SDC

Cluster Analysis Methods to Support Population Health Improvement Among US Counties

Elizabeth A. Pollock, PhD; Ronald E. Gangnon, PhD; Keith P. Gennuso, PhD; Marjory L. Givens, PhD, MSPH

ABSTRACT

Context: Population health rankings can be a catalyst for the improvement of health by drawing attention to areas in need of relative improvement and summarizing complex information in a manner understood by almost everyone. However, ranks also have unintended consequences, such as being interpreted as "hard truths," where variations may not be significant. There is a need to improve communication about uncertainty in ranks, with accurate interpretation. The most common solutions discussed in the literature have included modeling approaches to minimize statistical noise or borrow strength from covariates. However, the use of complex models can limit communication and implementation, especially for broad audiences.

Objectives: Explore data-informed grouping (cluster analysis) as an easier-to-understand, empirical technique to account for rank imprecision that can be effectively communicated both numerically and visually.

Design: Cluster analysis, specifically k-means clustering with Wasserstein (earth mover's) distance, was explored as an approach to identify natural and meaningful groupings and gaps in the data distribution for the County Health Rankings' (CHR) health outcomes ranks.

Setting: County-level health outcomes from the 2022 CHR.

Participants: 3082 counties that were ranked in the 2022 CHR.

Main Outcome Measure: Data-informed health groups.

Results: Cluster analysis identified 30 health groupings among counties nationwide, with cluster size ranging from 9 to 184 counties. On average, states had 16 identified clusters, ranging from 3 in Delaware and Hawaii to 27 in Virginia. Number of clusters per state was associated with number of counties per state and population of the state. The method helped address many of the issues that arise from providing rank estimates alone.

Conclusions: Public health practitioners can use this information to understand uncertainty in ranks, visualize distances between county ranks, have context around which counties are not meaningfully different from one another, and compare county performance to peer counties.

KEY WORDS: cluster analysis, communication, population health outcomes, ranking, uncertainty

Author Affiliations: *Department of Population Health Sciences, University of Wisconsin Population Health Institute, University of Wisconsin-Madison, Madison, Wisconsin (Drs Pollock, Gennuso and Givens); and Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin (Dr Gangnon).*

This work was made possible by support from the Robert Wood Johnson Foundation.

The authors have no financial relationships relevant to this submitted work.

The authors gratefully acknowledge contributions and content review from our colleagues Dr Nicholas Brigham Schmuhl and Dr Christine Muganda.

A protocol approval was not needed by an ethics committee to conduct this research.

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).

Correspondence: *Keith P. Gennuso, PhD, Department of Population Health Sciences, University of Wisconsin Population Health Institute, 524 WARF Office Building, 610 Walnut St., Madison, WI 53726 (gennuso@wisc.edu).*

DOI: 10.1097/PHH.0000000000002034

Population health rankings can be a catalyst for the improvement of health by drawing attention to areas in need of relative improvement through a synthesis of community health data.[1] County Health Rankings (CHR), a population health improvement platform affiliated with the authors, specifically aims to raise awareness of the factors that influence health and its variation from place to place. Each county has been ranked within its state—from most to least healthy—based on the health outcomes of each county and the factors that contribute to health.[2] Once aware of problem areas, communities can use CHR resources to access and enact evidence-informed policies and programs to improve overall health and reduce inequities. Fundamentally, health rankings are a tool to help health professionals, local community leaders, and the general public make informed decisions about the health of their communities.[3]

Ranking can be an effective, albeit controversial, method of calling attention to differences in society—from the oft-cited *US News and World Report* rankings of colleges and universities to international rankings of economy, education, or technology. Rankings have the ability to summarize complex information about a topic in a manner that can be understood by almost everyone.[4] Population health rankings, in particular, help set agendas—stimulating awareness, motivation, and debate over means to improved health outcomes, establish broad responsibility for population health, and emphasize the need for multisectorial collaboration to improve outcomes.[5]

While there are advantages to ranking, there are also disadvantages or unintended consequences for public health practice. While rankings can often confidently identify those at the top or the bottom, such as the most and least healthy counties, they frequently cannot distinguish meaningful differences in the middle of the distribution.[6,7] Value judgments are inherent in any rankings methodology, and underlying data are often limited due to validity, reliability, or completeness concerns.[4] Ordering of counties within states, where variations may not be practically, or even statistically, significant, are often interpreted as "hard truths," leading users to make comparisons or conclusions about health differences that may not exist in reality. Some have argued that the influence of randomness and error on rankings in general is so large that ranking should not be considered "evidence-based" or used in policy or funding decisions.[1-3,8-14] As an example, a county health department could choose to take action on a specific health dimension after observing a neighboring county's single rank is more favorable than theirs or a state agency could choose to fund one county's project over another's on the basis of lower health rank, when in these instances these two counties may not actually have meaningfully different health experiences.

