implementation. A compendium of technical resources, including samples of automated programs for analyzing and reporting subcounty data, was also developed.

Results: The teams summarized the common themes shared by all projects as well as unique technical considerations arising during the project implementation. Furthermore, technical challenges and implementation challenges involved in subcounty data analyses are discussed. Lessons learned and proposed recommendations for prospective analysts of subcounty data are provided on the basis of project experiences, successes, and challenges.

Conclusions: This multistate pilot project offers 3 successful approaches for creating and disseminating subcounty data products to communities. Subcounty data often are more difficult to obtain than county-level data and require additional considerations such as estimate stability, validating accuracy, and protecting individual confidentiality. We encourage future projects to further refine techniques for addressing these critical considerations.

KEY WORDS: community health, disparities, health assessment, small area, subcounty

Author Affiliations: Office of Public Health Practice, New York State Department of Health, Albany, New York (Dr Nguyen, Mr Michaels, and Ms Li); Department of Epidemiology and Biostatistics, School of Public Health, University at Albany, Rensselaer, New York (Dr Nguyen, Mr Michaels, and Ms Li); Office of Health Equity, California Department of Public Health, Sacramento, California (Dr Bustamante-Zamora); Hospital Industry Data Institute, Missouri Hospital Association, Jefferson City, Missouri (Dr Waterman); Division of General Medical Sciences, Department of Medicine, Washington University School of Medicine, St Louis, Missouri (Dr Nagasako); BJC HealthCare Center for Clinical Excellence, St Louis, Missouri (Dr Nagasako); and University of Wisconsin Population Health Institute, University of Wisconsin-Madison, Madison, Wisconsin (Drs Givens and Gennuso).

This research was funded by County Health Rankings & Roadmaps, a collaboration between the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. Additional research support (E.N.) was provided by BJC HealthCare and NIH/NCATS Washington University-ICTS grants UL1 TR000448 and KL2TR000450. The authors thank Mat Reidhead, MA; Ian Brissette, PhD; Anne McCarthy, MPH; Mycroft Sowizral, PhD; Asante Hilts, DrPH, MPH; Neil Maizlish, DrPH; Peter Oh, DrPH; and Madhurima Gadgil, MPH.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH/NCATS.

In the past decade, county-level data for health-related measures have been made broadly available nationwide.¹ However, to further support

The authors declare that they have no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (<http://www.JPHMP.com>).

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Correspondence: Trang Q. Nguyen, MD, DrPH, Office of Public Health Practice, New York State Department of Health, Corning Tower, R 910, Albany, NY 12237 (trang.nguyen@health.ny.gov).

Copyright © 2020 The Authors. Published by Wolters Kluwer Health, Inc.

DOI: 10.1097/PHH.0000000000001167

communities in assessing and effectively reducing local health burden, subcounty data are greatly needed for identifying health disparities within counties that county-level data may not detect.² Subcounty data are also needed for targeting, monitoring, and evaluating public health interventions.³ Granular knowledge about the distribution of health burdens enables public health practitioners to focus limited resources on the most acutely affected geographies and groups to achieve optimal impact.⁴

Multiple terms are used to describe subcounty-level data.^{5,6} In this context, “small-area data” refers to aggregate data for towns, zip code tabulation areas (ZCTAs), census tracts, and other geographies that can be smaller than counties. Likewise, “subpopulation data” refers to aggregate data for age groups, sexes, races, ethnicities, or other groups that, in part, comprise the total populations of counties.

Hospitals and local health departments are increasingly interested in accessing subcounty data.³ The Patient Protection and Affordable Care Act requires private nonprofit hospitals to regularly conduct community health needs assessments that include implementation plans for addressing identified health issues.^{7,8} Similarly, the national Public Health Accreditation Board requires accredited health departments and those pursuing accreditation to regularly conduct community health assessments and develop community health improvement plans.⁹ In past decades, studies have explored small-area data analysis for specific health outcomes and/or small geographical areas.¹⁰⁻¹² More recently, public health analysts have explored data at the subcounty level and some have begun to support these assessments and plans by improving the availability of subcounty data.¹³⁻²⁰ However, these projects either covered a specific geographic area (eg, one city, one county) or included a limited number of data sources and measures.

