Chapter 8: Data Interpretation and Dissemination

At the heart of chronic disease epidemiology is the quest to understand past disease trends and their association with potential risk and protective factors so to inform public health decisions and actions. This mission requires the development of appropriate methods to measure the impact of disease trends and associated risk factors. This data-driven approach is the foundation of evidence-based public health (Chapter 5). Data collection is just the first step, followed by analysis and dissemination of surveillance data that address real-world problems. This chapter covers key concepts and tips common for analysis and dissemination of chronic disease data, though not unique to it. As the lead chronic disease epidemiologist, your statistical knowledge, your skill in assessing bias in observational data, and your understanding of the strengths and limitations of each surveillance method (Chapter 7) and their impact on the resulting data will be crucial to appropriately interpreting data results. In other words, you will be able to objectively state what the results mean and do not mean. In addition to data interpretation, you will synthesize results to identify what public health actions the results support. You will translate the results into everyday language that policy makers and other decision makers can readily use.

Review Concepts Critical for Analyzing and Interpreting Data

CDC provides an online, self-study course in the principles of epidemiology in public health practice.\(^1\) Its six lessons and glossary cover descriptive and analytic epidemiology (lesson 1), disease concepts, surveillance (lesson 5), and applied biostatistics (lesson 3 on measures of risk). The website of the National Association for Public Health Statistics and Information Systems provides statistical measures commonly used with birth and death data.\(^2\) University of California, Los Angeles shares a good resource for selecting the most appropriate statistical test to use.\(^3\) For convenience, defined below are a few core statistical measures and concepts.

- **Type of measurement: Quantitative** data are information that can be measured (counted) and therefore can be represented with numbers. Chronic disease epidemiologists often use quantitative data. In contrast, **qualitative** data cannot be measured though it can be described or observed, such as colors or textures. Therefore, qualitative data represents information in non-numeric form, such as narrative descriptions in the form of text, audio words, or images.

---


Examples of qualitative data more common in public health are focus group notes or answers to open-ended interview questions. Qualitative data can provide insight or context that fills a gap in the quantitative data. However, qualitative findings will not be generalizable beyond the group who provided the information.

- **Measuring data quality**: Two qualities of measurement necessary for understanding and interpreting accurately data results are validity and reliability. **Validity** is the degree to which the measurement actually measures what it is intended to measure. **Reliability** refers to the degree to which a measurement produces the same values when identical measurements are repeated in the same population. If a measure is not valid, then what is being measured is not known. If a measure is not reliable, then one cannot measure changes over time.

- **Measuring magnitude of events or disease occurrence in a population**: **Incidence rate** is the rate of new cases of a disease or condition among a population over a defined period of time. Incidence rate is also a proportion, because the persons in the numerator are also in the denominator. A commonly used incidence rate is a death rate from the state mortality system of death certificates. **Prevalence** is the rate of existing cases of a disease, condition or behavior (e.g., tobacco use) among a population at a point in time or over a defined period of time. The most commonly used chronic disease prevalence rates come from BRFSS. BRFSS prevalence estimates can approximate the current (or historical) burden of a disease or risk factor in a population but not usually the number of new cases. Prevalence is a function of incidence and duration of a condition. Both can be used as measures of risk. **Adjusted rates** are rates that have been standardized in such a way to allow for fair comparisons of rates over time or among different populations defined by the geography where they reside. The most common in epidemiology is age-adjusted rates, because the rate of chronic diseases increases with increasing age. The direct method of age adjustment: Age-specific rates for a population at a particular time and place are applied to a standard age distribution, such as the U.S. population in 2000. Age-specific rates for a second population are applied to the same standard age distribution. In effect, the two resulting artificial or hypothetical rates allow for a fair comparison, because the resulting rates are independent of the underlying age distribution of the population they represent.

- **Measuring change in risk**: **Relative reduction** is the percent change from the baseline period compared to the next period of interest (e.g., 50% relative decrease from a prevalence of 20% in 2000 to 10% in 2010). It can also be the percent change among the experimental group compared to the control group. It is calculated as follows:

\[
\frac{(\text{Event Rate among Experimental Group} - \text{Event Rate among Controls})}{\text{Event Rate among Controls}}
\]

or

\[
\frac{(\text{Baseline Rate} - \text{Current Rate})}{\text{Baseline Rate}}
\]

**Absolute reduction** is the difference between the rate during the baseline period or among the control group and the rate during the current period or among the experimental group (e.g.,
10% absolute difference from the prevalence of 20% in 2000 to 10% in 2010, sometimes described as a 10-point change).