Studies have attributed double-digit shifts in rank to statistical noise in individual component data. Small population sizes, small numbers of counties, homogeneity of outcomes within states, and heterogeneity of outcomes within counties all contribute to this problem.[6,14-16] In contrast, large population sizes can lead to an overemphasis on statistically significant, but practically unimportant, differences between regions.[17] From an equity perspective, rural places and places with large disparities in outcomes may be more likely to lack a clear assessment of their relative status, resulting in more ambiguity and potentially misleading policy messages. Furthermore, because ranks are inherently comparative, error in one region affects every rank.[8,11] Rankings can change not only because of improvement or degradation in outcomes for the county itself, but also due to improvement, degradation, or statistical noise in the outcomes for other counties.[1]

Since rankings on an ordinal scale do not necessarily reflect practical or even statistically significant differences, CHR has always recommended examining the data which underlie the rank, particularly since no model—especially one based on free and publicly available data—can be perfectly precise. However, there is a need to improve communication about error and reliability in the ranks, giving end users the tools to correctly interpret the rankings.[12,18]

There are many ways to convey uncertainty to audiences, such as confidence intervals, potential ranges for ranks, quantiles, or data-informed groupings. Suggestions regarding better communication of error and real-world significance or applicability of rankings have been discussed in the literature. For instance, ranges and visual distributions (eg, gradient and density plots) have been suggested as effective alternatives to point estimates of rankings.[14,19-22] Modeling approaches to minimize statistical noise, such as Bayesian modeling procedures, have also been discussed.[6,14,15] Finally, models that borrow strength from other features such as time trends, geographic comparisons, covariates, joint outcomes, longitudinal data, data-driven weights, spatial smoothing, and more have also been explored as solutions.[1,3,6,15,23-25] However, incorporating additional features such as these into health outcome models cannot eliminate the problems caused by "noisy" data at the county level.[6,15]

While documenting the imprecision of ranks and avoiding misperceptions regarding actual differences in health is very important, the use of complex models can create limitations and difficulties of communication and implementation. A major appeal of health rankings is their simplicity.[3,8,26] Reports of rankings that overemphasize methodological caveats may be difficult to understand and raise concerns about validity, thus limiting their effectiveness as tools for improving population health.[3] Ultimately, given the air of authoritativeness given to rankings by some audiences, there is a crucial need for organizations that generate rankings to responsibly address the issue of error in the ranks in a methodologically robust but easily communicated, way.[5]

Despite many options to convey uncertainty, communicating uncertainty remains a difficult task.[20,27,28] Statistical assessments of uncertainty (confidence intervals, standard errors, *P* values, etc) often lead to confusion, misinterpretations, and decreased trust in information for broad audiences and users.[19,21,28] Users of CHR data range from community members to local health departments, policymakers, academic researchers, and more, which supports the need to prioritize communication to broad audiences. As an alternative, data-informed grouping may be an easier-to-understand technique that accounts for rank imprecision and can be effectively communicated using both numeric and visual displays.[27]

Methods

Commonly used ways to group data, such as quartiles (or other quantiles), are easy to calculate and communicate but insensitive to the magnitude of differences in health outcomes between counties. A county at the boundary of a quartile is frequently more similar to some counties in the adjacent quartile than to some counties within its own quartile. Cluster analysis, on the other hand—a name given to a collection of methods for grouping objects (such as a county) based on similarity—empirically identifies natural, meaningful gaps in the health outcomes across counties by using data to inform the number and size of groups. Thus, this method should better reflect any natural breakpoints in the health outcome distribution, and properly constructed clusters will be more reflective of the underlying distribution of health outcomes than groupings created using quantiles.

A cluster analysis approach was applied to values of a composite index (ie, a *z* score) consisting of a weighted combination of the 5 measures that make up the CHR health outcomes rank (premature death, poor/fair health, poor physical health days, poor mental health days, and low birthweight) for every county that received a rank in the 2022 CHR, representing all 50 states. These measures have been long-established in CHR's model of population health for more than a decade and were selected based on many factors, including reflection of important aspects of health that can be improved, validity, reliability, and use by others in the field, availability at the county level, and more.[3] Rather than calculating within-state *z* scores like the standard CHR methodology, which limits the ability to compare counties across the country or with peers in neighboring states or regions, we utilized the national distribution of *z* scores, which also has the added benefits of more data power with less statistical noise and only one overall decision on cluster cutoff (vs. 50 separate decisions). This is especially helpful for states with very few counties. K-means clustering was used to identify the optimal grouping of the counties for each possible number of groups (1, 2, . . ., 100). In K-means clustering, k random centroids of the data are created, and each data point is assigned to the nearest centroid (creating a cluster). The centroid of each cluster is then moved to the average of the data in the cluster and the process is repeated until no data points change groups. Wasserstein (earth mover's) distance, a measure of the distance between 2 probability distributions, was used to quantify the loss of information due to clustering. The number of clusters was chosen as the smallest value of k for which the Wasserstein distance was below the 95th percentile of its permutation distribution. There are many valid methodological options to choose from when using cluster analysis, and while the following outlined decision points were found to work the best for our purposes, others are encouraged to tailor their choices accordingly. A state's number of clusters will vary with the number of counties within a state, the population sizes of those counties, and the heterogeneity (or homogeneity) of outcomes across counties within a state.

Clustering was performed on the *z* score point estimates alone, ignoring the statistical uncertainty regarding these estimates. Quantifying the error in the *z* score is a challenging modeling problem, particularly given that some of the individual measures that comprise the *z* scores are, themselves, modeled estimates for which full information on the joint distribution of the errors is unavailable. Failure to account for uncertainty in the *z* scores is likely to result in an overestimate of the number of clusters, but additional study is required to quantify this effect.[29] Analyses were performed using the Ckmeans.1d.dp and transport packages in R.[30-32] Results from the state-by-state cluster analyses were summarized and examined for associations with relevant state-level characteristics; Supplemental Digital Content Figure 3, available at http://links.lww.com/JPHMP/B415.

