PUBLIC HEALTH INSTITUTE
OF WESTERN MASSACHUSETTS
PARTNERS FOR HEALTH EQUITY

Data and Action Framework:

Interpreting and Contextualizing Data

This guide aims to provide you with some best practices for understanding and interpreting quantitative data such as from the COVID-19 Community Impact Survey (CCIS) and Community Health Equity Survey (CHES), as well as other data sources. By following these recommendations, you can enhance your data literacy and make well-informed decisions based on reliable information.

Do's:

- ✓ Be curious and critical when examining the data source.
- ✓ Consider who is represented by the data to ensure it captures the experiences of diverse populations.
- ✓ Analyze the data with an awareness of potential biases and consider who may not be represented due to methods of data collection.

Don'ts:

- ✓ Don't solely rely on one data source for understanding an issue. Look for additional context whether that be other related data or input from community members with lived experience.
- ✓ Avoid overlooking potential biases in data analysis and interpretation.
- ✓ Don't assume that the data represents the entire population without considering who shared their data and the methods used to collect the data.
- ✓ Avoid making hasty conclusions without considering the broader context surrounding the data.

Understanding the Data Source:

- **Who collected the data and why?** Identify the individuals or organization(s) who collected the data and the purpose behind the data collection. This can provide context about what researchers wanted to understand from the data. Consider the assumptions, hypotheses, or theories of the researcher that might bias the data.

For background on CCIS and CHES, please visit the [MDPH Community Health Equity Initiative.]()

- **How was the data collected?** The way the data was collected can give insight into who is represented by the data demographically. Surveys, for instance, may be conducted through in-person interviews, online forms, telephone calls, and other ways. Each method of collecting data can reach different sections of the population that may be different by age, race, gender, sexual orientation, etc.

    - Consider who might be missing from the data due to sampling methods and participation barriers, such as language needs or access to technology.
- **How were people identified and selected for data collection?** There are different sampling strategies researchers can use to determine how people are identified and selected to participate in a survey or other data collection strategies. They each have their own benefits and limitations. The following are two common types to consider.

<table>
<tr><th colspan="2">Convenience sample</th></tr>
<tr><td colspan="2">With this strategy, researchers collect data from participants who they have access to and are willing to participate. The CCIS administered in 2020 was an example of this because the survey was sent out through various networks and was open to anyone in Massachusetts to participate. Findings represent those who participated, and there is limited ability to draw conclusions or make inferences to broader populations without carefully considering how participants were identified and selected and who they represent.</td></tr>
<tr><td>Benefit
There can be concerted effort to intentionally get to specific communities that are typically underrepresented in survey data and other population-based data collection methods.</td><td>Important consideration
There is very little control over who participates with this strategy; therefore, the sample is biased, meaning that those who participate are different than those who do not, which impacts your findings. This is a form of bias known as sampling bias.
“Sampling bias describes the scenario in which some individuals within a target population are more likely to be selected for inclusion than others. For example, if participants are asked to volunteer for a study, it is likely that those who volunteer will not be representative of the general population. This threatens the generalizability of the study results because volunteers tend to be more health conscious than the general population.”[1]
When the distribution of the sociodemographic information of the survey respondents does not match that of the general population - There are statistical methods to account for differences in the sociodemographic make-up of survey respondents compared to the general population (e.g. weighting). However, these methods do not account for biases in who participated.</td></tr>
</table>

<table>
<tr><th colspan="2">Random sample</th></tr>
<tr><td colspan="2">In this strategy, participants are randomly selected from a specific population to participate in the data collection process. An example of this is conducting a survey in a neighborhood by visiting every third house.</td></tr>
<tr><td>Benefit</td><td>Important consideration</td></tr>
</table>

---

[1] https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/biases

Randomization reduces sampling bias and enables conclusions and inferences to be made from the study sample to the population of interest. There are benefits in the data analysis process to help researchers feel more confident that their data is generalizable, or applicable to the general population.	You may miss important perspectives from historically and systemically underrepresented populations.