Since 2010, the County Health Rankings & Roadmaps (CHR&R) have provided overall health measures for nearly every county nationwide.²¹ To address the need for subcounty data compatible with CHR&R measures that cover a broad range of health factors and outcomes, 3 pilot projects were conducted in 2015 by Washington University in St Louis in partnership with the Missouri Hospital Association, New York State Department of Health, and the California Department of Public Health. The overall aims of the pilot projects were to (1) provide local data from multiple sources for a broad range of measures to support community health needs assessments and development of community health improvement plans and (2) develop analytical capability for subcounty data analyses and presentation to support public health activities. Detailed technical methods, analytic techniques, sample data, and programs (SAS and R) from

the 3 projects have since been shared as a white paper for data analysts to utilize in future projects.²² This article aims to summarize key considerations and lessons learned from the pilot projects to adopt and adapt the CHR&R model and measures to generate data products for subpopulations and small areas below the county level.

Methods

The pilot teams' processes shared many common steps, despite having varied data sources, measures, and outputs (Figure). Herein is a summary of the key stages and important considerations. The pilot project in New York was approved by the New York State Department of Health Institutional Review Board (IRB reference #15-041). For the pilot project in Missouri, employment of the aggregate data utilized as model inputs was governed by Hospital Industry Data Institute master data use agreements and academic personnel participation was reviewed by the Washington University School of Medicine Human Research Protection Office.

Conceptual development for data sources and measures

Identifying target audiences

The projects identified target audiences that included local public health departments, partner government agencies, hospital associations, hospital community benefits organizations, health planning organizations, community-based organizations, population health and strategic planning personnel, and academic researchers. The New York project supported the New York State Prevention Agenda, which requires local health departments, hospitals, and other community partners to collaborate on community health assessments and community health improvement plans every 3 years. In California, the measures were produced by the Healthy Communities Data and Indicators Project (HCI) of the Office of Health Equity. The HCI regularly produces subcounty data on the social determinants of health and disparities to inform stakeholders such as the California Health in All Policies Task Force and “Let’s Get Healthy California” state health improvement plan. The Missouri ZIP Health Rankings Project²³ aimed to assist hospitals, public health departments, community-based organizations, and funders with appropriately targeting scarce community health improvement resources.²⁴

Selecting measures

Criteria for selecting health measures may include the geographical units of data that are available, the

