Methods for Summarizing Data

by Deborah Rosenberg, PhD

In public health, multiple population groups, multiple risk factors, and multiple outcomes are often simultaneously of interest. In addition, data analysts must often address the domains of health services and health systems as well as health status. For maternal and child health professionals, census data, vital records data, Medicaid data, hospital discharge data, WIC data, client tracking system data, focus group and other qualitative data, and national and/or local sample survey data are all relevant for describing and monitoring the health of women, infants, children, adolescents, and families. The challenge for MCH data analysts is to summarize this array of data in ways that facilitate carrying out the core functions of assessment, assurance, and policy development.

The process of summarizing data within and across many domains and population groups can seem daunting unless a well defined analysis plan is articulated and implemented. Developing an analysis plan is an essential step in information-based decision making. The aim is to design a systematic approach for reducing the data burden for a given MCH analysis or report. Without such an approach, it is likely that data will not be successfully translated into the information needed by program planners, managers, and policy-makers.

DEVELOPING AN ANALYSIS PLAN

Making analytic choices that result in a cogent portrayal of health status, services, and the relationship between them is as much an art as a science. As a general rule, data are more reliable and easier to interpret as the degree of summarization increases but, conversely, data are more targeted and specific as the degree of summarization decreases. The objective is to weigh the advantages and disadvantages of different strategies and choose the one that strikes the best balance between specificity and interpretability.

Targeted Specific	Reliable Easy to Interpret
← ← ← ← ← ← ↔	→ → → → → →
Unsummarized	Summarized

Data summarization strategies can reduce the data burden either directly by limiting the amount of data reported, or indirectly by increasing its interpretability. Often, both approaches are used in combination.

In order for data summarization to be effective, the analysis plan should be framed according to each of the following:

- Purpose of the analysis
- Audience for the analysis
- Data availability and data quality

Each of these is discussed below.

Purpose of the Analysis

Routine surveillance and monitoring, statewide or community based needs assessment, quality assurance activities, program evaluation, and special studies each require development of a distinct analysis strategy. As a starting point, each analysis should be situated within a matrix of pertinent population groups and domains of interest:

Population Groups	**Domains**		
	Health Status	Health Services	Health Systems
Women			
Infants			
Children			
Adolescents			
Families			

One analysis might be situated in just one cell of the matrix, while another might be situated in all of the cells.

For a comprehensive statewide MCH needs assessment, for example, it is likely that all of the cells of the matrix are appropriate. Many indicators will be of interest as well as comparisons across person, place, time, and risk. In such a wide-ranging analysis, data summarization will be necessary both in terms of restricting the amount of data and using methods that increase interpretability. A certain amount of detail may have to be sacrificed as a safeguard against losing sight of the major findings of the needs assessment.

For community-based needs assessment, all of the cells of the matrix may again be appropriate and many indicators will again be of interest. The data burden, however, will be lighter because the focus is on each area separately rather than on comparisons across areas. Therefore, the degree of data summarization may not have to be as great as for a statewide analysis.

In contrast to needs assessment, intensive study of a particular health problem may situate an analysis in only one cell of the matrix, and only a few indicators will be of interest as well. In this circumstance, it is possible and desirable to disaggregate (leave unsummarized) data for many person, place, time and risk variables. The goal here is to explore one issue in great detail.

For program evaluation, an analysis will, by definition, cross at least the two domains of health status and health services, though often only one population group will be pertinent. Identifying improvement or deterioration in health status over time will be the primary focus, and therefore, data summarization will likely be necessary for other dimensions such as socio-demographic characteristics or geography.

These brief examples illustrate the linkage between data summarization and the purpose of an analysis. You can see that how widely or narrowly the analytic net is cast, whether many or only a few indicators are of interest, and the particular questions being asked, all govern the data summarization process.

The Audience for the Analysis

A particular analysis and report may be intended for use by the professionals working in state or local health agencies; it may be intended for legislators; it may be intended for stakeholders in communities; it may be intended for the general public. The degree and type of data summarization must be matched to the needs of each group. The general public may only need to see statewide figures, whereas legislators may want information stratified by county, and community leaders may want it stratified by even smaller units of geography. One audience may want a comprehensive report, another may want a sound bite. The MCH data analyst should be able to produce quality information for each of these circumstances.

Regardless of the audience, though, the data analyst should always examine the data in its most unsummarized form. If decisions about how much summarization, and on which variables, are made without first looking at the "raw" data, important differences and disparities might be overlooked. For example, a decision might be made to present an indicator for children ages 0-17 stratified by race/ethnicity, when in fact, variation by age is greater than variation across race/ethnic groups. In light of this, a more appropriate data summarization decision would have been to present the data stratified by age and not by race/ethnicity. This would have been clear had the data been examined in their most refined form.

Decisions about how to summarize data, then, should rely less on untested assumptions and more on empirical findings. Depending on the audience, the analyst has to make choices about the degree of summarization, but these choices should be guided by a thorough understanding of the loss of information entailed in each decision.

Availability of Data and Data Quality

Choices about which data to include in an analysis are highly dependent on both the availability of data and on its quality. As MCH professionals, we can all name many indicators that would enhance our ability to carry out the core functions of public health, but for which data are not routinely collected. For example, it would be very informative to report data on the incidence of child injury by type of injury and age. Since incidence data are not readily available, however, hospitalization for injury or injury mortality are typically reported instead, despite the limitations this imposes on our analysis.

Similarly, although it is well accepted that analyzing the content or quality of prenatal care is critical to understanding the effectiveness of this service, generally only measures of the quantity of prenatal care received, such as the timing of the first visit and the total number of visits, are available for inclusion in analysis.

Data quality also limits our analytic choices. For example, gestational age is a critical variable for understanding perinatal morbidity and mortality, but birthweight is more commonly chosen as an indicator because it has been shown to be more accurately documented.

How we summarize data is affected by data availability and data quality. Sometimes we have to present more data than we would like because we need to use several less than optimal measures in an attempt to approximate the information contained in an indicator that is itself unavailable. Sometimes we use data of poor quality because no high quality alternative is available, and perhaps most problematic, sometimes we mis-specify the questions we ask because of the constraints in the data.

Variables, Methods, and Presentation

Framed by the purpose of the analysis, the audience for which it is intended, and the availability and quality of data, the specifics of analysis planning can proceed. The opportunity for data summarization occurs in three successive phases:

Phase I: Selection of variables. This phase includes the selection and definition of the primary indicator or indicators that will be the central focus of the analysis. While the formulation of some MCH indicators, such as the infant mortality rate, are well established, defining numerators and denominators for other indicators is a crucial part of the data summarization process. Similarly, selection and definition of the person, place, time, and risk variables that will be used to refine the analysis of the primary indicator(s) occurs in this phase.

Phase II: Selection of analytic methods. This phase involves decisions about how the indicators and other variables selected will be examined. For example, some indicators might be presented as counts and some as rates; some might be presented as overall averages while others might be stratified by person, place, time, or levels of risk; some indicators might be combined into a composite index; some indicators might be presented in their original form while others might be transformed into categories, ranks, or scores. In addition, comparisons might be made intuitively, or formal statistical testing might be conducted.

Phase III: Selection of presentation format. This phase involves designing a report that effectively communicates the results of the analysis. Written narrative, tables, charts, graphs, and maps are each effective formats depending on the type of data being presented. Organization of each, including defining sections of text, choosing chart types (pie, bar, line, etc.), defining rows and columns for tables, horizontal and vertical (x and y) axes for graphs, and deciding on map layers are important aspects of data summarization.

In the variable selection phase (Phase I), data summarization choices are made to restrict the actual amount of data to be analyzed and reported. The number of indicators can be restricted as well as the number of person, place, time, and risk strata. For instance, from a pool of 50 potential indicators of interest, it might be decided to include only 10 in a given report. Similarly, from a pool of many potential person, place, time, and risk variables, only a few might be selected. The drive to summarize data dictates the exclusion of many variables, and so those that are selected must, if possible, be broadly representative of those that are not.

In the analytic methods phase (Phase II), data summarization choices are made to both restrict the amount of data and to increase its interpretability. Restricting the amount of data is accomplished by limiting the degree of stratification for each of the indicators, drawing from the list of the other selected variables. Possibilities for reporting each indicator might include:

- By age, race/ethnicity, county, and year
- By age, race/ethnicity, and county
- By age, and race/ethnicity
- By age
- Overall for the state for one year

Moving from the most refined to the more summarized levels, a minimal loss of information is assumed (or preferably verified with preliminary analyses). For instance, if you choose to report an indicator without stratifying by county, the assumption is that the indicator does not vary by county in a way that is meaningful, either epidemiologically or from a program or policy perspective.