Results

Using the 3082 counties that were ranked in the 2022 CHR, 30 groupings of counties (clusters) nationwide were identified. The smallest cluster contained 9 counties, the largest 184 counties, with an average cluster size of 103 counties, as shown in Table 1 and Figure 1. Figure 2 shows the distribution across the country of the unclustered health outcomes *z* score values, and Figure 2 shows the distribution of the 30 identified clusters. Less healthy clusters tended to be in the South, Southwest, and Appalachian regions of the country. Healthier clusters tended to be in the Northeast, Upper Midwest, and Pacific Northwest regions. Comparing Figure 2a and b demonstrates how clustering helped to simplify the presentation of counties across the nation without losing the overall distributional information; Supplemental Digital Content Figure 1, available at http://links.lww.com/JPHMP/B413.

TABLE 1
Number of Clusters and Counties Within Clusters, Nationwide, and by State

State	Number of Ranked Counties (n)	Number of Clusters (n)	Average Number of Counties per Cluster (n)	Minimum Number of Counties in a Cluster (n)	Maximum Number of Counties in a Cluster (n)
United States	3082	30	102.7	9	184
Texas	244	23	10.6	1	31
Georgia	159	24	6.6	1	13
Virginia	133	27	4.9	1	12
Kentucky	120	21	5.7	1	13
Missouri	115	24	4.8	1	19
Kansas	104	19	5.5	1	16
Illinois	102	20	5.1	1	12
North Carolina	100	25	4.0	1	14
Iowa	99	12	8.3	1	17
Tennessee	95	16	5.9	1	13
Indiana	92	19	4.8	1	11
Ohio	88	21	4.2	1	11
Minnesota	87	14	6.2	1	19
Michigan	83	18	4.6	1	14
Mississippi	82	19	4.3	1	10
Nebraska	79	15	5.3	1	10
Oklahoma	77	19	4.1	1	10
Arkansas	75	19	3.9	1	9
Wisconsin	72	16	4.5	1	12
Alabama	67	19	3.5	1	9
Florida	67	21	3.2	1	7
Pennsylvania	67	15	4.5	1	16
Louisiana	64	17	3.8	1	10
New York	62	12	5.2	1	11
South Dakota	61	21	2.9	1	9
Colorado	59	21	2.8	1	5
California	58	17	3.4	1	8
West Virginia	55	19	2.9	1	8
North Dakota	48	15	3.2	1	7
Montana	47	19	2.5	1	6
South Carolina	46	19	2.4	1	6
Idaho	43	15	2.9	1	8
Washington	39	14	2.8	1	7
Oregon	35	14	2.5	1	7
New Mexico	32	15	2.1	1	6
Utah	28	14	2.0	1	3
Alaska	24	16	1.5	1	3
Maryland	24	13	1.8	1	3
Wyoming	23	12	1.9	1	4
New Jersey	21	11	1.9	1	4
Maine	16	11	1.5	1	3
Nevada	16	9	1.8	1	3
Arizona	15	11	1.4	1	3
Massachusetts	14	8	1.8	1	3

(*continues*)

TABLE 1
Number of Clusters and Counties Within Clusters, Nationwide, and by State (*Continued*)

State	Number of Ranked Counties (n)	Number of Clusters (n)	Average Number of Counties per Cluster (n)	Minimum Number of Counties in a Cluster (n)	Maximum Number of Counties in a Cluster (n)
Vermont	14	8	1.8	1	3
New Hampshire	10	6	1.7	1	3
Connecticut	8	5	1.6	1	2
Rhode Island	5	4	1.3	1	2
Hawaii	4	3	1.3	1	2
Delaware	3	3	1.0	1	1

On average, states had 16 identified clusters, ranging from a minimum of 3 clusters in Delaware and Hawaii to a maximum of 27 clusters in Virginia, as shown in Table 1. The average number of counties within a cluster at the state level was 4 counties, with a minimum of 1 county per cluster and a maximum of 31 counties per cluster. Delaware, the state with the least number of counties at 3, had the lowest average counties per cluster, with 3 clusters comprising one county each, while Texas, the state with the most counties at 244, had the highest average counties per cluster at 11. Differences between states with similar numbers of counties were observed. For example, North Carolina, with 100 counties, and Iowa, with 99 counties, had 25 and 12 identified clusters, respectively. This demonstrates how the cluster analysis worked differently for relatively heterogenous verses homogenous states, respectively, identifying more clusters in North Carolina whose distribution is more spread out with more distances between ranks, while identifying fewer clusters in Iowa, which has a very tight distribution of z scores. As another example, New York, with 62 counties, and South Dakota, with 61 counties, had 12 and 21 identified clusters, respectively, showing a state with many healthy counties like New York having fewer classified clusters than a state with some healthy but many unhealthy counties like South Dakota. This may also speak to a small population issue, with some small counties in South Dakota having more extreme values. Figure 1 depicts this information, with one bar per cluster, and each cluster's z score along the national distribution illustrating the distribution of health outcomes of the clusters in each state. Figure 3 shows an example of identified clusters within a state, Wisconsin, with 16 clusters. Clusters were visually separated by color and horizontal position, with the length of bar denoting the cluster's z score range and within-state rank range on the right axis. Counties within the same cluster were ordered alphabetically to de-emphasize individual rank point estimates and emphasize the cluster ranks; Supplemental Digital Content Figure 2, available at http://links.lww.com/JPHMP/B414.