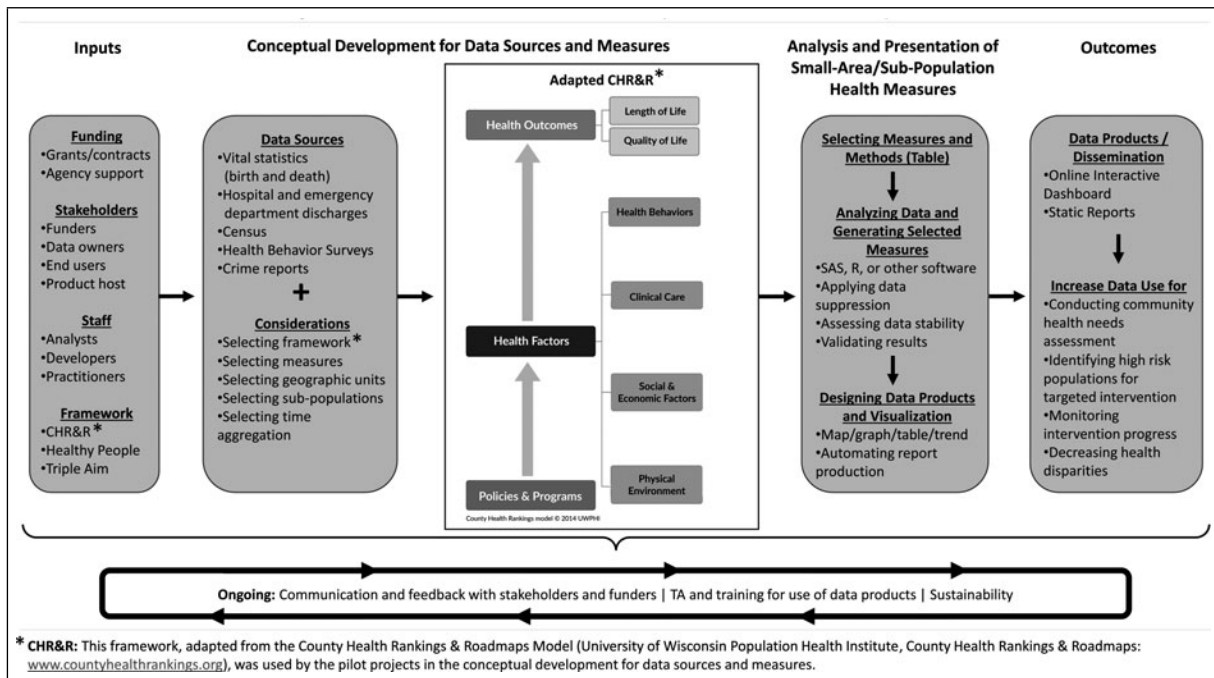


FIGURE General Model of Subcounty Health Data Development Used with permission. Abbreviations: CHR&R, County Health Rankings & Roadmaps; TA, technical assistance.

volume of the measured events, the extent of data time aggregation necessary to obtain stable estimates, local priorities, the sustainability and longevity of the data source, and the cost. The 3 projects were able to create 22 out of 35 ranked measures from the CHR&R model based on these criteria. The sources included birth, death, and hospitalization administrative data, health survey data, US Census data, and modeled data (see Supplemental Digital Content Table A, available at <http://links.lww.com/JPHMP/A652>). When possible, the data were further disaggregated by county subpopulations such as race/ethnicity, sex, poverty level, and disability status to inform on disparities.

Some ranked measures, such as alcohol-impaired driving deaths, were not selected because of event rarity. Other measures were selected as proxies due to unavailability of data at the subcounty level that met the specifications of measures from the CHR&R model. For example, the New York project substituted teen births (original measure) with teen pregnancies, and hospitalizations for ambulatory care-sensitive conditions were substituted with preventable hospitalizations. The California project adjusted the age groups for unemployment and educational attainment measures according to those available in existing data sources. The Missouri team selected multiple measures that correlated (as proxies) with existing CHR&R measures using a 2-step process that included an examination of face and criterion validity.²⁴

Selecting geographic units

Selection of geographic units of analysis depended on data availability, time range of data aggregation necessary to obtain stable estimates, and the familiarity and usability by the target audience. Several solutions were implemented to improve the stability of estimates with the desired geographic granularity, including the use of state-specific data sources that capture larger samples or populations than those available from national data sources. For example, in contrast to calculating preventable hospitalizations among Medicare patients (aged 65 years and older), New York used a statewide all-payer database to generate a similar measure for all adult patients. For certain measures, pilot projects used more years of data for subcounty measures than CHR&R used for county-level measures. Even within a single state, all measures were not available in the same geographical units. Therefore, New York and California used a variety of small-area geographic units, including zip codes (ZCTAs), minor civil divisions, and census tracts (see Supplemental Digital Content Table A, available at <http://links.lww.com/JPHMP/A652>). Geographic aggregation of zip codes into minor civil divisions was used for survey data by New York. In contrast, Missouri developed a methodology to produce all measures at the zip code level.

Analyzing and presenting small-area/subpopulation health measures

Generating subcounty statistics

The 3 projects followed different methods to analyze data and generate estimates for each measure using statistical software (eg, SAS, R) (Table). The New York team aggregated individual records from administrative data (eg, births, deaths, emergency department visits, and hospitalizations) to generate counts, rates, and percentages for selected measures, by county population characteristics (eg, race/ethnicity, age group, Medicaid status, and education levels) and by geographic areas. For survey data, zip codes were aggregated to generate estimates for minor civil divisions.