- **Examining trends** is a basic analysis of surveillance data by time to detect changes in incidence or prevalence of risk and protective factors and health outcomes. Assess if the change in risk is an artifact of an increase in population size, improved diagnostic procedures, and/or enhanced reporting and other reporting biases.

- **Measures of association** between an exposure or risk factor and a health outcome include risk ratio (relative risk), rate ratio, odds ratio, and population attributable risk. An association is any observed relationship or pattern between to measured quantities. An association is not the same as causation.\(^4\) **Risk ratio** is the ratio of incidence proportions of two groups (the percentage of persons with a disease in group one divided by the percentage of persons with a disease in a second group). The **rate ratio** is the ratio of incidence rates of two groups. **Odds ratios** provide a reasonable approximation of a risk ratio when the study design is a case-control study. In a case-control study, one cannot calculate risk, because the size of the population from which the cases (the persons with the outcome of interest) is not known. **Confounding** is the distortion of an association between an exposure (risk factor) and a health outcome by a third variable related to both the exposure and the outcome. **Interaction** is modification of the effect of the exposure on the outcome by a third variable. For example, the risk of mesothelioma from exposure to asbestos greater among smokers than non-smokers. **Population attributable risk** measures the public health impact of an association between an exposure and outcome. Also known as attributable proportion or attributable risk percent, it represents the expected reduction in disease if the exposure could be removed. The definition is the difference between risks for the exposed group and unexposed group divided by the risk for the exposed group times 100%. For example, Colorado analyzed the population attributable risks of maternal smoking and inadequate weight gain during pregnancy on low birth weight.\(^5\)

- **Statistically significance** is a measure of how likely a result could have occurred by chance alone, and a **confidence interval** is a range of values for a measure or estimate (e.g., rate or odds ratio) constructed so that the range has a specified probability of including the true value of the measure in the population. Often the probability, which is selected in advance of running the calculations, is 95 percent. For example, if an epidemiologist took 100 random samples from the population and each time measured obesity in the sample, 95 times out of a 100, the true obesity prevalence in the population would be within the 95% confidence interval. If the confidence intervals for estimates (in the same population in two time periods or two populations in the same time period) do not overlap then these two estimates are statistically significantly different. However, this difference might not be clinically relevant or

---


meaningful in practical terms. Also, using confidence intervals to test for statistical significance is sometimes a conservative approach.

Understand Concepts Critical for Disseminating Data Results
There are circumstances where you as the lead chronic disease epidemiologist cannot or should not release data results. Not releasing data is referred to as “data suppression”. The two main reasons to suppress data are: 1) to protect the confidentiality of the person or persons and 2) to prevent the use of unreliable data or low quality data. To protect confidentiality, follow the confidentiality guidelines and data use agreements—standards established by the health department, data source, or surveillance system to protect the identity of the individuals or jurisdictions represented within the data (Chapter 6). Public health practitioners can be held personally liable for not following these standards. The exact threshold for suppression varies in confidentiality guidelines and might be based solely on the numerator or both the numerator and the denominator in an analyzed table of results. For example, a confidentiality guideline could specify that any count in a table that is less than 5 must be suppressed, meaning deleted from the table of results, regardless of the total count in the table or the total population the analyzed results represent. Some guidelines further specify that if the reader can calculate the number suppressed from the table of results that more information must be suppressed or further aggregated. To prevent the use of low quality data, there are sample size guidelines—standards established by the surveillance system to ensure that only representative (and reliable) data is shared. For example, the Behavioral Risk Factor Surveillance System suppresses data with an unweighted sample size of the denominator is less than 50 respondents or the relative standard error is greater than 0.3. In addition, caution is warranted in interpreting estimates based on cell sizes (numerators) less than 50.

Translate Data for Optimal Messaging to Your Intended Audience
For data to effectively drive action, the data results from surveillance systems must be easy for staff to use appropriately and easy for the public to understand. Your role is to determine the audience(s) for the data results and their specific needs, select appropriate formats for sharing the results, identify the key messages for each audience, and share information accordingly. You are translating numeric, statistical information into everyday language while also interpreting and revealing the meaning of the results. Telling the results as a story that frames the issue and uses social math provides the kind of everyday language that busy professionals (colleagues, policymakers) immediately grasp and will use.

Frame your information to tell a story
The London-based data-journalist and “information designer” David McCandless advises data purveyors to “design information so it makes more sense, tells a story, and allows us to focus only
on information that is important.” Answer these questions before developing your message from the data, a message that connects the problem, solution, values, and action.