At the state level, the number of clusters was strongly associated with the number of counties within a state, as well as the population of the state (R-squared 0.76 and 0.32, P value < .0001 and .023, respectively, data not shown). When cluster results were compared to 2021 CHR data, there were a similar number of overall national clusters (31 in 2021; 30 in 2022), with a correlation coefficient of 96% in county-specific assigned cluster (data not shown); Supplemental Digital Content Table 1, available at http://links.lww.com/JPHMP/B416.

Discussion and Conclusion

It is important for initiatives such as CHR to communicate about the error and reliability around ranks, allowing end users to more meaningfully interpret similarities or differences in health among counties and to better inform population health improvement efforts. Though there is not consensus about how to best approach this goal, data-informed grouping may be a possible avenue for accounting for distance between ranks and rank imprecision, while effectively communicating uncertainty both numerically, with assigned clusters and rank ranges as seen in Figure 3 and Supplemental Digital Content Table 1, available at http://links.lww.com/JPHMP/B416., and visually, with grouped counties and their associated spread across the distribution as seen in Figures 1 and 3. Clustering utilizes the data to create empirically informed breaks in the values, identifying meaningful gaps between counties and natural clusters, with the number and size of clusters being informed by the level of uncertainty in the underlying distribution. These convenient groupings provide a rough sense of the distribution and provide more transparency and statistical context around which counties are not reasonably different from one another. This can result in more responsible data communication to public

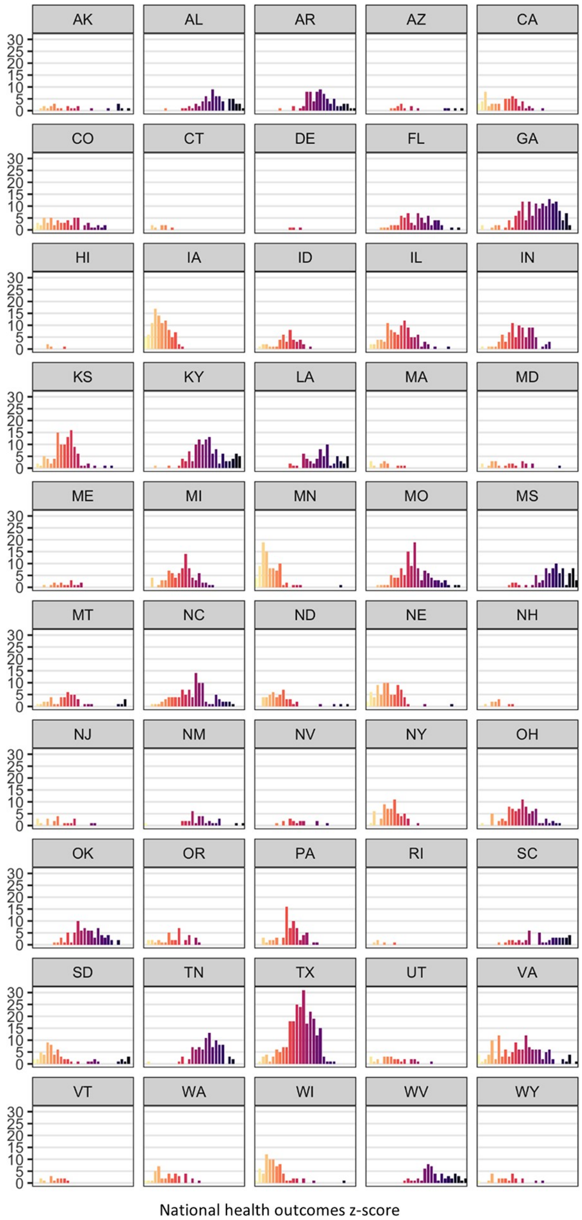

FIGURE 1 Number of Clusters and Counties Within Clusters, by State, Across *z* Score Values+
+Lower value indicates better health outcomes, and higher value indicates worse health outcomes.

health practitioners. While ordinal ranks imply that each of the 3082 counties were different from one another in terms of their health, these findings indicate to users that there were closer to 30 distinct groups of counties, with health experiences being indistinguishable within them.