The California team used an application programming interface (API) to automate downloads of American Community Survey tables that were subsequently used to aggregate demographic groups within selected geographies. For some measures, model-based estimates were produced for multiple small areas (including ZCTAs, cities, congressional districts, and assembly districts) via a subcontract award with the

California Health Interview Survey Neighborhood Edition.

The Missouri team conducted principal components analyses to derive composite zip-level scores corresponding to CHR&R subdomain scores.²⁴

Applying data suppression

Increased granularity of reporting increases the risk that individuals with certain health outcomes could be identified. To protect individuals' confidential information and to address problems arising from skewed or potentially miscoded underlying data, estimates were sometimes suppressed (ie, not reported). Three types of data suppression were applied: primary, secondary, and tertiary.

Primary suppression rules, usually based on minimum volume thresholds for estimate numerators and/or denominators, vary depending on the data source and criteria used in each pilot project are described in Supplemental Digital Content Table A (available at <http://links.lww.com/JPHMP/A652>). Secondary suppression was applied when primary suppression affected only one subpopulation or one geographic area in a county; suppressing an

TABLE
Summary of Project Data Sources, Subcounty-Level Estimates and Methods

	State		
	California	Missouri	New York
Data sources	Demographic statistics database Health behaviors survey Housing data Aggregate crime reports data	Hospital inpatient, outpatient, and emergency department discharge data Socioeconomic deprivation index database ²⁵ Census-based data set	Health behaviors survey Vital statistics (birth and death) Health care discharge data
Subcounty-level data	<i>Demographic:</i> Disability status Poverty level Race/ethnicity Sex <i>Geographic:</i> Assembly districts Census tract City/town Congressional districts ZCTA	<i>Demographic:</i> The clinical, SES, and census-based data used were available at both zip code and county levels (see ref 20 Supplemental Digital Content file, available at: http://links.lww.com/JPHMP/A652) <i>Geographic:</i> Zip code	<i>Demographic:</i> Age group Education Medicaid status Race/ethnicity <i>Geographic:</i> Minor civil division (where data available) Zip code
Methods	Aggregating data from existing sources such as the American Community Survey; eg, aggregating demographic groups (sex or age) within smaller geographies (eg, census tract) Generating model-based estimates for subcounty geographies	Ranking zip code health factor and health outcome composite scores generated from estimates for multiple measures	Aggregating individual record data into rates or percentages for individual measures For survey data (Behavioral Risk Factor Surveillance System survey), zip codes were aggregated into minor civil divisions

Abbreviations: SES, socioeconomic status; ZCTA, zip code tabulation area.

additional cell (ie, secondary suppression) prevents the identification of data for the primary-suppressed cell, which could otherwise be calculated by subtracting the sum of the unsuppressed cells from the county's total count. Tertiary suppression was applied to remove outlier estimates that resulted from coding errors (eg, in patients' demographic information) or from skewed age group distributions among cases that caused age adjustment to produce extreme values. In addition, the Missouri team applied Winsorization²⁶ criteria (top coding) to estimates to minimize the risk of identifying individuals.

Assessing data stability

Disaggregating data result in equal or smaller (but never larger) counts and in proportions with equal or smaller (but never larger) denominators. Sub-county estimates are therefore less stable than county-level estimates. General guidelines state that, for count measures, estimates with a relative standard error (RSE) greater than 30% should be considered unreliable/unstable.²⁷ This usually occurs when there are fewer than 10 events in the numerator.²⁸ For measures using survey data, guidelines state that an estimate can be considered unreliable/unstable when the width of the 95% confidence interval is greater than 20% or the RSE is greater than 30%.²⁹

The research teams found that most end users preferred data products to include as much data as possible, even when estimates were unstable. Academic and technical users may already understand the limitations of making inferences based on unstable estimates and may prefer to receive all subcounty estimates. Nontechnical users may not have prior understanding of the limitations and therefore may prefer to receive only stable subcounty estimates that can be used in practice. Accordingly, the New York team flagged unstable estimates with asterisks and explained estimate stability in the Methods section of its reports. The California research team provided confidence intervals and RSEs in its data products, along with materials on how to interpret data stability.