The aim is to combine only as many strata as necessary to gain reliability in terms of sample size, and ease of interpretation in terms of reducing the amount of data the audience needs to digest. Aggregating across too many strata may obscure meaningful differences, and although the audience will not be overwhelmed, neither will it have useful information. This principle holds whether data will be summarized across person, place, time, or across risk. Over-summarization of socio-demographic characteristics and risk can obscure differences that have implications for program design, and in combination with geography, may have implications for resource allocation. Over-summarization across time can mask patterns of improvement or deterioration, and conceal the impact of critical events such as implementation of new programs or policies.

One strategy for reducing the amount of data reported without sacrificing too much information is to let person, place, time, and risk variables operate jointly. For instance, data can be collapsed in one of these dimensions, and the danger of obscuring meaningful differences minimized if related data from the other dimensions drive the process. To illustrate, if geographic areas are to be combined, some possible approaches are:

- Combine areas according to levels of risk—for example, combine those areas with similar low birthweight rates, or similar childhood immunization rates.

- Combine by socio-demographic factors—for example, combine those areas with similar racial/ethnic composition, similar income, or similar educational levels.

- Combine by level of services/programs in the area—for example, combine those areas with similar numbers of providers per capita, or similar numbers of WIC sites.

Instead of restricting the amount of stratification, another strategy that restricts the amount of data is to report only values that highlight extremes, disparities, or changes. For example, if indicators are being examined by geographic area, only the two areas with the highest rates and the two areas with the lowest rates on each indicator might be reported. Or, only indicators with a three-fold difference across areas might be reported. Attention might be focused on those geographic areas within which there are large disparities between demographically defined population groups. If trends over time are of interest, perhaps only areas that show a 10% improvement or deterioration over time on an indicator might be reported.

Other data summarization techniques used in the methods phase (Phase II) leave the actual amount of data reported unchanged, but reduce the data burden by transforming the data in ways that make it easier for an audience to assimilate. For example, grouping indicator values into discrete categories alleviates the data burden by replacing many distinct values with a few summary ones. Ranking and scoring methods, like categorization, increase the interpretability of data without reducing the amount of data reported. Unlike categorization, though, these methods do not even reduce the number of distinct values, but the ordering and labeling inherent in the transformed values (such as integer ranks) adds information and meaning that the original values could not convey.

Categorization, ranking, and scoring enhance the reader's ability to synthesize information.
Care must be taken, however, not to over-use these methods. Report cards, for example, while they are easy to understand, often suffer from over-summarization, yielding a view of a health issue that is too rudimentary to be useful from a program planning or policy perspective.

Index construction is a hybrid of restricting the amount of data and enhancement of data interpretability. It permits two or more indicators to be represented by one composite measure. The number of data values reported is reduced, but rather than excluding indicators from the analysis, the information they contain is merely reflected in a somewhat different form.

In the presentation phase (Phase III), data summarization choices are also made to facilitate interpretability. In general, visual methods such as graphs and maps handle data from many sources and of different types in an integrated and comprehensible form. Suppose, for example, a state has 64 counties. Whereas it may be difficult to scan 64 numbers in a table and make judgements about their relative standing, looking across the heights of bars in a bar chart may be much easier. Text and tables, however, while they may require more effort to absorb, do provide more detailed and specific information.

It is important not to ignore or underestimate the role of text in the data summarization process. Data may have been selected appropriately in Phase I, creative and useful methods applied in Phase II, and effective tables, charts, graphs and maps created to present it in a clear and concise fashion in Phase III, but data summarization is incomplete without explicit narrative interpretation. The data alone cannot fully communicate a message. MCH professionals must take on the responsibility for carrying the data summarization process through to its conclusion by writing text that gives shape to the story that the data only outlines.

The table below reviews the three phases of analysis planning and data summarization:

<table>
<tr><th colspan="3">Summarizing Data: Reducing the Data Burden</th></tr>
<tr><td colspan="2">Restricting the Amount of Data</td><td>Increasing Interpretability</td></tr>
<tr><td>Phase I: Variables</td><td>Phase II: Methods</td><td>Phase III: Presentation</td></tr>
<tr><td>Limit the number of indicators

Limit the number of person, place, time, and risk variables</td><td>Limit the amount of stratification

Transform variables into:
Discrete categories
Ranks
Scores

Construct indices

Use statistical testing</td><td>Text

Tables

Charts

Graphs

Maps</td></tr>
</table>

METHODS FOR INCREASING INTERPRETABILITY

Categorization

We saw earlier that categorization can reduce the data burden by condensing many values into a few summary categories. In addition, when multiple indicators are being analyzed, imposing common labels helps to integrate information across the various measures. To illustrate, the first of the two tables below shows the low birthweight rates and child injury mortality rates for 10 counties, and the second of the two tables shows the same data in categorized values. The counties are sorted by the low birthweight rates, and categories are defined as above or below the median on each indicator.

Original Values		
County	Low Birthweight Rate (%)	Child Injury Mortality Rate (per 100,000)
A	4.17	46.60
B	5.68	45.94
C	6.08	47.69
D	6.41	36.55
E	6.49	81.50
F	6.96	31.38
G	7.20	42.83
H	7.75	39.36
I	8.44	30.79
J	21.05	52.64

Summary Categories: Above (+) and Below (-) the Median		
County	Low Birthweight Rate (%)	Child Injury Mortality Rate (per 100,000)
A	-	+
B	-	+
C	-	+
D	-	-
E	-	+
F	+	-
G	+	-
H	+	-
I	+	-
J	+	+

You can see that it is easier to interpret the second table. For example, overall, it does not appear that a high value on one health status indicator necessarily means a high value on the other. Specifically, though, the second table shows that County J is high on both measures while County D is low on both. The distinct values in the first table made it difficult to see this pattern.

In addition to imposing common values, categorizing with intuitive labels is also valuable when the meaning of indicator values is unclear. While some MCH indicators, such as infant mortality, have values that most audiences understand, many indicators do not have values that are so readily interpretable in their original form. For example, many people know without the aid of a label that an infant mortality rate of 15 per 1,000 live births is unacceptably high, but this is not the typical situation. Particularly when there is no standard or goal, such as a national objective, that has been identified and publicized, the meaning of indicator values is not clear. Translating indicator values into intuitive labels such as the following helps the audience find the meaning:

- high, medium, low
- above average, below average
- excellent, good, fair, poor

When possible, a benchmark or external standard, such as a national objective, determines the categorization process. When no such external reference exists, data are categorized according to their relative position in the observed distribution. In "Descriptive Epidemiology and Statistical Estimation", (Module I) we described in some detail various methods for categorizing data, such as defining boundaries such that there are equal numbers of observations in each category or such that there are equal portions of the range of values in each category; we also demonstrated that natural breakpoints or clustering in the data may define category boundaries, and that statistics such as the mean, median, or standard deviation may be used as well.

In the above table (previous page), the median of the observed values was used to define two categories for the 10 low birthweight rates and child injury mortality rates. Following are two approaches to assigning the 10 county low birthweight rates to four categories. As you look at the data again, notice the extreme gap between the highest rate of 21.05 in County J and the next highest rate of 8.44 in County I. Also notice that only County A with its low birthweight rate of 4.17 has met the Year 2000 Objective of 5 low birthweight births in every 100 live births.

2 Possible Ways to Categorize 10 Low Birthweight Rates

	Method I		Method II	
Category	County	Rate (%)	County	Rate (%)
1	A	4.17	A	4.17
	B	5.68		
			B	5.68
	C	6.08	C	6.08
2	D	6.41	D	6.41
	E	6.49	E	6.49
	F	6.96	F	6.96
3	G	7.20	G	7.20
	H	7.75	H	7.75
			I	8.44
	I	8.44		
4	J	21.05	J	21.05

To the extent possible, Method I divides the 10 observations equally into the four categories. This results in the rates of 8.44 and 21.05 being in the same group, a questionable approach. Method II uses the Year 2000 Objective of 5 % low birthweight births to define the boundary between categories 1 and 2, and also isolates the highest rate of 21.05 because of its extreme value.

Regardless of the chosen breakpoints, defining discrete categories results in loss of information. The values 1-4 cannot possibly be as descriptive as the ten original values for each county.

In either method, for example, the rates of 6.08, 6.41 and 6.49 in Counties C, D and E are grouped together, implying that they are the same. Assessment of an acceptable loss of detail should determine the number of categories to use for a particular indicator; as we've already seen, loss of information can be offset by the gain in interpretability that categorization provides.