This clustering technique helps address many of the issues that arise from providing ranks by themselves. The method can be a first step in addressing the imprecision of ranks, accounting for some uncertainty in the underlying distribution and, importantly, communicating that counties within states are not all different from one another. For example, while Douglas County, Wisconsin, may have received an ordinal rank of 47 out of 72 counties, the cluster analysis in Figure 3 shows that Douglas County was not meaningfully or statistically different than Price, Oconto, Monroe, Kenosha, Clark, or Lincoln Counties, and therefore more reasonably had a rank between 44 and 50 out of 72. This impacts how Douglas County would interpret and act on their health status relative to those around them. Douglas County was also clustered with St. Louis County, Minnesota, the county directly across the state border and commonly compared to Douglas in practice, illustrating the added benefit of clustering across state lines. Even for states with very few counties, this approach can offer some utility beyond ordinal ranks. For instance, Hawaii's 4 assigned counties were grouped into 3 clusters, showing that Honolulu County and Maui County have similar health experiences. The approach also showed that Honolulu and Maui, in cluster number 5, were comparable in health to places like Santa Barbara County, California and Fairfax City, Virginia. Kauai County, Hawaii, in cluster 6 with similar health to Los Angeles County, California, was closer in health to Honolulu and Maui than Hawaii County, Hawaii, which sat in cluster 10 and had similar health to Cook County, Illinois (Chicago). Furthermore, though Delaware's 3 counties were assigned 3 different clusters; therefore, having the same grouping as ranks, this method still adds information that Kent County, within cluster 14, was less similar in health than New Castle and Sussex Counties, within clusters 11 and 12, respectively, were to one another. Additional plausible local comparators are also added through this approach for a state with such few counties, including New Castle's comparability in health to the District of Columbia, Sussex's to Cape May County, New Jersey, and Kent's to Caroline County, Maryland, all right across Delaware's border. Figures 1-3 also visually demonstrate these points, with Figures 1 and 3 showing the spread of counties across the *z* score distribution and their distances from one another, and Figure 2 showing comparable counties across state or regional lines.

While the dividing line between clusters may still imply more difference than may be meaningful between 2 counties, this method provides information about the distance between ranks that simple rank numbers alone do not. Furthermore, the well-documented issue of ranks being unable to distinguish counties in the middle of the distribution is accounted for with clustering. While clustering is still best at highlighting the extreme ends of the spectrum,

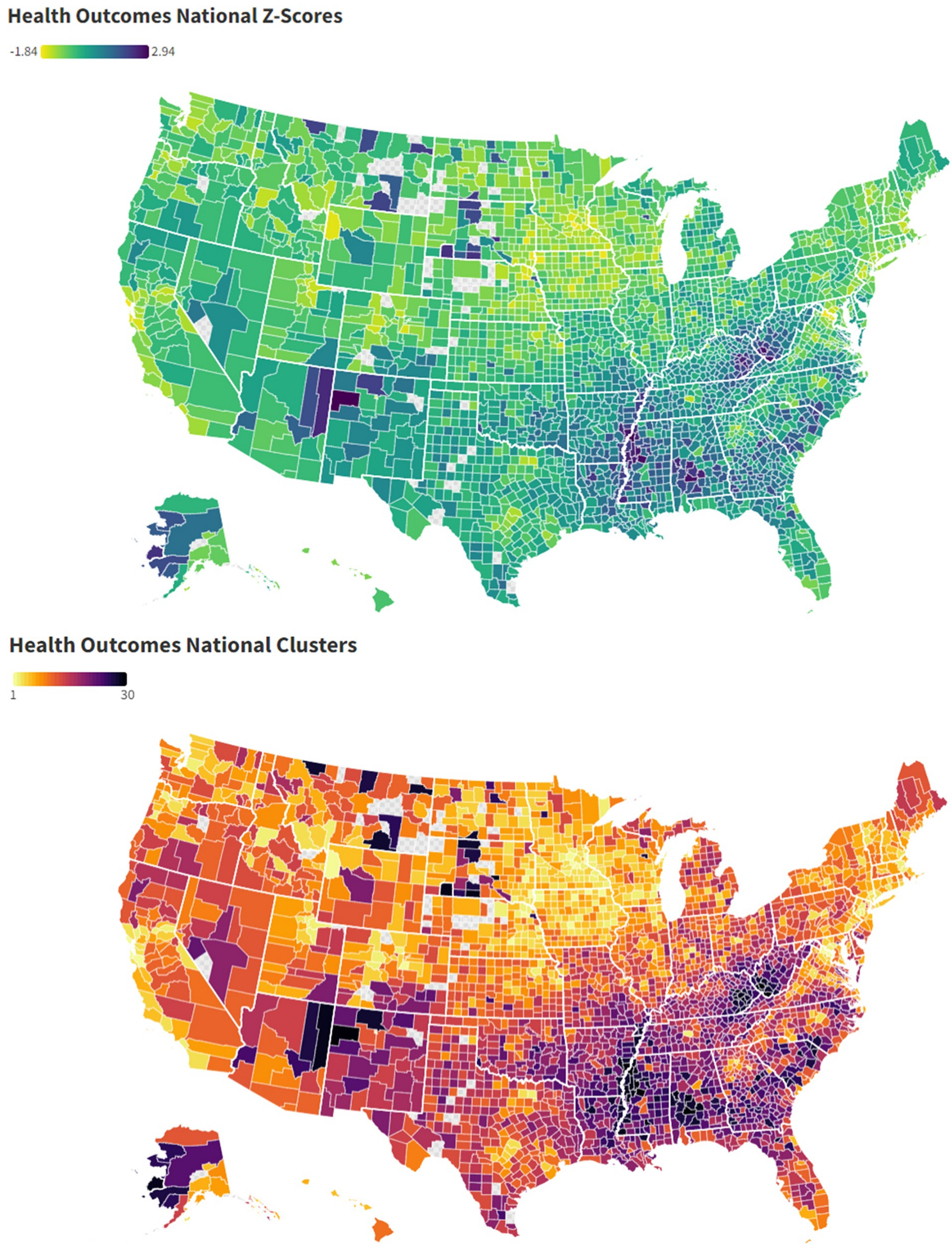

FIGURE 2 Geographic Distribution of (a) the National Health Outcomes *z* Score Values and (b) the 30 National Health Outcomes *z* Score Clusters+
+lower value indicates better health outcomes, and higher value indicates worse health outcomes.

