



# Ethics in Data Collection and Use

## A Guide for Prevention Professionals

Kyle D. Barrington, Ph.D.  
CAPT Associate, West Resource Team

This training was developed under the Substance Abuse and Mental Health Services Administration's Center for the Application of Prevention Technologies task order. Reference #HHSS283201200024I/HHSS28342002T.

The views expressed in this training do not necessarily represent the views, policies, and positions of the Substance Abuse and Mental Health Services Administration or the U.S. Department of Health and Human Services.

# Agenda – Our Roadmap

- Welcome and learning objectives
- Defining ethics
- Types of data
- Brief history of data collection
- Legal and ethical considerations

# Agenda – Our Roadmap

- Vulnerable populations
- Ways to collect de-identified data
- Presenting data in reports
- Wrap-up

# Objectives

- Analyze and apply the definition of ethics
- Explore the different types of data
- Discuss the history of data collection
- Describe the legal and ethical issues related to data collection
- Identify vulnerable populations
- Contrast the different ways to de-identify data
- Summarize how to use data in reports

# Presenter



**Kyle Barrington**  
CAPT Associate

Kyle provides technical assistance and training to community-based, governmental, and tribal organizations. He has conducted 175 program evaluations using a 'mixed-methods' approach and his work has been published in peer-reviewed journals.

# Pre-Test Question 1

## Why are British sailors called Limeys?

- A. Because a favorite port of call was Lima.
- B. Because they ate limes.
- C. To honor the first Admiral of the Fleet, William de Limebourne, sailors are called “Limeys” in his honor.

# Pre-Test Question 2

**If you collect information about a participant's substance misuse history you must comply with HIPAA.**

A. True

B. False



# Pre-Test Question 3

**The first rule of data collection is...**

- A. Ensure the data are valid
- B. Collect signed, active consent forms
- C. Do no harm
- D. Seek grantor approval

# Pre-Test Question 4

**Which federal law requires consent forms to be used even when collecting anonymous data?**

A. PPRA

B. FERPA

C. HIPAA

D. None, no federal law covers anonymous data

# Pre-Test Question 5

**If you survey middle school students regarding alcohol, tobacco or other drug use you should seek \_\_\_\_\_ from the students:**

- A. Active consent
- B. Passive consent
- C. Assent agreement
- D. All of these
- E. None of these as parents must give consent

# Pre-Test Question 6

**Post office zip codes are considered what kind of data?**

- A. Personally identifiable information (PII)
- B. Personal health information (PHI)
- C. Personally protected information (PPI)
- D. None of these

# Pre-Test Question 7

**For practitioners, the best practice is to \_\_\_\_\_ the data.**

- A. De-identify
- B. Disaggregate
- C. Blur
- D. Redact

# Ethics

## What do We Mean by Ethics?



# Definition and Usage

- “The rules of conduct recognized in respect to a particular class of human actions or a particular group or culture”<sup>1</sup>
- Ethics governs professional interactions
- Morality governs personal interactions
- Laws govern society as a whole

# Practitioner's Notes - Ethics

Ethics govern \_\_\_\_\_ interactions.

Morals govern \_\_\_\_\_ interactions.

Legal, or laws, govern \_\_\_\_\_ interactions.





# Practitioner's Notes - Ethics

Ethics govern **professional** interactions.

Morals govern **interpersonal** interactions.

Legal, or laws, govern **social** interactions.



# Types of Data

## Identifying the Type of Data You Collect



# Data Types - Statisticians' Point of View

- Nominal
- Ordinal
- Interval
- Ratio

# Data Types - Evaluators' Point of View

- Qualitative
- Quantitative
- Mixed

# Data Types – Lawyers’ Point of View

- Protected personal information (PPI)
- Personal health information (PHI)
- Personally identifiable information (PII)
- Personally identifiable data (PID)

# Data Types - Practitioners' Point of View

- Confidential
- Sensitive
- Public

# Practitioner's Notes - Data

There are many \_\_\_\_\_ of data.

\_\_\_\_\_ data can be shared with anyone.