Validating results

It is important to assess the distribution of estimate values for each measure to identify outliers and determine whether they are acceptable. Outliers can result not only from truly extreme health disparities in specific communities but also from errors in raw data coding (eg, race/ethnicity miscoded on individual records) or overadjustments by statistical procedures (eg, age-adjusting estimates for small geographies with skewed underlying age distributions). Each

team examined outlying estimates carefully to identify and exclude erroneous data (eg, coding errors, differentially adjusted estimates) while avoiding exclusion of "real" data with truly outlying values that reflect true health burdens in the population.

The teams reviewed univariate distributions of subcounty estimates for count data. Estimates with values exceeding the respective measure's statewide 90th percentile were compared with county-level estimates. Extreme values were investigated individually and suppressed if deemed invalid (tertiary suppression). For age-adjusted estimates, age distributions of the underlying populations were checked as well. For the composite scores, zip codes were suppressed if 2 or more of their scores were larger than 3 standard deviations and could not be sufficiently explained.

Designing data outputs and visualizations

The research teams based their data product designs and media on end users' needs and input. During the development processes, each team conducted active discussion (eg, focus groups) with key stakeholders to collect and review their feedback and suggestions regarding data product designs.

It was found that users preferred to receive data products with data visualizations (eg, graphs, maps, trends, data tables) in formats that are easily accessible (eg, PDF, online query). Users also preferred to receive data in formats that they could further manipulate or use to generate their own visualizations (eg, Excel). In addition, teams identified needs to provide users with simple technical guidance and explanations of methods and limitations to support accurate interpretation of data products.

Standardizing and automating report production

It is important to consider output and visualization designs before developing a production process because factors such as data set structure and the formation of technical programs in SAS or R may depend on formats and features of the outputs and visualizations. The 3 research teams incorporated standardization and automation with these considerations in mind.

The California and New York teams used SAS and R to automate data processing, including data import, data aggregation, calculation of rates, calculation of reliability measures, data visualization (eg, graphs, maps, tables), final formatting, and export. When available, APIs were used to download data from secondary sources. All measures followed a standardized data output format. The Missouri team worked

with partners to develop an online data platform that includes functions such as mapping, report building, and downloadable public use files.³⁰

Disseminating reports and other data products

To support and enhance utilization of final data products, researchers can develop dissemination strategies to reach key audiences, such as Web site publication, direct distribution, in-person presentations, or a combination of methods. Project teams worked with stakeholders and leveraged their networks to share data products with key audiences.

In California, data were published on the California Department of Public Health Web site and announced to a large number of stakeholders. In Missouri, the launch of the exploreMOhealth³⁰ data platform was announced via press releases, region-specific fact sheets, presentations at a statewide meeting, and a webinar. In New York, the reports were e-mailed directly to local health departments and then publicly released on the New York State Association of County Health Officials Web site.³¹ The team also held a webinar for local health departments, hospitals, regional health planning groups, and other partners to introduce the reports and help with interpreting them.

Planning for sustainability

All 3 projects aimed to provide data products to support end-user needs in applied settings after the initial product release. Sustainability requires teams to strategically use available data and resources, have well-documented operational processes, and secure funding.

One consideration with implications for sustainability is the ongoing availability and cost of data needed to derive subcounty measures. Most data for the pilot projects were freely available and frequently refreshed to support periodic product updates. However, population counts for small areas sometime need to be purchased from commercial sources. The Missouri research team transitioned from using commercial data sources to a publicly available source to maintain production without incurring untenable costs. Sustained operations also involve additional costs, including providing user outreach and training, implementing processes to collect user feedback, administering support, and conducting ongoing research and development to improve measures and data products.