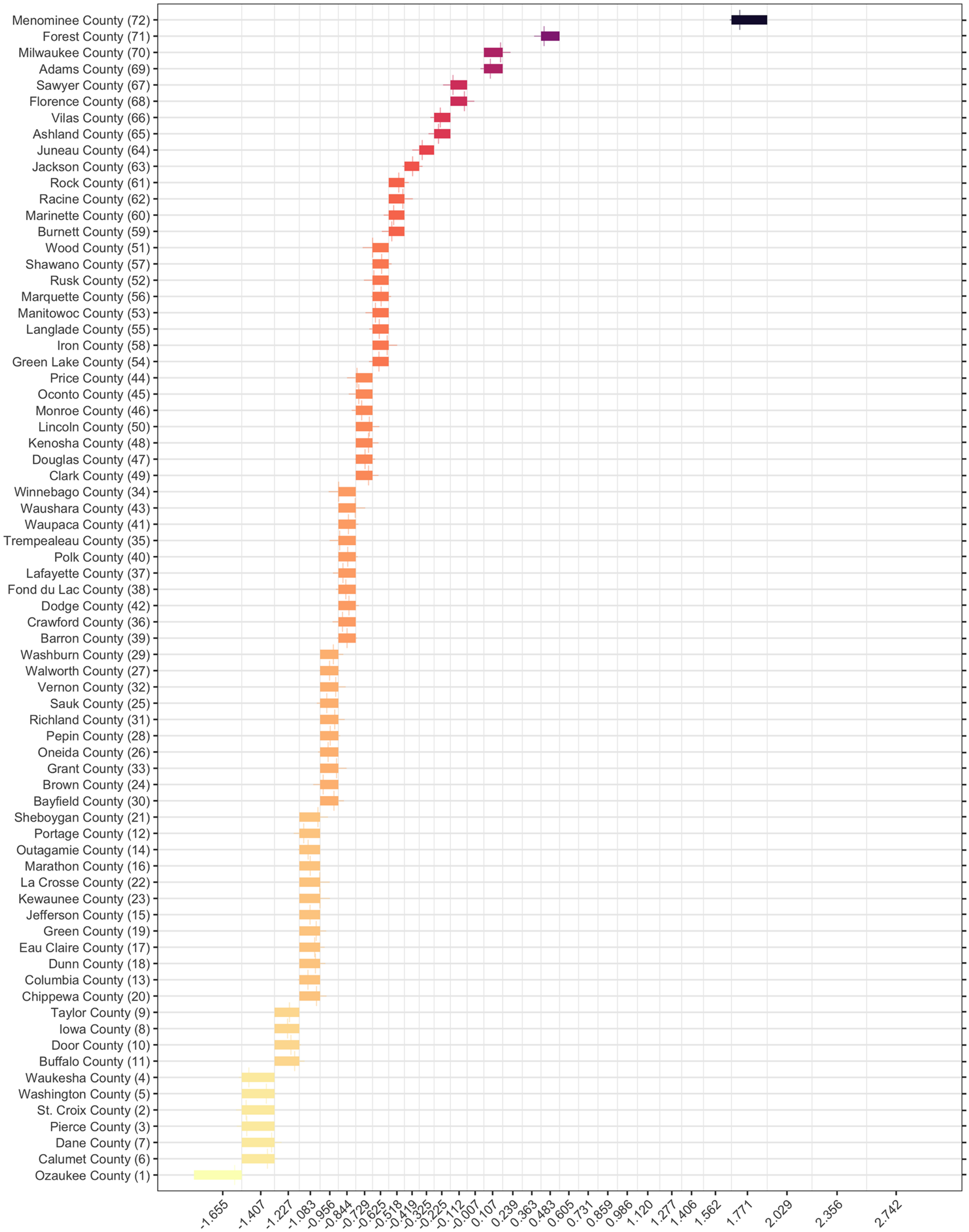

FIGURE 3 Example of Within-State Clusters for Wisconsin, by County, Across *z* Score Values+
+lower value indicates better health outcomes, and higher value indicates worse health outcomes.

showing more unique clusters among both the best- and worst-performing counties, clustering illustrates this "middle-mush" issue through identification of larger clusters covering more counties within states toward the middle. Clustering also provides a tool for comparing county performance with "peer counties," to which community and public health leaders can look for similar experiences or challenges. Using the national distribution of health outcomes provides a higher resolution picture of the groups of counties and extends these possible comparisons to counties across state borders and even regions, extending the possible places to which counties can compare. Finally, cluster analysis may make the comparison of overall health outcomes over time more feasible. Clustering using the national distribution will tend to be more stable year-to-year than looking at a county's single rank value 1 year to the next and may demonstrate more authentic change when counties move between clusters.