\_\_\_\_\_ data should only be shared within the organization.

\_\_\_\_\_ data should only be shared with those who have a legal right to know AND in a \_\_\_\_\_ manner.



# Practitioner's Notes - Data

There are many **types** of data.

**Public** data can be shared with anyone.

**Sensitive** data should only be shared within the organization.

**Confidential** data should only be shared with those who have a legal right to know AND in a **secure** manner.





# History of Data Collection

## A Very Brief History



# Data Collection – Origins

- Biblical reference to a census
- Roman Emperors used survey data in 6 AD
- 1753 James Lind conducted first reported clinical trial
- 1805 Benjamin Rush published results of a qualitative study on alcohol misuse
- Straw polls first used in 1824 presidential election
- ‘Identified populations’ became common in 1943

# Data Collection – Consequences

- Landis' Facial Expressions Experiment (1924)
- Tuskegee Syphilis Study (1932 – 1972)
- Nazi Medical Experiments (1939 – 1945)
- Oak Ridge National Laboratory (1942 – 1945)
- Willowbrook Study (1963 – 1966)
- Humphrey's "Tearoom sex" study (1960s)
- Major League Baseball's Mitchell Report (2008)

# Practitioner's Notes - History

Data collection is nearly as old as \_\_\_\_\_ history.

Lind's initial study, to determine a treatment for scurvy, had \_\_\_\_\_ participants.

Most abuses related to data collection occurred in times of \_\_\_\_\_ and were \_\_\_\_\_.



# Practitioner's Notes - History

Data collection is nearly as old as human history.

Lind's initial study, to determine a treatment for scurvy, had 12 participants.

Most abuses related to data collection occurred in times of peace and were unintentional.



# Legal and Ethical Considerations

- **Health Insurance Portability and Accountability Act of 1996 (HIPAA)**
- **Family Educational Rights and Privacy Act (FERPA)**
- **Protection of Pupil Rights Amendment (PPRA)**

# HIPAA Guidelines



# Health Insurance Portability and Accountability Act (HIPAA)

- Created to protect PHI
- Applies to covered entities who electronically transmit claims transaction information
- HIPAA typically does not apply to prevention programs or community coalitions, but following HIPAA's privacy and safety rules is considered a "best practice"



# HIPAA and Data Collection

- Two types of data:
  - Identifiable
  - De-identified
- HIPAA is concerned only with identifiable data.
- When it comes to de-identified data, there are no restrictions to the use or disclosure.

# Complying with HIPAA

- Consent form that expressly states what data is collected and when that data may be disclosed.
- Denote specific situations when you are legally required to violate HIPAA.
- Secure a Release of Information (ROI) prior to release of PHI.

# Federal Educational Rights and Privacy Act



# FERPA Guidelines

- Protects the privacy of student education records.
- Requires written permission from the parent or eligible student prior to releasing education records.
- Exceptions include evaluation purposes.

# FERPA and Data Collection

- Two types of data:
  - Identifiable
  - De-identified
- FERPA is concerned only with identifiable data
- When it comes to de-identified data, there are no restrictions to the use or disclosure

# Complying with FERPA

- Request consent prior to beginning the service.
- Ensure parents sign consent forms for release of information if your efforts are not exempt.
- Secure assent forms from students participating in the program.
- Ensure parents and students understand how you will use the data.

# Protection of Pupil Rights Amendment (PPRA)



# PPRA

# PPRA Guidelines

- Applies to any school receiving U.S. Department of Education (ED) funding.
- Governs administration of survey, analysis or evaluation that concerns protected questions.
- Governs the use of consents.
- Provides protection for eight protected areas.



# Eight Protected Areas in PPRA

1. Political affiliations or beliefs of the student or the student's parent
2. Mental or psychological problems of the student or the student's family
3. Sex behavior or attitudes
4. Illegal, anti-social, self-incriminating, or demeaning behavior

# Eight Protected Areas in PPRA (cont.)

5. Critical appraisals of other individuals with whom respondents have close family relationships
6. Legally recognized privileged or analogous relationships, such as those of lawyers, physicians, and ministers
7. Religious practices, affiliations, or beliefs of the student or student's parent
8. Income (other than that required by law to determine eligibility for participation in a program or for receiving financial assistance under such program)

# PPRA and Data Collection

- If ED funding is used for the program being evaluated and if the student is required to submit the survey, you must use an active consent.
- If ED funding is not used for the program being evaluated or student is not required to submit the survey, you can use a passive consent.