- What story do the data tell? What problem and solution, if any, do the data suggest?
- What is your end game? What question were you trying to answer with the data analysis? What is your call to action?
- Who is your audience? What do they value? What is their context? What decisions are they facing?
- Numeracy or statistical literacy of the audience? Can they translate percentages and rates? If not, the data will need to be presented as simply as possible, in a clear language with compelling context and visual displays of the results that would tell the same story, if the words were removed. What is your understanding of the concept and meaning, not just the tool of data analysis and the recipe for calculating a rate? Do you understand the concept well enough to tell it simply?

Use social math and framing to help tell your story

Social math is “the practice of translating statistics and other data so they become interesting to the journalist and meaningful to the audience.” As outlined in CDC’s Framing Guide for Communicating about Injury, there are several steps for creating social math and a compelling story:

1. Consider the message frame. For example, CDC’s frame for those who work in the field of injury is “We want a society where people can live to their full potential.” This frame is a value that you and the audience have in common. When selecting a frame, consider what you want to accomplish and whether you can make interesting connections, comparisons or metaphors.

2. Make a strong and dramatic statement of the problem.
   - Select relevant examples appropriate for the target audience.
   - Make the statistic meaningful to the audience. It can be effective to break data down by time (e.g., 400,000 deaths per year) and by place to localize the information as much as possible. However, do not provide a long list of statistics about the problem. People want to know about the solutions, what they cost, and how they will get done.
   - Find useful comparison statistics, such as a statistic about a familiar thing. For example, compare the daily individual cost of a program to the cost of a daily latte coffee. The comparison can be dramatic or unusual. However, avoid invalid, unfair comparisons.

---


3. Get to the solution sooner, and use positive, action-oriented statements about the solution.
   - Reinforce the science without jargon.
   - If personal responsibility and/or community action are common values, tie the solution to them.
   - Ensure that the message has a call to action and engages the audience to act.

When creating social math, make sure the data results and concepts:9
   - Are 100% accurate.
   - Make sense and are related.
   - Are appropriate for the audience.
   - Are married to the story.
   - Are visual, if possible.
   - Are dramatic.
   - Engages the audience in fixing the problem.
   - Do not depict mayhem.
   - Are used sparingly.

To recap, the steps to arrive at an effective story that connects the problem, solution, values, and action: Consider the frame, relevant examples, statistics, and comparison; limit what to present; avoid invalid, unfair comparisons; check facts and visuals.

**Present effectively**

Known for his engaging presentation, former Apple CEO Steve Jobs used a similar approach as described above. His ten favorite public speaking tips “to be insanely great in front of any audience” are:10

1. Plan your presentation with pen and paper.
2. Simplify complex information.
3. Tell a story with a villain and a hero. [This tip might not always work for public health. Public health professionals and government officials are careful in describing the problem and the potential solutions. They are careful not to demonize or victimize people and businesses.]
4. Personalize benefits. The audience needs to know “what’s in it for me?”
5. Stick to the rule of three. It is easier for people to remember three key points. [Three verbal points are common in U.S. culture. A story has a beginning, middle, and end. Jokes have three repetitions. Sermons have three points.]
6. Evoke a higher sense of purpose.
7. Create visual slides.

---


8. Make numbers meaningful.
9. Use plain English.
10. Practice before you present.

Consider trying presentation software such as Prezi, which allows a viewer to see a whole visual representation before the embedded detail slides. The online presentation tool is at http://prezi.com/ and has tips and examples to help you organize your ideas and present your information effectively. Look at the examples listed below for ways to summarize and present data results visually.

Mapping examples

- Central Indiana interactive mapping: http://www.savi.org/savi/QuickInformation.aspx
- Indiana mapping and social networking: www.communitycommons.org
- U.S. County Profiles with maps of life expectancy, obesity, and sufficient physical activity: http://www.healthmetricsandevaluation.org/us-health/county-profiles

Data visualization examples

www.indianacancer.org/category/blog/infograph/
www.informationisbeautiful.net/visualizations/
www.gapminder.org/
www.dashboard.imamuseum.org/
http://www.healthmetricsandevaluation.org/tools/data-visualizations
Summary

This chapter reviews key concepts in analyzing, interpreting, and disseminating data and suggests ways to match the data and its message to the intended audience. It focuses on your essential role in surveillance and communication.

- Surveillance: Keep examples of data products and presentation, regardless of the topic, to help you quickly and meaningfully disseminate results of chronic disease surveillance regularly and widely in a variety of formats.

- Communication: If simply communicating complex or statistical information is not your strength, consider one of these suggestions. Read science writers who communicate technical information in everyday language. Review the resources in the chapter as often as needed. Review CDC’s website.

- Consultation: Talk to your health communication specialist in your department.