While this chosen method of cluster analysis helps to achieve many identified issues pertaining to the imprecision of ranks, there are still some limitations to such an analysis. As mentioned previously, the decision of how many clusters to ultimately use in a cluster analysis is somewhat arbitrary, with many methodologic options from which to choose. Our decision to use Wasserstein (earth mover's) distance was made to use a more data-driven, statistical test to choose the optimal number of clusters. Yet, many other choices for evaluating or choosing the number of clusters could have also been employed, potentially arriving at different results, and other users of cluster analysis are encouraged to tailor their choices to those that best suit their purposes. Furthermore, while this method of clustering accounts for some of the underlying uncertainty in the distribution of county-level health outcomes, some lingering imprecision remains. Error in the individual measures that comprise the health outcomes z score was not accounted for, as this error is incredibly complex to capture, given many of the measures are themselves modeled estimates. Therefore, these clustering methods do not fully capture the imprecision in the estimates and are likely overestimating the distinct number of groups within states. Finally, while this analysis demonstrated the application of this method on a case study of an already-established composite index of CHR's health outcomes, these results, and the accompanying confidence in groupings, are entirely dependent on the selection of measures that underlie the chosen composite. We present just one example of a possible use-case for this method. For instance, basing this analysis on CHR's health factors (a composite z score of 30 measures) would produce a different set of data-informed groupings among counties. Preliminary analyses indicate this application of cluster analysis methods results in similar but distinct clusters and patterns across the country. It is important to consider the many ways counties can be compared, making use of existing contextual factors impacting the health of counties, such as population demographics, socioeconomic characteristics, community resources, and more. Additionally, users could create their own composite indices of other measures specific to their purposes, using summary z scores to calculate relevant groupings based on any number of underlying concepts.

Understanding how to best communicate these clusters will be crucial for their implementation and impact, and future work will focus on this matter. Communication issues to be addressed include whether to replace or complement ranks with these groups and how to best present them. For example, if a county was ranked 10th but assigned to the third cluster in the state spanning ranks 8-14, would that county's cluster be most effectively communicated as a rank range (ie, 8-14), a group number (ie, group no. 3), or a rank tie (ie, rank no. 8 for all counties in that cluster)? Other communication implications include whether to provide both numbers and visual representations of clusters; how to best visualize clusters on graphics and maps; whether to provide interactive options to select different numbers of clusters and see how counties are affected by that choice; and more. These questions could be explored through pilot tests or focus groups to understand the use or utility of these methods for various audiences. Future work will also explore methods for more fully accounting for underlying statistical uncertainty in measures.

Population health rankings can be a powerful catalyst for improving health in communities. However, unintended consequences of presenting ranks alone need to be mitigated by improved communication and tools for users to interpret their rank more meaningfully. Data-informed groupings, such as clusters, can be one such

Implications for Policy & Practice

When health professionals, local community leaders, and the general public look to use population health rankings to make informed decisions about the health of their communities, it is important to consider:

- Rankings are a powerful catalyst for improving health in communities, but unintended consequences of presenting ranks alone should be mitigated by improved communication and tools for users to interpret their rank more meaningfully.
- Data-informed grouping, or clustering, is an effective and simple approach to account for and communicate the uncertainty of ranks, using both numeric and visual displays.
- Public health practitioners can use this information to understand uncertainty in ranks, visualize distances between county ranks, have context around which counties are not meaningfully different from one another, and compare county performance to peer counties.

approach to effectively account for and communicate uncertainty of ranks, both numerically and visually. Community members and public health practitioners can use the information these clusters provide to understand uncertainty in their rank, visualize distance between ranks and distribution of counties, and have context around which counties are similar or reasonably different from one another. Finally, the tool can also be used to compare county performance with peer counties, to which community and public health leaders can look for similar experiences or challenges, even across state borders and regions.