Comment

The level of the data will influence the choice of how to define discrete categories. Categorizing a set of birthweight values for individual infants, for example, will require a different strategy than categorizing a set of aggregate low birthweight rates for counties such as that shown in the above examples. For individual level birthweight data, the clinically relevant categories of very low birthweight (< 1500 grams), moderately low birthweight (1500-2499 grams), and "normal" birthweight (>= 2500 grams) are typically used.

Test Yourself

Question: Below are the sorted values of the child injury mortality rates for the 10 counties. The overall child injury mortality rate for the 10 counties combined is 42.77 per 100,000 children. Describe a possible categorization strategy.

County	Child Injury Mortality Rate (per 100,000)
I	30.79
F	31.38
D	36.55
H	39.36
G	42.83
B	45.94
A	46.60
C	47.69
J	52.64
E	81.50

Answer: There are many possible categorization strategies for this indicator. One might be as follows:

Category	County	Child Injury Mortality Rate (per 100,000)
1	I	30.79
	F	31.38
2	D	36.55
	H	39.36
3	G	42.83
	B	45.94
	A	46.60
	C	47.69
4	J	52.64
5	E	81.50

This strategy highlights both low and high rates, including emphasizing the potential significance of the two highest rates of 52.64 and 81.50. The overall rate of 42.77 is used to define the breakpoint between the two middle categories.

While the common labels that discrete categories impose, such as "high" and "low", increase the comparability across indicators, they do so in a fairly crude fashion. A more refined transformation is required if precise comparisons are desired. Here again are the low birthweight rate and the child injury mortality rate for County J along with the median on each indicator for the 10 counties:

	Low Birthweight Rate (%)	Child Injury Mortality Rate (per 100,000)
County J	21.05	52.64
Median of the 10 Counties	6.73	44.38

Earlier we saw that both of County J's rates were labeled "+", indicating that each was above its respective median, but although we know they are both high, it is impossible to judge which of County J's indicators is actually "farther" from the median. Is a 14 % difference (21.05 v. 6.73) better or worse than a 8 per 100,000 difference (52.64 v. 44.38)?

It is difficult to achieve precision with a few discrete categories, and therefore, it may be preferable to transform the original values to a common metric or scale using ordinal or continuous (uncategorized) measures such as integer ranks or z-scores. These methods transform the original values such that there is little or no loss of information—the same number of values exist after the transformation as existed before. For example, the 10 indicator values for the 10 counties can be transformed into 10 ranks or 10 z-scores.

Whereas a category label such as "high" or "+" might have a somewhat different meaning from one indicator to the next, an integer rank such as "1" or a score such as "1.3" will have the same meaning across indicators. Some ranking and scoring methods are described below.

Integer Ranking

Simple ranking assigns integers to the sorted values of the indicator of interest. For instance, if 10 counties are to be ranked according to the percent of children living in poverty, the integers 1 through 10 would be assigned. By definition, the differences between the ranks are uniform: the distance between rank 2 and rank 4 is equal to the distance between all other values that are two ranks apart.

Integer ranking can be applied to individual or aggregate level data, although it is most often used for aggregate values. Here are the 10 county low birthweight rates transformed into 10 integer ranks:

Integer Ranking of 10 Low Birthweight Rates		
County	%	Rank
A	4.17	1
B	5.68	2
C	6.08	3
D	6.41	4
E	6.49	5
F	6.96	6
G	7.20	7
H	7.75	8
I	8.44	9
J	21.05	10

Continuous scoring methods such as percentile rescaling, z-scores, and z-tests also assign ordered values to the original data, but unlike integer ranking these values are not uniformly spaced. Instead, scoring methods assign new values that preserve the original distribution of the data, leaving the relative distances between values unchanged.

Percentile Rescaling

The goal of percentile rescaling is to translate the position of a value on one scale to exactly the same position on another. For example, this approach will translate a value at the 25th percentile along the original distribution of values, and find a value at the 25th percentile along a new range. This transformation can be accomplished in two steps:

First, let's examine the formula for finding the percentile of a known value:

$$\text{Percentile} = \frac{\text{Original Value} - \text{Lowest Original Value}}{\text{Highest Original Value} - \text{Lowest Original Value}} \times 100$$

The numerator in this equation is a measure of how far along the original range a given value is—its distance from the lowest value on this range. The denominator is a measure of the entire original range—the distance from the lowest to the highest value. The ratio of these two, then, is the percentile of the value of interest.

Returning to the example of the 10 county low birthweight rates, remember that the range was from the lowest rate of 4.17% in County A to the highest rate of 21.05% in County J. Let's consider County C's rate of 6.08 and determine its position along this range. All of the terms on the right-hand side of the equation are known, and we solve for the unknown percentile:
:

$$\text{Percentile} = \frac{6.08 - 4.17}{21.05 - 4.17} \times 100$$

$$= \frac{1.91}{16.88} \times 100$$

$$= 11.3$$

County C's rate of 6.08 is a distance of 1.91 beyond the lowest value of 4.17, and 1.91 is 11.3% along the whole range of 16.88.

Second, to place 6.08 on a scale with a minimum value of 1 and a maximum value of 10, for example, we need to find the value located at 11.3% along this new range. The calculation is set up the same as before except that the equation is now written in terms of the new values rather than the original ones:

$$\text{Percentile} = \frac{\text{New Value} - \text{Lowest New Value}}{\text{Highest New Value} - \text{Lowest New Value}} \times 100$$

This time, the value of the percentile is known, along with the highest and lowest values of the new range, and we solve for the unknown new value. For County C's low birthweight rate, we have:

$$11.3 = \frac{\text{New Value} - 1}{10 - 1} \times 100$$

Then, the equation is algebraically reorganized to solve for the new value:

$$\text{New Value} = \frac{(10-1) \times 11.3}{100} + 1$$

$$= \frac{9 \times 11.3}{100} + 1$$

$$= 2.02$$

This new value of 2.02 is at exactly the same point along the range of 1 to 10, as is the original value of 6.08 along the range of 4.17 to 21.05, as is 11.3 along the range of 0 to 100. Similarly, the values of 5.5, 12.61, and 50 are at the same point (50th percentile). The equivalence across the three ranges is illustrated below:

Equivalent Values on Three Different Scales

	*		
1.00	2.02	5.5	10.00

	*		
4.17	6.08	12.61	21.05

	*		
0.00	11.30	50	100.00

Here are the results of applying the percentile rescaling process to all 10 of the county low birthweight rates :

Integer Ranking and Percentile Rescaling of 10 Low Birthweight Rates			
County	Original Values (%)	Rank	New Scale 1-10
A	4.17	1	1.00
B	5.68	2	1.80
C	6.08	3	2.02
D	6.41	4	2.19
E	6.49	5	2.24
F	6.96	6	2.49
G	7.20	7	2.62
H	7.75	8	2.91
I	8.44	9	3.28
J	21.05	10	10.00

You can see that the percentile rescaled values are in the same order as the values of the original data. Unlike simple integer ranking, however, the distances between values are not uniform; the distances between the rescaled values mirror their relative distances in the original data. For example, in the above table, the distance between the integer ranks for Counties B and D is, of course, 2 (4-2), while the distance between the rescaled values for these two counties is only 0.39 (2.19-1.80). Conversely, the percentile rescaling is able to preserve the extreme gap between the highest rate of 21.05 and the next highest of 8.44 with a distance of 6.72 (10.0-3.28), while the integer rank difference is only 1 (10-9).

With percentile rescaling, a common range is chosen for all indicators. In the example for the 10 county data, the lowest value (at 0% along the range) of any indicator would be assigned the new value of "1" and the highest value (at 100% along the range) of any indicator would be assigned a new value of "10". All of the intermediate values would likewise be assigned equivalent values—a value at the 11.3th percentile *on any indicator* would be assigned a new value of 2.02 as was the low birthweight rate for County C.

This form of rescaling can be applied to individual or aggregate level data, although it is typically used with aggregate values.