Developing and documenting standardized processes and workflows to accurately produce small-area estimates were necessary for maintaining ongoing production and assisted in the production of

similar projects. For example, the New York team documented technical methods and SAS programs as protocols so that new staff could easily modify and apply them to another subcounty data project.³²

Securing budgetary support for sustained operations was a de facto consideration for all 3 research teams. For the Missouri team, pilot project funding did not support ongoing delivery of data and reports. Therefore, the Missouri team identified and negotiated shared funding through 2 foundations to support both ongoing operations and the development of a shared Web-based reporting platform.

Results

One lesson learned from these pilot projects was that there are trade-offs between estimates' geographic and demographic granularities, and both stability and suppression. Especially when working with subcounty data, maintaining privacy is essential. Because regulations may vary depending on the data source and where data are to be displayed, it is important for researchers to follow data suppression policies and assess identifiability risks. It is still possible to release granular estimates by applying necessary suppression criteria and including accompanying information about estimate stability and other data limitations.

In addition, proxy measures are often needed because of the lack of availability of measures at the subcounty level. The CHR&R model is a well-known framework for assessing community conditions and health outcomes at the county level. When attempting to replicate this model at the subcounty level, the pilot projects often needed to identify proxy measures to help fill information gaps when subcounty-level data did not match the county-level measure. The pilot projects identified proxies by using the same measure but with a different universe (eg, population 18 years and older in substitution of Medi-Cal enrolled population), using a similar measure (eg, teen pregnancy in substitution of teen births), and using a new measure that correlates with the original measure (eg, hospital utilization rates for mental health in substitution of the average number of mentally unhealthy days reported in past 30 days).

Obtaining health outcomes data at the subcounty level is perhaps the biggest technical challenge in subcounty data analyses. In general, there are 3 options for subcounty health-related data, each with differing limitations:

1. *Survey oversample*: A limitation of this approach is that it is expensive and self-reported.

2. *Modeled estimates* (small-area estimation): A limitation of this approach is that it generalizes population characteristics and will not capture variation due to public programs (eg, local tobacco control programs), which could make model estimates substantially different from direct estimates.^{33,34}
3. *Administrative data* (eg, health claims data): A limitation of this approach is that the data are not population based but rather reflect the population receiving health care.

Another technical challenge is validating outliers. Widely accepted criteria for distinguishing real outliers (where health burdens are truly extreme) from erroneous outliers (eg, caused by data entry errors, or skewed adjustment) do not exist. Future research on validation algorithms or best practices could greatly contribute to subcounty data analyses.

These technical challenges also impact user engagement. The general public can be surprised and disappointed if data cannot be provided for their cities or neighborhoods for multiple health outcomes. This issue can be magnified by different users' needs for more granular geographical levels and subpopulation data (ie, race and ethnicity). Geographic areas of interest to users can include, for example, neighborhoods, voting districts, hospital service areas, and school districts.

Different users will also require different levels of estimate stability. One known method to increase the stability of small-area estimates is aggregating data from adjacent geographies.³⁵ However, in this case, the analyst's aggregation choices may or may not align geographically with how communities define themselves. In some cases, a less stable estimate with clearly stated limitations will suffice while other users may require increased estimate stability. Especially as multisector collaborations to improve health increase, finding ways to locate and combine relevant data across multiple sectors and from multiple source types, and to display these data in geographic areas that are meaningful to different stakeholders are becoming increasingly important. When the unit of analysis is a subcounty area, this work is more challenging.

Locating and analyzing subcounty data across sectors is resource intensive. While there is great interest in the use and availability of these types of data, funding and sustaining services can be challenging. Producing subcounty data on a regular basis requires full-time dedicated staff. However, sustainable funding to produce small-area estimates for health outcomes is often lacking, whether with subcontractors or through the development of in-house capabilities.

Implications for Policy & Practice

Future subcounty data projects should have clearly defined frameworks/models, goals, and target audiences.