References

1. 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 ManagePract*. 2011;17(5):406-412. doi:10.1097/PHH.0b013e318211b49f.
2. Remington PL, Booske BC. Measuring the health of communities—how and why? *J Public Health ManagePract*. 2011;17(5):397-400. doi:10.1097/PHH.0b013e318222b897.
3. Remington PL, Catlin BB, Gennuso KP. The county health rankings: rationale and methods. *Popul Health Metr*. 2015;13(1):11. doi:10.1186/s12963-015-0044-2.
4. Saisana M, Saltelli A, Tarantola S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J R Stat Soc Ser A Stat Soc: Series A (Statistics in Society)*. 2005;168(2):307-323. doi:10.1111/j.1467-985X.2005.00350.x.
5. Oliver TR. population health rankings as policy indicators and performance measures. *Prev Chronic Dis*. 2010;7(5):A101. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938395/. Accessed June 10, 2021.
6. Athens JK, Catlin BB, Remington PL, Gangnon RE. Using empirical Bayes methods to rank counties on population health measures. *Prev Chronic Dis*. 2013;10:E129. doi:10.5888/PCD10.130028.
7. Saisana M, d'Hombres B, A S. Rickety numbers: volatility of university rankings and policy implications. *Res Policy*. 2011; 40(1):165-177. doi:10.1016/j.respol.2010.09.003.
8. Kanarek N, Tsai HL, Stanley J. Health ranking of the largest US counties using the community health status indicators peer strata and database. *J Public Health ManagePract*. 2011;17(5):401-405. doi:10.1097/PHH.0b013e318205413c.
9. Hendryx M, Ahern MM, Zullig KJ. Improving the environmental quality component of the county health rankings model. *Am J Public Health*. 2013;103(4):727-732. doi:10.2105/AJPH.2012.301016.
10. Arndt S, Acion L, Caspers K, Blood P. how reliable are county and regional health rankings? *Prev Sci*. 2013;14(5):497-502. doi:10.1007/s11121-012-0320-3.
11. Arndt S. Just how useful are health rankings? *Health Serv Res*. 2015;50(5):1403-1406. doi:10.1111/1475-6773.12350.
12. Waldrop AR, Moss JL, Liu B, Zhu L. Ranking states on coverage of cancer-preventing vaccines among adolescents: the influence of imprecision. *Public Health Rep*. 2017;132(6):627-636. doi:10.1177/0033354917727274.
13. Purtle J, Peters R, Kolker J, Diez Roux AV. Uses of population health rankings in local policy contexts: a multisite case study. *Med Care Res Rev*. 2019;76(4):478-496. doi:10.1177/1077558717726115.
14. Jewett PI, Zhu L, Huang B, Feuer EJ, Gangnon RE. Optimal Bayesian point estimates and credible intervals for ranking with application to county health indices. *Stat Methods Med Res*. 2019;28(9):2876-2891. doi:10.1177/0962280218790104.
15. Athens JK, Remington PL, Gangnon RE. Improving the rank precision of population health measures for small areas with longitudinal and joint outcome models. *PLOS ONE*. 2015;10(6):e0130027. doi:10.1371/journal.pone.0130027.
16. Hood CM, Gennuso KP, Swain GR, Catlin BB. County health rankings. *Am J Prev Med*. 2016;50(2):129-135. doi:10.1016/j.amepre.2015.08.024.
17. Arndt S, Acion L, Caspers K, Diallo O. Assessing community variation and randomness in public health indicators. *Popul Health Metr*. 2011;9(1):3. doi:10.1186/1478-7954-9-3.
18. Zhang S, Luo J, Zhu L, et al. Confidence intervals for ranks of age-adjusted rates across states or counties. *Stat Med*. 2014;33(11):1853-1866. doi:10.1002/sim.6071.
19. AM VDB, S VDL, Freeman ALJ, et al. Communicating uncertainty about facts, numbers and science. *R Soc Open Sci*. 2019;6(5):181870. doi:10.1098/rsos.181870.
20. Bisantz AM, Marsiglio SS, Munch J. Displaying uncertainty: investigating the Effects of Display Format and Specificity. *Hum Factors*. 2005;47(4):777-796. doi:10.1518/001872005775570916.
21. Correll M, Gleicher M. Error bars considered harmful: exploring alternate encodings for mean and error. *IEEE Trans Vis Comput Graph*. 2014;20(12):2142-2151. doi:10.1109/TVCG.2014.2346298.
22. Jackson CH. displaying uncertainty with shading. *Am Statistician*. 2008;62(4):340-347. doi:10.1198/000313008X370843.
23. Courtemanche C, Soneji S, Tchernis R. Modeling area-level health rankings. *Health Serv Res*. 2015;50(5):1413-1431. doi:10.1111/1475-6773.12352.
24. Davis W, Gordan A, Tchernis R. Measuring the spatial distribution of health rankings in the United States. *Health Econ*. 2021;30(11):2921-2936. doi:10.1002/hec.4416.
25. Wallace M, Sharfstein JM, Kaminsky J, Lessler J. Comparison of US county-level public health performance rankings with county cluster and national rankings: assessment based on prevalence rates of smoking and obesity and motor vehicle crash death rates. *JAMA Network Open*. 2019;2(1):e186816-e186816. doi:10.1001/jamanetworkopen.2018.6816.
26. Hazelkorn E, Gibson A. Global science, national research, and the question of university rankings. *Palgrave Commun*. 2017;3(1):1-11. doi:10.1057/s41599-017-0011-6.
27. Bisantz AM, Cao D, Jenkins M, et al. Comparing uncertainty visualizations for a dynamic decision-making task. *J Cogn Eng Decis Mak*. 2011;5(3):277-293. doi:10.1177/1555343411415793.
28. Belia S, Fidler F, Williams J, Cumming G. Researchers misunderstand confidence intervals and standard error bars. *Psychol Methods*. 2005;10(4):389-396. doi:10.1037/1082-989X.10.4.389.
29. Kumar M, Patel NR. Clustering data with measurement errors. *Comput Stat Data Anal*. 2007;51(12):6084-6101. doi:10.1016/j.csda.2006.12.012.
30. Wang H, M S. Ckmeans.1d.dp: optimal k-means clustering in one dimension by dynamic programming. *R J*. 2011;3(2):29-33. doi:10.32614/RJ-2011-015.
31. Schuhmacher D, diagrams BB (aha and power), Bonneel (networkflow) N, et al. transport: computation of optimal transport plans and Wasserstein distances. https://CRAN.R-project.org/package=transport. Published online August 12, 2022. Accessed October 17, 2022.
32. Team RC. R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria.. http://www.R-project.org/. https://cir.nii.ac.jp/crid/1574231874043578752. Published online 2016. Accessed October 17, 2022.