Comment and Example

The rescaling process can be carried out in just one step, by solving the equations for the original values and the new values simultaneously as follows:

$$\frac{\text{New Value} - \text{New Low}}{\text{New High} - \text{New Low}} = \frac{\text{Original Value} - \text{Original Low}}{\text{Original High} - \text{Original Low}}$$

$$\text{New Value} - \text{New Low} = \frac{\text{Original Value} - \text{Original Low}}{\text{Original High} - \text{Original Low}} \times \text{New High} - \text{New Low}$$

$$\text{New Value} = \left(\frac{\text{Original Value} - \text{Original Low}}{\text{Original High} - \text{Original Low}} \times \text{New High} - \text{New Low} \right) + \text{New Low}$$

In this formulation, County C's low birthweight rate is rescaled as follows:

County C

$$\frac{\text{New Value} - 1}{10 - 1} = \frac{6.08 - 4.17}{21.05 - 4.17}$$

$$\text{New Value} - 1 = \frac{1.91}{16.88} \times 9$$

$$\text{New Value} = (0.113 \times 9) + 1$$

$$= 2.02$$

Consider a few additional examples of using this formulation, beginning with the two endpoints of the range—County A and County J with low birthweight rates of 4.17 and 21.05 respectively:

County A

$$\frac{\text{New Value} - 1}{10 - 1} = \frac{4.17 - 4.17}{21.05 - 4.17}$$

$$\text{New Value} - 1 = \frac{0}{16.88} \times 9$$

$$\text{New Value} = (0 \times 9) + 1$$

$$= 1$$

County J

$$\frac{\text{New Value} - 1}{10 - 1} = \frac{21.05 - 4.17}{21.05 - 4.17}$$

$$\text{New Value} - 1 = \frac{16.88}{16.88} \times 9$$

$$\text{New Value} = (1 \times 9) + 1$$

$$= 10$$

And, here is percentile rescaled value for County H with a low birthweight rate of 7.75:

County H

$$\frac{\text{New Value} - 1}{10 - 1} = \frac{7.75 - 4.17}{21.05 - 4.17}$$

$$\text{New Value} - 1 = \frac{3.58}{16.88} \times 9$$

$$\text{New Value} = (0.212 \times 9) + 1$$

$$= 2.91$$

Test Yourself

Question: For the 10 county data, transform County C's child injury mortality rate of 47.69 to its equivalent percentile position on a scale of 1-10.

Answer: The rescaled value for County C's child injury mortality rate is as follows:

County C

$$\frac{\text{New Value} - 1}{10 - 1} = \frac{47.69 - 30.79}{81.50 - 30.79}$$

$$\text{New Value} - 1 = \frac{16.90}{50.71} \times 9$$

$$\text{New Value} = (0.333 \times 9) + 1$$

$$= 4.00$$

Question: Compare the percentile rescaling results for County C's low birthweight rate and child injury mortality rate with the results from using the categories above and below the median, and integer ranking.

Answer: Here are the results for County C using the three methods discussed so far:

County C	Low Birthweight Rate = 6.08 %	Child Injury Mortality Rate = 47.69 per 100,000
Above (+) and Below (-) the Median	-	+
Integer Ranks	3	8
Percentile Rescaled Values	2.02	4.00

County C's low birthweight rate is below the median of the 10 counties while its child injury mortality rate is above the median of the 10 counties. This dichotomous categorization gives the impression that County C is doing well with regard to low birthweight, but poorly with regard to child injury mortality. The disparate integer ranks of 3 and 8 give the same impression. In contrast, because of the relative distribution of the 10 counties on each indicator, the percentile rescaled values of 2.02 and 4.00 are closer together, both less than half the distance along the range of 1 to 10.

Z-Scores

The classic method for "standardizing" a set of values (finding a common metric or scale) is calculation of z-scores. Instead of translating data to a fixed range as with percentile rescaling, z-scores are anchored by the mean and standard deviation of the original values, and rescaled such that the new mean is 0 and the new standard deviation is 1. The resulting z-scores correspond to points on the standard normal curve, with a theoretical range of approximately -3 to +3. The actual range for each indicator, however, will be different.

Z-scores can be calculated for individual level or aggregate level data. In either case, each value is treated as an "individual" member of a sample. The sample size, then, is equal to the total number of "individuals". When z-scores are calculated for aggregate data, such as county rates and percents, the county is the "individual" (unit of analysis), the group of counties is the sample, and the sample size is therefore the total number of counties. The formula for calculating z-scores is as follows:

Individual Level Data

$$z_i = \frac{X_i - \overline{X}}{\sqrt{\dfrac{\sum_{i-1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}}}$$

where X_i = the original value for individual i

and $\overline{X}$ = the mean of the distribution of individuals

and n = the total number of individuals

Aggregate Level Data

$$z_i = \frac{p_i - \overline{p}}{\sqrt{\dfrac{\sum_{i=1}^{n}(p_i - \overline{p})^2}{(n-1)}}}$$

where p_i = the original proportion for county i (or region, census tract, etc.)

and $\overline{p}$ = the mean of the distribution of counties (or regions, census tracts, etc.)

and n = the total number of counties (or regions, census tracts, etc.)

For the 10 county data, the mean of the 10 low birthweight rates is:

$$\frac{4.17 + 5.68 + 6.08 + 6.41 + 6.49 + 6.96 + 7.20 + 7.75 + 8.44 + 21.05}{10} = 8.02$$

And the standard deviation is:

$$\sqrt{\frac{(4.17 - 8.02)^2 + (5.68 - 8.02)^2 + \ldots + (8.44 - 8.02)^2 + (21.05 - 8.02)^2}{9}} = 4.72$$

Therefore, the z-score for County C's low birthweight rate is:

$$z_C = \frac{6.08 - 8.02}{4.72} = -0.41$$

And, the z-score for County H's low birthweight rate is:

$$z_H = \frac{7.75 - 8.02}{4.72} = -0.06$$

Here are the results of calculating z-scores for all 10 of the county low birthweight rates :

Integer Ranking and z-score for all of the county low birthweight rates:			
County	Original Values (%)	Rank	z-score
A	4.17	1	-0.82
B	5.68	2	-0.50
C	6.08	3	-0.41
D	6.41	4	-0.34
E	6.49	5	-0.32
F	6.96	6	-0.23
G	7.20	7	-0.17
H	7.75	8	-0.06
I	8.44	9	0.09
J	21.05	10	2.76

Like integer ranking and percentile rescaling, the z-scores are in the same order as the values in the original data. The distances between z-scores are also not uniform, but mirror the relative distances between the original values. The distance between the z-scores for counties B and D is 0.16 (-0.34–-0.50) compared to the integer rank difference of 2 (4-2). And, the distance between the z-scores for Counties I and J is 2.67 (2.76–0.09) compared to the integer rank difference of 1 (10-9).

Different than percentile rescaling, z-scores not only maintain the relative distances between values *within* each indicator, they also maintain relative differences *across* indicators. Using the varying standard deviations rather than a fixed range, values at the same percentile will have different z-scores if the standard deviations of the indicators are not equivalent. Generally, z-scores should not be interpreted as tests of statistical significance. Being anchored by the mean of the distribution, though, they can identify values as above or below average—negative versus positive scores. If the distribution is skewed, however, these negative and positive scores may be misleading. Notice that the distribution of the 10 county low birthweight rates is indeed skewed. Because of County J's extreme rate of 21.05 %, the mean of 8.02 used in calculating the z-scores is not a very good measure of the center of the distribution. Recall that in a normal distribution, the values of the mean and the median are the same, but we saw earlier that the median of the low birthweight rates is 6.73, far below this mean of 8.02. As a result, only Counties I and J have positive z-scores (scores higher than the mean), while the eight other counties have negative ones (scores lower than the mean).

Comment and Example

The coefficient of variation (CV), or the ratio of the standard deviation to the mean, measures the relative variability in a set of values. The larger the CV, the wider the range of z-scores. A value at the same percentile of two distributions will have a larger z-score if its distribution has a larger CV than the other distribution. To illustrate, the mean of the low birthweight rates for the 10 counties, as we've seen, is 8.02% and the standard deviation is 4.72; the mean of the child injury mortality rates is 45.52 per 100,000 children and the standard deviation is 14.5. The low birthweight rates have a larger CV than do the child injury mortality rates:

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}}$$

$$\frac{4.72}{8.02} = 0.59 > \frac{14.5}{45.52} = 0.32$$

Therefore, the z-score for a low birthweight rate at a given percentile along the range of low birthweight rates will be larger than the z-score for a child injury mortality rate at the same percentile along the range of child injury mortality rates.

Test Yourself

Question: For the 10 county data, calculate a z-score for County C's child injury mortality rate of 47.69. The mean of the 10 county rates is 45.52 and the standard deviation is 14.50.

Answer: The z-score for County C's child injury mortality rate is as follows:

$$z_C = \frac{47.69 - 45.52}{14.50} = 0.15$$

Question: Compare the z-scores for County C's low birthweight rate and child injury mortality rate with the results of percentile rescaling, the categories above and below the median, and integer ranking.