# Complying with PPRA

- Determine if students are required to submit a survey, then determine if active or passive consents are required.
- Provide a copy of the survey being used for parents to review.
- Provide, upon request, copies of instructional materials.

# Discussion – Topic 1



A community coalition collects date of birth, name, and sexual orientation data on participants. Is this information protected under HIPAA?

# Discussion – Topic 2



You are implementing a school-based curriculum to improve students' life skills, funded by the U.S. Department of Education, and you administer an survey. Students must complete the survey for a grade. Can you use passive consents?

# Practitioner's Notes - Laws

There are \_\_\_\_\_ laws that cover data collection.

If you treat all client data as \_\_\_\_\_ you can avoid most legal and ethical concerns.

Only the \_\_\_\_\_ law specifically requires protection for anonymous data.



# Practitioner's Notes - Laws

There are many laws that cover data collection.

If you treat all client data as confidential you can avoid most legal and ethical concerns.

Only the PPRA law specifically requires protection for anonymous data.





# Vulnerable Populations

## Who They Are and How We Protect Them



# Vulnerable Populations

Special safeguards should be in place for:

- A. Pregnant women
- B. Medically compromised persons
- C. Children
- D. Cognitively impaired persons
- E. Prisoners
- F. Students and employees
- G. Educationally disadvantaged individuals
- H. Veterans<sup>2</sup>

# Consider the Consequences

1. I use drugs to win sports competitions.
2. What drugs do you currently use?
3. Have your parents hit or slapped you?
4. Which prison guard is the most abusive?
5. Do you like your teacher?

# Practitioners Notes – Vulnerable Populations

Because of previous \_\_\_\_\_ by \_\_\_\_\_, vulnerable populations need additional \_\_\_\_\_.

Persons classified as “vulnerable” may be more susceptible to the influence of \_\_\_\_\_.

When working with vulnerable populations a data collector should secure \_\_\_\_\_ and \_\_\_\_\_.



# Practitioner's Notes – Vulnerable Populations

Because of previous **abuses** by **data collectors**, vulnerable populations need additional **safeguards**.

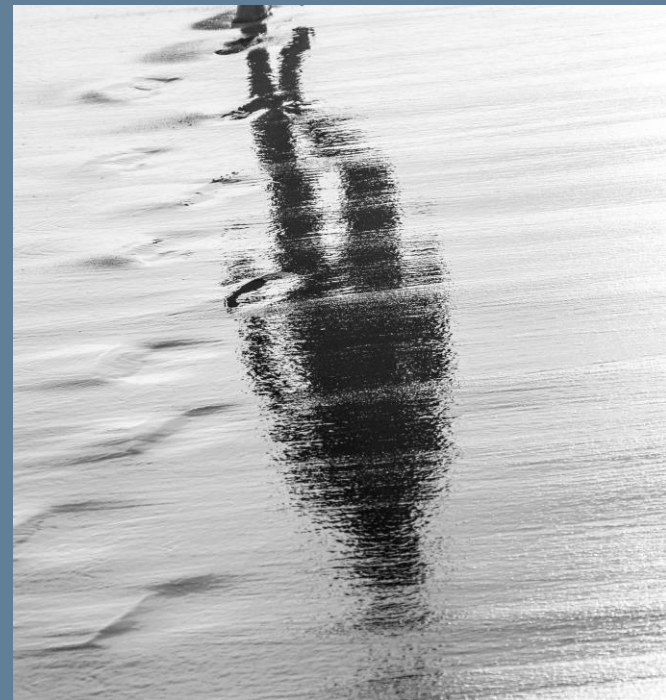
Persons classified as “vulnerable” may be more susceptible to the influence of **authority**.