- Researchers should consider technical considerations early on and throughout subcounty data projects.
- Data product designs should be discussed while analyzing data and generating estimates so that, at later steps, results can be organized and structured to streamline the production of data products.
- Decisions on how or whether to present unstable data should consider the needs of end users, users' ability to interpret unstable estimates, and institutional policy.
- Automating analyses and production processes can improve the consistency of reported data and the quality of data visualizations.
- Clearly documenting methods and processes further supports project sustainability, facilitates staff transitions, and enables project methods to be adapted for subsequent work.
- Technical assistance on how to use the data products and interpret the results should be provided when releasing data products.
- At all stages, input from key stakeholders can help inform the project considerations.
- To increase sustainability, investigators interested in pursuing subcounty health data projects should design the project to align with organizational interests.
- Funders should provide support for multiyear projects to ensure greater impact and sustainability.

For example, having dedicated staff to work on indicator data projects might be difficult to secure in local or state health departments, as budget earmarks and competing priorities can make it difficult to dedicate staff exclusively to a single project.

Conclusions

Generating, communicating, and sustaining subcounty data are critical steps for advancing health and equity. The generally limited availability of resources for implementing evidence-based public health interventions highlights the needs for small-area data so that interventions can be more effectively targeted. This article outlines opportunities and challenges practitioners may face in this work, shares lessons learned, and offers 3 pilot projects that successfully developed and disseminated small-area data as useful models for future projects.