Answer: Here are the results for County C using the four methods discussed so far:

County C	Low Birthweight Rate = 6.08 %	Child Injury Mortality Rate = 47.69 per 100,000
Above (+) and Below (-) the Median	-	+
Integer Ranks	3	8
Percentile Rescaled Values	2.02	4.00
z-scores	-0.41	0.15

We described earlier how the percentile rescaled values for the two indicators were more similar than either the integer ranks or the dichotomous categories, although all three methods indicate that County C's child injury mortality rate is relatively "worse" than its low birthweight rate. The z-score of -0.41 for low birthweight is below its mean while the one of 0.15 for child injury mortality is above its mean, indicating the same pattern as in the other methods. Both measures, though, are less than one standard deviation away from their respective means, implying that the difference between them is not great.

Z-Tests

Z-tests are designed to test hypotheses about summary statistics and therefore, by definition, they are only applied to aggregate level data. In "Measures of Association and Hypothesis Testing", we described z-tests in some detail, both for comparing two independent proportions or rates, or for comparing a proportion or rate to a standard. In the context of data summarization, z-tests may be used to increase the interpretability of data by comparing an entire set of values to a common standard; identifying statistical significance is of secondary importance. Different than z-scores, z-tests do not assume that a set of aggregate values are "individual" members of the same sample; instead, each aggregate value is properly considered a summary of values from a distinct sample with its own distribution. The sample size, then, varies according to the number of individuals at risk in each sample.

A set of z-tests is in a sense a set of "weighted" z-scores since the z value assigned to a given county, for example, is "weighted" by the number of individuals at risk in that county. The varying reliability of the rates in different counties is thus taken into account. The "scores" are the results of the separate statistical tests of the difference between each county indicator and a standard; they are not points on one curve. The standard used may be the overall or weighted mean of the observed data, or an external standard such as observed data from a state or the nation, or a state or national goal.

The formula for calculating z-tests is as follows:

Binomial (percents) Poisson (rates)

$$z_i = \frac{p_i - \bar{p}_{tot}}{\sqrt{\dfrac{\bar{p}_{tot}\left(100 - \bar{p}_{tot}\right)}{n_i}}} \qquad z_i = \frac{r_i - \bar{r}_{tot}}{\sqrt{\dfrac{\bar{r}_{tot}\ (*)}{n_i}}}$$

where p_i or r_i = the indicator for county i

and $\bar{p}_{tot}$ or $\bar{r}_{tot}$ = either the overall (weighted) indicator for all counties combined, or a "standard"

and n_i = the total number of individuals at risk in County i

and $* = 1, 100, 1{,}000, \ldots, 100{,}000$, etc., according to the units in the denominator

For the 10 county data, the overall low birthweight rate (weighted mean) is:

$$\frac{14 + 193 + 41 + 129 + 12 + 38 + 124 + 93 + 59 + 24}{336 + 3397 + 674 + 2013 + 185 + 546 + 1723 + 1200 + 699 + 114} \times 100$$

$$= \frac{727}{10887} \times 100$$

$$= 6.68$$

The numerator is the sum of the low birthweight births in each county, and the denominator is the sum of the total live births in each county.

The z-test for County C, then, is:

$$z_C = \frac{6.08 - 6.68}{\sqrt{\frac{6.68(100 - 6.68)}{674}}}$$

$$= -0.62$$

And, the z-test for County H is:

$$z_H = \frac{7.75 - 6.68}{\sqrt{\frac{6.68(100 - 6.68)}{1200}}}$$

$$=1.48$$

Here are the results of calculating z-tests for all 10 of the county low birthweight rates :

Integer Ranking and z-tests for 10 County Low Birthweight Rates				
County	Number of Live Births	Original Values (%)	Rank	z-tests
A	336	4.17	1	-1.84
B	3397	5.68	2	-2.33
C	674	6.08	3	-0.62
D	2013	6.41	4	-0.48
E	185	6.49	5	-0.10
F	546	6.96	6	0.26
G	1723	7.20	7	0.87
H	1200	7.75	8	1.49
I	699	8.44	9	1.87
J	114	21.05	10	6.15

Unlike integer ranking, rescaling with percentiles, or z-scores, z-tests do not necessarily correspond to the order in the original data; adjusting for the varying population sizes may alter the ordering. You can see that the z-tests for Counties A and B are not in the same order as their original values or the integer ranks shown above. This is because the relatively large number of live births in County B added "weight" to its score. For z-tests, then, the distance between scores is a function of both the distances in the original data and of the varying population sizes.

Even when population sizes are essentially equal, z-scores and z-tests will be slightly different because of the differing assumptions underlying each method and the resulting differences in the calculational methods used. Remember that z-scores treat aggregate values as though they were values for individuals; z-tests, on the other hand, treat aggregate values as statistics that summarize distinct samples of values for individuals.

Weighted by the varying population sizes in the 10 counties, the overall mean of 6.68 used to calculate the z-tests for the low birthweight rates is a better measure of the center of the distribution than the unweighted mean of 8.02 used earlier in the calculation of z-scores. This value is very close to 6.73, the median of the distribution, and therefore, five of the counties (A-E) have negative z-tests and the other five (F-J) have positive ones.

Because z-tests treat each aggregate value separately, z-tests do not result in a truly common metric across indicators. The ranges in scores can be quite different for different indicators. One indicator may measure the occurrence of a fairly common health event, another may measure the occurrence of a rare event. The z-tests are sensitive to these differences.

Further, one indicator may be based on all live births, another on children < 18 years old, another on the total population. The z-tests are also sensitive to the differences in size of these denominators just as they are sensitive to the differences in the size of populations across geographic areas. The more common the health event and the larger the denominator, the larger the z-test result will be.

Comment and Example

Here are the z-tests for 2 indicators, one measured per 100 (common), the other per 1,000 (rare), each compared to a standard of 9 (per 100 or per 1,000), and each based on a population at risk of 500.

Relatively Rare Event
Relatively Small Population

$$z = \frac{12-9}{\sqrt{\frac{9}{500} \times 1{,}000}}$$

$$= 0.71$$

Relatively Rare Event
Relatively Large Population

$$z = \frac{12-9}{\sqrt{\frac{9}{1{,}000} \times 1{,}000}}$$

$$= 1.00$$

Now, here are the z-tests for 2 indicators, both measured per 100, each compared to a standard of 9 per 100, one based on a population at risk of 500, and the other on a population at risk of 1,000.

Relatively Common Event
Relatively Small Population

$$z = \frac{12-9}{\sqrt{\frac{9(100-9)}{500}}}$$

$$= 2.34$$

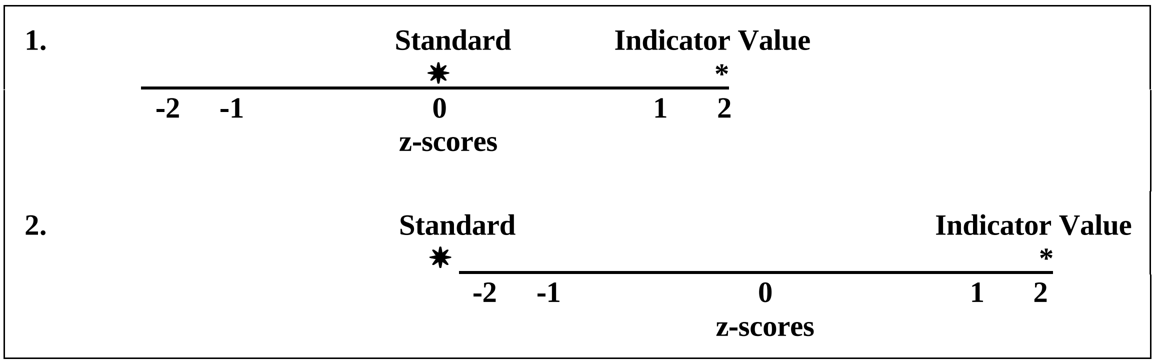
Relatively Common Event
Relatively Large Population

$$z = \frac{12-9}{\sqrt{\frac{9(100-9)}{1{,}000}}}$$

$$= 3.31$$

Notice that the smallest of the four z-tests is that of 0.71 for the rare event measured in a small population; the largest of the z-tests is that of 3.31 for a common event measured in a large population.

Thus far, all of the scoring methods discussed have assigned values according to some internal feature of the observed data, either the median, the range, the unweighted or weighted mean, and/or the standard deviation. One advantage of using z-tests is that they can be anchored to an external standard, which is independent of the distribution of the observed data. Scoring with measures internal to the observed data can only tell us how a particular area is doing in comparison to the other areas being analyzed; the overall standing of all of the areas is not considered. For instance, a county's rate may be two standard deviations "worse" than the mean rate of all of the counties, but this score has quite a different public health meaning depending on whether all of the counties have already met or surpassed an objective for this indicator, some of the counties have met or surpassed the objective while others have not, or none of the counties has yet met or surpassed the objective. An example of this issue is depicted in the graph below:



For indicator #1, approximately half of the counties have already met the standard while the other half have not; for indicator #2, none of the counties has yet met the standard. Reporting a z-score of 2 for both indicators for the one county obscures this important difference. Similarly, having a negative z-score has a very different interpretation depending on the indicator; for indicator #1, counties with negative z-scores have met the standard, but for indicator #2, counties with negative z-scores have not met the standard.