Interpreting the Data:

- **Who responded to your survey?** Consider who responded to the survey compared to the population you are interested in prioritizing for action. If they are significantly different, consider using additional data sources that include your population of interest.
- **What do other data sources say?** If there is any uncertainty about a data point, you can use **data triangulation.** Data triangulation is the process of looking at different sources of information to understand if the data point is reasonable and make sense compared to the other sources. Triangulation provides context that can help with telling the full story about the data.
- Consider **ground truthing** the findings from the data. Ground truthing is the process of walking through the findings with people with lived experience to gain a deeper understanding and gather important context for the findings. Ground truthing can provide context such as how much the findings resonate as true, if they reflect the lived experience of your population of interest (or not), or other important information.

Resources:

  - Introduction to Triangulation (UNAIDS)
  - The Truth, the Whole Truth, and Nothing but the Ground-Truth: Methods to Advance Environmental Justice and Researcher–Community Partnerships

- **Break down the data to look at different groups (disaggregate) to identify disparities**, or differences, between population groups. You may disaggregate data by race, gender, geography, income, employment status, and many other factors. Understanding these differences is important for beginning to explore how health inequities may manifest in the data. However, be cautious of the narrative developed solely on the disparities in the disaggregated data. There are other factors and context that should be considered to get the full picture. See next section on "Contextualizing Data".
  - "To disaggregate data is to break down combined summary (or aggregated) data into smaller units of data... In order to be able to disaggregate data, you must have access to the variables or pieces of data that you want to disaggregate by." [2] For example, if you want to look at access to services by race, you would need to look at the "access to services" variable for each racial

---

[2] https://react-data-decoder.vercel.app/

category you are examining. Therefore, you need access to racial identity data in the data set. Sometimes, data sources will have disaggregated data in them.

- **When determining what you can and cannot say about the data once you've analyzed it:**
    - **Be mindful of recall/response bias and social desirability bias** when interpreting self-reported data. Data from CCIS and CHES is self-reported by the participants.

Recall/Response Bias	Social Desirability Bias/ Reporting Bias
This form of bias happens when participants are asked questions about events and experiences that happened in the past and participants may not be able to accurately recall them. As a result, participants may respond in either an underestimate or overestimate of the experience, which can impact the results.[3]	This bias refers to when participants respond to questions with the option that makes them look good to others. For example, youth may report low frequency of drinking alcohol because they know it is not a desirable behavior for their age.[3]

-
    - **Use extra caution when trying to draw conclusions from data based on small numbers** of respondents for a given question or among a subpopulation of interest.

Contextualizing the Data:

- Examine the **broader context** surrounding the data, considering demographic, social, economic, and cultural factors that may influence the findings. For example, referring to the Data and Action Framework Narrative, the COVID-19 pandemic led to changes in many people's behaviors. Data collected during this time may be very different from previous years.
- Go beyond descriptive analysis **to explore the root causes** and drivers of health disparities.
- Consider framing the data with background about the historical, social, economic, and systemic factors that contribute to inequities.
- As stated earlier, ground truthing the data can provide important context directly from people with lived experience.

Resources:

- Analyzing Root Causes of Problems: The "But Why?" Technique
- The Storytelling Power of Numbers

---

[3] https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/biases

Communicating Findings:

- **Use plain language and avoid technical jargon**. For example, the term "social determinants of health" is not understood by all people. Instead, you might consider saying something like, "social factors, such as neighborhood safety or education, that impact one's health."
- **Use visual aids, storytelling, and community examples** to make the data more accessible and relatable.
- **Use asset-based framing**. Asset-based framing means to frame the data, and the story you would like to tell with the data, with a focus on assets, solution-based approaches, and/or inclusive calls to action rather than focusing solely on the needs or what is wrong. As repeatedly shared with us in our work with data in the community, it is disheartening to repeatedly hear how badly a racial/ethnic group or community you belong to is doing. We often do this to show need, but don't realize that we are inadvertently causing harm and potentially impeding action by perpetuating a sense of hopelessness. Starting with assets (e.g. great work going on in the community to address a given issue, anecdotal stories of change) or potential calls to actions first helps to support action.

Resource:

Anchor and Credential Solutions, Not Problems (The Case Made)

- **Connect and share back with community partners** that contributed to the process of ground truthing, meaning-making of the data, or any other related efforts. This requires planning ahead in the design phase of your process to be able to collect contact information and input on the best communication methods to reach your community partners.

By following these recommendations and taking health equity into account, you can interpret and contextualize data in a meaningful way, leading to better-informed actions.