References

- County Health Rankings & Roadmaps. County Health Rankings & Roadmaps; a Robert Wood Johnson Foundation program. County Health Rankings national data. <http://www.countyhealthrankings.org>. Published 2017. Accessed December 27, 2017.
- Webber WL, Stoddard P, van Erp B, et al. A tool for providing data on small areas: development of neighborhood profiles for Santa Clara County, California, 2014. *Public Health Rep*. 2016;131(1):35-43.
- Stoto MA, Davis MV, Atkins A. Making better use of population health data for community health needs assessments. *EGEMS (Wash DC)*. 2019;7(1):44.
- Malec D. *Statistical Small Area Estimation: Some Examples and Current Projects at NCHS*. Atlanta, GA: Centers for Disease Control and Prevention; 2013. <https://ncvhs.hhs.gov/wp-content/uploads/2014/05/130501p02.pdf>. Accessed February 13, 2020.
- Comer KF, Gibson PJ, Zou J, Rosenman M, Dixon BE. Electronic health record (EHR)-based community health measures: an exploratory assessment of perceived usefulness by local health departments. *BMC Public Health*. 2018;18(1):647.
- Giovenco DP, Spillane TE. Improving efficiency in mobile data collection for place-based public health research. *Am J Public Health*. 2019;109(S2):S123-S125.
- irs.gov. Requirements for 501(c)(3) Hospitals Under the Affordable Care Act—Section 501(r). <https://www.irs.gov/charities-non-profits/charitable-organizations/requirements-for-501-c3-hospitals-under-the-affordable-care-act-section-501r>. Updated September 20, 2019. Accessed February 7, 2020.
- irs.gov. Community health needs assessment for charitable hospital organizations—Section 501(r)(3). <https://www.irs.gov/charities-non-profits/community-health-needs-assessment-for-charitable-hospital-organizations-section-501r3>. Updated September 20, 2019. Accessed February 7, 2020.
- Public Health Accreditation Board. Standards and Measures Version 1.5. https://phaboard.org/wp-content/uploads/PHABSM_WEB_LR1-1.pdf. Published 2013. Accessed February 7, 2020.
- Kelsall JE, Diggle PJ. Kernel estimation of relative risk. *Bernoulli*. 1995;1(1/2):3-16.
- Eaton N, Shaddick G, Dolk H, Elliott P. Small-area study of the incidence of neoplasms of the brain and central nervous system among adults in the West Midlands region, 1974-86. Small Area Health Statistics Unit. *Br J Cancer*. 1997;75(7):1080-1083.
- Wakefield J, Elliott P. Issues in the statistical analysis of small area health data. *Stat Med*. 1999;18(17/18):2377-2399.
- Baltimore City Health Department. Neighborhood health profile reports. <https://health.baltimorecity.gov/neighborhood-health-profile-reports>. Accessed February 2, 2018.
- Boston Foundation. The Boston Indicators Project. <http://www.bostonindicators.org/topics/neighborhoods-and-housing?q&sortBy&date&sortOrder=desc&page&2>. Accessed February 2, 2018.
- healthyplan.la. Plan for a Healthy Los Angeles. Health profiles. <http://healthyplan.la/interactive/neighborhoods>. Accessed February 2, 2018.
- City of Minneapolis. Minneapolis neighborhood profiles. <http://www.ci.minneapolis.mn.us/residents/neighborhoods/index.htm>. Accessed February 2, 2018.
- New York City Department of Health and Mental Hygiene. New York City community health profiles. <http://www1.nyc.gov/site/doh/data/data-publications/profiles.page>. Published 2015. Accessed February 2, 2018.
- Santa Clara County Public Health Department. City and small area/neighborhood profiles. <https://www.sccgov.org/sites/phd/hi/hd/Pages/city-profiles.aspx>. Accessed February 2, 2018.
- City of Toronto. Neighbourhood profiles. <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=ae17962c8c3f0410VgnVCM1000071d60f89RCRD>. Accessed February 2, 2018.
- The National Equity Atlas. PolicyLink, University of Southern California Program for Environmental and Regional Equity. <http://nationalequityatlas.org>. Published 2016. Accessed March 20, 2018.
- Remington PL, Catlin BB, Gennuso KP. The County Health Rankings: rationale and methods. *Popul Health Metr*. 2015;13:11.
- Generating sub-county health data products: methods and recommendations from a multi-state pilot initiative. County Health Rankings & Roadmaps Web site. <https://www.countyhealthrankings.org/sites/default/files/resources/Working%20Paper%20All%20files.pdf>. Published September 17, 2019. Accessed November 7, 2019.
- Our methods. County Health Rankings & Roadmaps Web site. <http://www.countyhealthrankings.org/explore-health-rankings/our-methods>. Accessed June 26, 2018.
- Nagasako E, Waterman B, Reidhead M, Lian M, Gehlert S. Measuring subcounty differences in population health using hospital and census-derived data sets: the Missouri ZIP Health Rankings Project. *J Public Health Manag Pract*. 2018;24(4):340-349.
- 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.
- Wilcox RR. Winsorized robust measures. In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, eds. *Wiley Stat-Ref: Statistics Reference Online*. John Wiley & Sons, Ltd; 2017. doi:10.1002/9781118445112.stat06339.pub2.
- Klein RJ, Proctor SE, Boudreault MA, Turczyn KM. *Healthy People 2010 Criteria for Data Suppression*. Hyattsville, MD: National Center for Health Statistics; 2002. Statistical Notes No. 24.
- Rates based on small numbers—statistics teaching tools. New York State Department of Health Web site. <https://www.health.ny.gov/diseases/chronic/ratesmall.htm>. Revised April 1999. Accessed June 7, 2018.
- Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. *Vital Health Stat*. 2017;(175):1-22. https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf. Accessed November 19, 2018.
- exploreMOhealth. Home page. <https://exploremohealth.org>. Accessed June 25, 2018.
- Sub-county health data report for county health rankings-related measures 2016. New York State Association of County Health Officials Web site. <http://www.nysacho.org/i4a/pages/index.cfm?pageID=3810>. Published May 2016. Accessed June 7, 2018.
- New York State 2017 Health Equity Reports. https://www.health.ny.gov/statistics/community/minority/mcnd_reports.htm. Accessed June 25, 2018.
- Zhang X, Holt JB, Yun S, Lu H, Greenlund KJ, Croft JB. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. *Am J Epidemiol*. 2015;182(2):127-137.
- Srebotnjak T, Mokdad AH, Murray CJ. A novel framework for validating and applying standardized small area measurement strategies. *Popul Health Metr*. 2010;8:26.
- Talbot TO, Kumar S, Babcock GD, Haley VB, Forand SP, Hwang SA. Development of an interactive environmental public health tracking system for data analysis, visualization, and reporting. *J Public Health Manag Pract*. 2008;14(6):526-532.