Use of an external standard also provides a reference over time. For example, a county may have a z-score of 2 in comparison to other counties for five successive years, but along with the other counties, it may be making progress toward an objective, as measured by improving z-tests of 2.0, 1.5, and 1.0 over the same five year period. The z-tests anchored to an external standard, then, tell a different and important story.

Our discussion of z-tests has focused on their merits from an analytic and statistical perspective, but it must be acknowledged that accounting for varying population size is not always viewed as an advantage by public health professionals or other stakeholders. If 2 counties have exactly the same rate on a health indicator, but very different population sizes, their integer ranks, percentile rescaled values, and z-scores will also be identical, but their z-tests will be different. The question is always raised, "If our rate and their rate is the same, why do we have a worse score?"

Suppose, for example, that in one year, two counties each have an infant mortality rate of 9 infant deaths per 1,000 live births. This rate might hypothetically be ranked 25th out of 50 counties, might be at the 50th percentile, and have a z-score of 0. Now further suppose that one county is rural with very few live births each year, while the other is urban with many live births each year. The z-test for the rural county may not indicate that its rate is statistically different than a standard ($z <= 1.96$, $p >= 0.05$), but the z-test for the urban county may show a significant result for the same rate ($z > 1.96$, $p < 0.05$).

Some would argue that it is unfair to imply by way of these varying z-tests that the urban county is doing worse than the rural one when in fact their rates are identical. We have more confidence, however, in the accuracy of the rate observed in the urban county since it is based on a large sample size, whereas the accuracy of the rate in the rural county is in question since it is based on a small sample size. In fact, having a "worse" z-test despite identical rates could be used as evidence of the need for additional resources.

It should be noted that z-tests do not always show areas with large populations to be "worse" than their less populated counterparts; if two areas, one large and one small, have both met a standard, the z-test for the large area will imply that it is doing better than the small area even if their rates are the same; in other words, a z-test for a large area will always be farther away from zero than a z-test for a small area with the same rate, whether the values are in the positive or negative direction.

Summary of Categorization / Ranking / Scoring Methods

Discrete Categories	Some loss of information—many distinct values are condensed into only a few. Reduces the number of values an audience has to assimilate: easy to interpret if labels are intuitive such as "high" and "low".
Integer Ranking	Preserves the order of the sorted data as well as the number of values, but imposes a uniform set of distances between data points. Audiences can readily understand the meaning of being ranked "1st", "5th", "50th", "100^{th}", etc.
Percentile Rescaling	Preserves the order of the sorted data, the number of values, and the relative distances between data points. A given percentile on any range of original values will be assigned the same position on the new common range. The new range can be for example, 1 to 10, 1 to 100, 0 to 5, -5 to +5.
z-scores	Preserves both the order of the sorted data and the relative distances between data points, and also accounts for differing variability across indicators. By definition, yields a center of 0, but the scores are somewhat less readily understood by general audiences.
z-tests	Both the order of the data and the relative distances between the data points are "adjusted" according to the varying population sizes in the areas of interest. Also sensitive to how common or rare the health event being measured; does not result in a truly common metric across indicators, although values and ranges are usually similar. The value of the scores is determined by the value used as the "standard"; an external standard adds equivalence across indicators and within an indicator over time.

Test Yourself

Question: For the 10 county data, calculate a z-test for County C's child injury mortality rate of 47.69 per 100,000. The overall mean of the 10 county data is 42.77 and the population of children < 18 in County C is 13,480.

Answer: The z-test for County C's child injury mortality rate is as follows:

$$z_C = \frac{47.69 - 42.77}{\sqrt{\frac{42.77}{13,480} \times 100,000}} = -0.28$$

Question: Compare the z-tests for County C's low birthweight rate and child injury mortality rate with the z-scores, the results of percentile rescaling, the categories above and below the median, and integer ranking.

Answer: Here are the results for County C using five methods:

County C	Low Birthweight Rate = 6.08 %	Child Injury Mortality Rate = 47.69 per 100,000
Above (+) and Below (-) the Median	-	+
Integer Ranks	3	8
Percentile Rescaled Values	2.02	4.00
z-scores	-0.41	0.15
z-tests	-0.62	0.28

The z-tests, as with the other methods, show that County C's low birthweight rate is lower than its child injury mortality rate, in comparison to the other nine counties. The dichotomous categories and the integer ranks imply that this difference is quite large; the percentile rescaled values, z-scores, and z-tests imply it is quite small. The z-tests, however, do imply a somewhat larger difference than do the z-scores— -0.62-0.28 is greater than -0.41-0.15 —because of the assumptions of the method and because low birthweight is a much more common event than child injury mortality.

Index Construction

In the broadest terms, an index is any measure that expresses relative values of a phenomenon. Any single ordinal or continuous variable meets this definition; the infant mortality rate, for example, can be considered an "index" of infant death.

More commonly, though, the term index implies that information from more than one data element has been combined into one composite measure. In addition, an index is usually meant to be used over time, and therefore, comparison to a standard that is fixed over time should be built in to maximize its effectiveness.

There are many different types of indices. An index may be based on either individual or aggregate level data, and may combine variables which are related to the same indicator, may combine variables which cross distinct indicators, but which are related to a larger construct, or may combine variables which cross constructs, but which are related as dimensions of a larger, complex system. Below are a few brief examples illustrating various types of indices.

1. Combining variables which are related to the same indicator.

An index of low birthweight might combine both the numbers of low birthweight births with the low birthweight rate (%) in order to simultaneously address the service burden and the rate of occurrence. Both a high number of low birthweight births and a high rate will be considered indicative of a problem. Here are the results for all 10 counties using integer ranks:

County	Integer Ranks for Number of Low Birthweight Births	Integer Ranks for Low Birthweight Rate (%)	Index Scores of Low Birthweight Mean of the 2 Ranks
A	2 (14)	1 (4.17)	1.5
B	10 (193)	2 (5.68)	6.0
C	5 (41)	3 (6.08)	4.0
D	9 (129)	4 (6.41)	6.5
E	1 (12)	5 (6.49)	3.0
F	4 (38)	6 (6.96)	5.0
G	8 (124)	7 (7.20)	7.5
H	7 (93)	8 (7.75)	7.5
I	6 (60)	9 (8.44)	7.5
J	3 (24)	10 (21.05)	6.5

Considering the number of low birthweight births along with the low birthweight rate draws a different picture than when the low birthweight rate was considered alone. Counties G, H, and County I, rather than County J have the highest index score (7.5 v. 6.5), and County D, although ranked only fourth in terms of its low birthweight rate, is tied with County J on the composite index.

Another index of low birthweight might consider the current low birthweight % in combination with the trend in the rate over the past several years of data. One possible set of categories could be created as follows:

1. Low birthweight rate below the median and rate decreasing
2. Low birthweight rate above the median and rate decreasing
3. Low birthweight rate below the median and rate constant or increasing
4. Low birthweight rate above the median and rate constant or increasing

Notice that the index scores are ordered such that category 1 is the "best" and category 4 is the "worst". In addition, the *trend* in the low birthweight rate takes precedence over the current level of the rate; counties with low birthweight rates below the median but which show deterioration or no improvement over time are assigned a worse score (3) than counties with low birthweight rates above the median but which show improvement over time which are assigned a better score (2). In other words, a constant or increasing rate is considered a more powerful marker of a problem than the current level of the rate itself.

There are many alternative ways to choose the number of categories and the ordering of the categories for an index. Using the mean of the integer ranks for the number of low birthweight births and the rate of low birthweight resulted in an index with 10 categories for the 10 counties, whereas using above and below the median and decreasing and increasing trends in low birthweight resulted in a four category index. Also, in the four category approach, the ordering of the categories might have given precedence to the current level of the rate rather than the trend in the rate—categories 2 and 3 could be reversed.

2. Combining variables which cross distinct indicators, but which are related to a larger construct.

A health status index might combine the values of many indicators into one measure. As a simple illustration, the low birthweight rates and the child injury mortality rates for the 10 counties will be combined. Here are the results for all 10 counties using integer ranks:

County	Integer Ranks for Low Birthweight Rate (%)	Integer Ranks for Child Injury Mortality Rate (per 100,000)	Index Scores of Health Status Mean of the 2 Ranks
A	1 (4.17)	7 (46.60)	4.0
B	2 (5.68)	6 (45.94)	4.0
C	3 (6.08)	8 (47.69)	5.5
D	4 (6.41)	3 (36.55)	3.5
E	5 (6.49)	10 (81.50)	7.5
F	6 (6.96)	2 (31.38)	4.0
G	7 (7.20)	5 (42.83)	6.0
H	8 (7.75)	4 (39.36)	6.0
I	9 (8.44)	1 (30.79)	5.0
J	10 (21.05)	9 (52.64)	9.5

Another index could be defined by combining a health status and a health services indicator to address a particular health problem. For example, the trend in a health status measure could be combined with the trend in the availability of relevant health services. Counties might then be assigned to one of the following categories:

1. health status improving, services constant or increasing
2. health status improving, services decreasing
3. health status deteriorating, services constant or increasing
4. health status deteriorating, services decreasing

In this index, the trend in health status is given precedence over the trend in available services; improvement or deterioration in the rate of occurrence is considered primary since health services are only one factor influencing any change in health status.

3. An index may combine variables which cross constructs, but which are related as dimensions of a larger, complex system.

A global index of MCH need might include measures of risk and of outcome, and also measures of access, availability and utilization of health services. In this scenario, a constellation of indicators is selected, index scores are generated to summarize all of the data, and geographic areas are grouped according to the index scores.

Selecting which component measures are to be used in constructing an index, and exactly how these measures will be used should be driven by a conceptual understanding of the relationship among indicators with respect to the particular issue being addressed. A given variable might be handled in any of several ways. For example, it might be:

- Used as one of the primary measures included in an index
- Used to standardize (adjust) the primary measures in an index
- Used to define strata of the component measures in an index
- Used to weight the component measures in an index
- Not used at all

Categorization, integer ranking, percentile rescaled values, z-scores, and z-tests may all be used in constructing indices. Once the selected measures have been transformed to a comparable scale, whether through creation of discrete or continuous scores, then an approach for combining the measures into a summary index can be developed.

As the examples above illustrated, the index itself may be categorical or continuous. If the component measures are categorized, then an algorithm can be developed to combine the categories into new ones. If the component measures are continuous scores, then an algorithm can be developed to combine these either into discrete categories or into a new set of continuous scores.

The algorithm is a set of rules for how to combine the component measures. Each component measure may be treated equally (unweighted) or differentially (weighted). If measures are to be weighted, usually the differential weights are defined conceptually, by considering the relative importance of each component measure in terms of, for example, the consequences of the outcome, how common it is, whether intervention strategies exist to address it, how effective the existing strategies are, and how feasible it is to implement interventions in terms of cost and other factors. If statistical procedures are used to combine measures, weighting may be part of the algorithm itself.

A typical algorithm for constructing an index is simply to sum (or compute the mean) of the scores across all of the component measures. If weighting is to be used, this is done either explicitly by multiplying each measure by its assigned weight prior to summing, or implicitly by defining discrete categories that result in summing across non-equivalent levels of each measure. Let's look at an example to see how this process works.

Suppose the index combining number of low birthweight births and the low birthweight rate seen earlier is created, but this time giving more weight to the low birthweight rate than to the number of low birthweight births. An explicit weight of 2 might be assigned to the rate, while the number of low birthweight births will have a weight of 1.

County	Integer Ranks for Number of Low Birthweight Births Weight=1	Integer Ranks for Low Birthweight Rate (%) Weight=2	Index Scores of Low Birthweight Mean of the 2 Ranks Weighted
A	2 (14)	2×1=2 (4.17)	2.0
B	10 (193)	2×2=4 (5.68)	7.0
C	5 (41)	2×3=6 (6.08)	5.5
D	9 (129)	2×4=8 (6.41)	8.5
E	1 (12)	2×5=10 (6.49)	5.5
F	4 (38)	2×6=12 (6.96)	8.0
G	8 (124)	2×7=14 (7.20)	11.0
H	7 (93)	2×8=16 (7.75)	11.5
I	6 (60)	2×9=18 (8.44)	12.0
J	3 (24)	2×10=20 (21.05)	11.5

The index scores are now somewhat different than before (refer to the table on page 163). County I now has the highest score, though not by very much. Counties G and H, which were tied before are now differentiated since County H's low birthweight rate is higher than County G's and this indicator is weighted more. In addition, these two counties are no longer scored higher than County J as they were when no weighting was applied. Finally, County D's score is relatively lower in comparison to the counties with the highest scores, because its large number of low birthweight births now has a diminished impact on the index score.

Now let's look at using discrete categories to implicitly weight the two indicators. The low birthweight rate will be divided into 5 categories with two counties in each (quintiles) and the number of low birthweight births into only 2 categories with five counties in each (below and above the median). Here are the results:

County	Median Ranks for Number of Low Birthweight Births	Quintile Ranks for Low Birthweight Rate (%)	Index Scores of Low Birthweight Mean of the 2 Ranks Weighted
A	1 (14)	1 (4.17)	1.0
B	2 (193)	1 (5.68)	1.5
C	1 (41)	2 (6.08)	1.5
D	2 (129)	2 (6.41)	2.0
E	1 (12)	3 (6.49)	2.0
F	1 (38)	3 (6.96)	2.0
G	2 (124)	4 (7.20)	3.0
H	2 (93)	4 (7.75)	3.0
I	2 (60)	5 (8.44)	3.5
J	1 (24)	5 (21.05)	3.0

Because the low birthweight rates are divided into more categories than the number of low birthweight births, they dominate the order of the index scores. The number of low birthweight births, however, has influenced the index scores somewhat. For instance, County D is tied with Counties E and F because, even though its low birthweight rate is lower, it is in the higher category for number of low birthweight births.

One final note on constructing indices: be careful before combining measures (or even before comparing measures), that the directionality of the indicators is comparable—make sure a high score on one measure has the equivalent meaning as a high score on another measure. On some indicators, such as number of physicians per capita in a community, a high score is "good" and a low score is "bad". Categories, ranks, rescaled values, and scores must reflect this. For example, the order of ranks for the number of physicians per capita and the signs (positive or negative) of z-scores or z-tests would have to be reversed in order to achieve consistency with other indicators such as low birthweight or child injury mortality. If this seeming detail is forgotten, index scores will be inaccurate and misleading.

You can see that there are a myriad of approaches to constructing indices. The analyst has a great deal of flexibility in accomplishing this task, but the underlying conceptual framework should not be allowed to get lost in the mechanical process.

Comment

Often, when summarizing data, multiple methods are used in combination. For example, several indicators with values for many geographic areas might first be rescaled to a common range, then combined into an index by calculating the mean of the rescaled values for each area, and finally turned into a few discrete categories with labels such as "areas in most need", "areas with moderate need", and "areas with least need".

Similarly, after initial ranking or scoring, an index might be constructed and then ranking or scoring might be applied again to the index values themselves.

Sometimes, two scoring methods are used at once to capture multiple perspectives: for instance, z-tests might be used to identify statistically significant departures from a standard, while integer ranks might be used as well to provide a more readily understood measure of relative position.

Mapping is often chosen as a presentation format to gain the advantages of index construction without losing the information on each component measure. The layering of information on the map in effect combines it into one composite measure, but the layers can be pealed away or re-combined (the index deconstructed or reconstructed) to yield slightly different views of the health issue being examined.

Index construction might also occur in stages. For example, indicators within given domains might be rescaled and combined into distinct indices, and then these indices may be combined into one grand summary measure. For example, one index might be created to summarize health status measures, one to summarize health services measures, and one to summarize sociodemographic measures. These indices could be reported separately or together.

SUMMARIZING DATA TO ALLOCATE RESOURCES

The results of data summarization are sometimes used as input into a formula for distribution of resources. Developing a distribution formula requires its own series of choices that correspond to conceptual, analytic, practical, and political considerations. Perhaps the most fundamental issue to be addressed is whether resources will be distributed according to the size of the population at risk, according to need as measured by "high" scores on indicators or indices, according to some combination of both population size and need, or according to a more complex method that accounts for other factors such as demographic characteristics, historical funding patterns, or status of the public health infrastructure.

For example, it might be decided to divide the resource pool into two portions such that funds are allocated separately, according to population size alone and according to population size weighted by need. The pool might be split evenly, or a 10-90 split, a 25-75 split, or some other division may be deemed appropriate. Unless population sizes across areas do not vary greatly, it is necessary to consider need on a per capita basis in order to insure equitable distribution of funds. It is possible, though, to leave the resource pool undivided, allocating 100% of the resources based on population size weighted by need.

It may also be important to designate a portion of the resource pool for funding areas that have a track record in implementing programs and policies that result in improved health status. These areas often score low on need indices and therefore are not targeted for infusion of resources. It can be argued, however, that their successes should be rewarded with funding, both to facilitate continued improvement, and as an incentive for other areas to adopt and adapt effective approaches.

In essence, developing a resource allocation formula is constructing an index, using whatever factors are considered relevant to the process. For example, all previously computed indicator or index scores might be rescaled to reflect the desired differential in weighting. Perhaps a range of 1 to 2 will be chosen so that the "worst" area receives twice the funding on a per capita basis as the "best" area. Perhaps a range of 1 to 5 will be chosen to increase the per capita differential in funding. Perhaps all areas will first be allocated a fixed amount based on population at risk and then only those areas whose z-scores are > 1 will be allocated additional funds based on excess need. Perhaps those areas with more than a 1% average annual increase (deterioration) and those with more than a 1% average annual decrease (improvement) will be given a weight of 2, while those areas with less than a 1% average annual change in either direction will be given a weight of only 1.

The resource allocation index or formula can also build-in rules that reflect political considerations. For instance, a baseline level of funding can be defined either in absolute dollars or relative to historical funding levels. The baseline might be defined so that no area receives less funding than it had in the past, or so that any decreases are limited to, say, a 5 or 10 % reduction. Similarly, a funding ceiling can be defined either in absolute dollars or relative to historical funding levels. The ceiling might be defined, for example, so that no area receives more than a 5 or 10% increase.

Suppose $2,000,000 is to be distributed to counties A through J for low birthweight prevention. In the past, this amount had been distributed based solely on the total number of births in each county. The following table shows the distribution of funds to each county according to its proportionate share of the 10,887 births that occurred in all ten counties combined:

County	Number of Births	% of Total Births	Funds Allocated ($)
A	336	3.1	62,000.00
B	3,397	31.2	624,000.00
C	674	6.2	124,000.00
D	2,013	18.5	370,000.00
E	185	1.7	34,000.00
F	546	5.0	100,000.00
G	1,723	15.8	316,000.00
H	1,200	11.0	220,000.00
I	699	6.4	128,000.00
J	114	1.0	20,000.00
TOTAL	10,887	100.0	~2,000,000.00

Now suppose that the following hypothetical rules for a new distribution formula are agreed upon:

1. Allocation of funds will be based on both population at risk and excess need as reflected in low birthweight z-tests.

2. All counties will receive funds from each category—the county with the smallest population will receive funds based on its population at risk and the county with the best low birthweight z-test will receive funds earmarked for excess need.

3. The $2,000,000.00 will be split evenly between population at risk and excess need, and the z-tests will be rescaled to a range of 1 to 2 so that no county will experience more than a 10 % reduction in resources.

The following table lays the basis for the distribution of funds according to the new rules:

County	a. Low Birthweight z-tests	b. # of Births	c. % of Total Births
A	-1.84	336	3.1
B	-2.33	3,397	31.2
C	-0.62	674	6.2
D	-0.48	2,013	18.5
E	-0.10	185	1.7
F	0.26	546	5.0
G	0.87	1,723	15.8
H	1.49	1,200	11.0
I	1.87	699	6.4
J	6.15	114	1.0
		10,887	100.0

(Table continued on next page)

County	d. Low Birthweight Rescaled z-tests	e. # of Births Weighted by Rescaled z-tests	f.* % of Total Births Weighted by Rescaled z-tests
A	1.06	356	2.7
B	1.00	3,397	25.4
C	1.20	809	6.1
D	1.22	2,456	18.4
E	1.25	231	1.7
F	1.31	715	5.4
G	1.38	2,378	17.8
H	1.45	1,740	13.0
I	1.50	1,049	7.9
J	2.00	228	1.7
		13,359	100.0

Column d. shows the z-tests rescaled to a range of 1 to 2 using the formula described on pages 150-151. County C's rescaled value of 1.20, for example, was calculated as follows:

County C

$$\frac{\text{New Value} - 1}{2 - 1} = \frac{-0.62 - (-2.33)}{6.15 - (-2.33)}$$

$$\text{New Value} - 1 = \frac{1.71}{8.48} \times 1$$

$$\text{New Value} = (0.20 \times 1) + 1$$

$$= 1.20$$

Column e. is the distribution of the populations at risk weighted by the rescaled z-tests. County C's weighted number of births, then, is calculated as 1.20 × 674 = 809 (column d. × column b.), and this is 6.1% of the weighted total ($(809 \div 13{,}359) \times 100$). Once these weighted percents are calculated for all 10 counties, the $2,000,000.000 can be allocated. Remember that half of the funds, or $1,000,000.00, will be allocated based solely on the population at risk and the other half will be allocated based on the z-tests weighted by the population at risk. The following table shows the results:

County	% of Total Births Unweighted	Funds Allocated Population at Risk ($)	% of Total Births Weighted by Rescaled z-tests	Funds Allocated Excess Need ($)
A	3.1	31,000.00	2.7	27,000.00
B	31.2	312,000.00	25.4	254,000.00
C	6.2	62,000.00	6.1	61,000.00
D	18.5	185,000.00	18.4	184,000.00
E	1.7	17,000.00	1.7	17,000.00
F	5.0	50,000.00	5.4	54,000.00
G	15.8	158,000.00	17.8	178,000.00
H	11.0	110,000.00	13.0	130,000.00
I	6.4	64,000.00	7.9	79,000.00
J	1.0	10,000.00	1.7	17,000.00
	100.0	~1,000,000.00	100.0	~1,000,000.00

After summing the funds allocated to each county from both halves of the resource pool, here is the comparison between the old and the new distribution:

County	Previous Allocation of Funds ($) Population at Risk Only	New Allocation of Funds ($) Population at Risk and Need 50-50 Split
A	62,000.00	58,000.00
B	624,000.00	566,000.00
C	124,000.00	123,000.00
D	370,000.00	369,000.00
E	34,000.00	34,000.00
F	100,000.00	104,000.00
G	316,000.00	336,000.00
H	220,000.00	240,000.00
I	128,000.00	143,000.00
J	20,000.00	27,000.00
	~2,000,000.00	~2,000,000.00

The hypothetical rules agreed upon for distributing funds were quite conservative, resulting in fairly minor differences from the old method. Counties A-C, with the lowest low birthweight rates would receive a slight decrease in funds, while Counties H-J, with the highest low birthweight rates would receive a slight increase in funds.

If a greater proportion of funds had been allocated based on excess need, if decreases and increases in funding had not been limited, and/or if the z-tests had been scaled to a range greater than 1-2, the results would have shown more pronounced differences compared to the old method. For example, if 100% of the funds were allocated based on the rescaled z-tests for the low birthweight rates, the results would be as follows:

County	Previous Allocation of Funds ($) Population at Risk Only	New Allocation of Funds ($) Population at Risk and Need 0-100 Split
A	62,000.00	54,000.00
B	624,000.00	508,000.00
C	124,000.00	122,000.00
D	370,000.00	368,000.00
E	34,000.00	34,000.00
F	100,000.00	104,000.00
G	316,000.00	356,000.00
H	220,000.00	260,000.00
I	128,000.00	158,000.00
J	20,000.00	34,000.00
	~2,000,000.00	~2,000,000.00

Now the shift in funding is greater, but the hypothetical rules have been violated. In order to increase funds to Counties H-J with their elevated low birthweight rates, the funding of Counties A and B is reduced by more than 10%—County A's funding is reduced by more than $6,200.00 and County B's funding is reduced by more than $62,400.00. Once again, analytic choices greatly influence results and must always be made with care and with a sound rationale that is developed through a broad based consensus building process.

SUMMARY

Summarizing data is a complex process, with summarization occurring within and across many dimensions simultaneously and sequentially. A well structured analysis plan is essential for controlling the process and insuring that the final product reflects the analytic goals. The data burden can be reduced both by limiting the amount of data used and by applying methods to increase the interpretability of the data. Categorizing, ranking, scoring, indexing, and multivariable statistical methods are tools for increasing the interpretability of data, as are graphing, mapping, and other presentation methods.

It is critical to state the assumptions and acknowledge the overall strategies that drive the choice of analytic methods and presentation formats. Different choices can lead to different results and potentially to different conclusions. It is essential, therefore, to examine the impact of, and the tradeoffs, that each choice carries with it. Sensitivity analysis—contrasting the results achieved using differing sets of choices—is strongly recommended.

Data summarization is not an end in itself. Its purpose is to achieve a coherent view of health problems—a view that, in maternal and child health, promotes action to improve the health of women, infants, children, adolescents, and families.