When working with vulnerable populations a data collector should secure **consents** and **assents**.



# De-identifying Data

## Ways to De-identify Data



# De-identifying Data

- Is the process used to prevent a person's identity from being connected with a specific data point.
- Severs the link between the data and the data provider.
- Is mostly used when participant(s) change in knowledge, attitude, and/or behavior need to be tracked over time.

# Ways to De-Identify Data

- Random numbers
- Coded surveys
- Tear sheets/raffle tickets
- Pseudo codes
  1. Full answer
  2. Touch tone code
  3. Rubric



# Pseudo Code Practice - 1

What code do you get if you use full answer:

1. Your favorite pet's name?
2. Favorite teacher's last name?
3. Mascot of your high school?



# Pseudo Code Practice - 2

What code do you get if you use Touch Tone Code:

1. Code your full name



# Pseudo Code Practice - 3

What code do you get if you use a rubric:

1. What is the second letter of your mother's maiden name?
2. What is the fourth digit of your social security number?
3. What is the last letter of the name of the state in which you were born?
4. What is the last digit of your cell phone number?
5. What is the second letter of the month in which you were born?
6. How many brothers do you have?
7. How many sisters do you have?



# Practitioner's Notes – De-Identified

The need to track a specific participant's change in knowledge, attitude, and/or behavior \_\_\_\_\_ creates issues for data collectors.

Creating a \_\_\_\_\_ allows a data collector to \_\_\_\_\_ between the \_\_\_\_\_ and the \_\_\_\_\_.

To meet this challenge many data collectors rely on \_\_\_\_\_.



# Practitioner's Notes – De-Identified

The need to track a specific participant's change in knowledge, attitude, and/or behavior **over time** creates issues for data collectors.

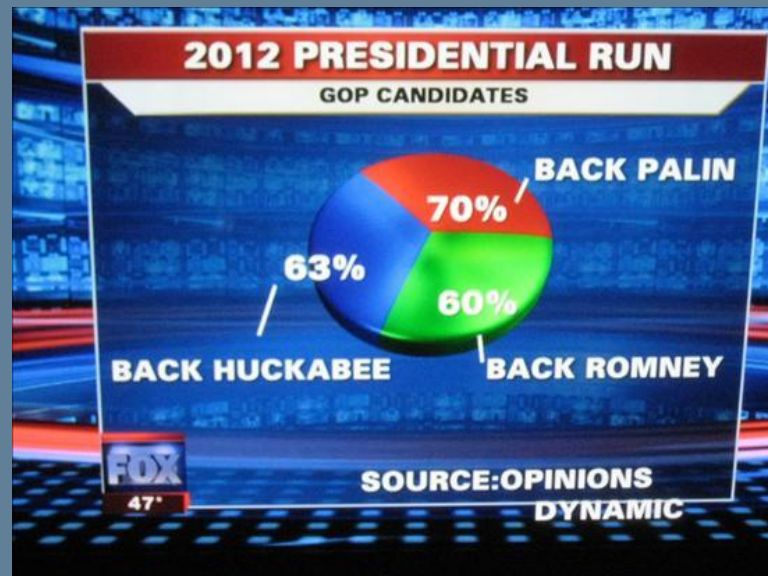
Creating a **pseudo code** allows a data collector to **sever the link** between the **data** and the **data provider**.

To meet this challenge many data collectors rely on **de-identified data**.



# Presenting Data in Reports

## Ethical Ways to Present Data



# Report Complete Numbers

- Identify the total number of eligible participants.
- Identify the total number of actual participants.
- Identify the number who participated in the evaluation of the program.

# Would this make a difference?

- 80.0% of participants reported decreases in past 30-day binge drinking rates.
  - Total number of eligible participants = 15,200
  - Total number of actual participants = 25
  - Total number of participants in evaluation = 10



# Use the Same Math

- If you are presenting longitudinal data, it is important you calculate each year's data in the same manner
- Ensure you use the same data points for your denominator and numerator for each calculation

# Would this make a difference?

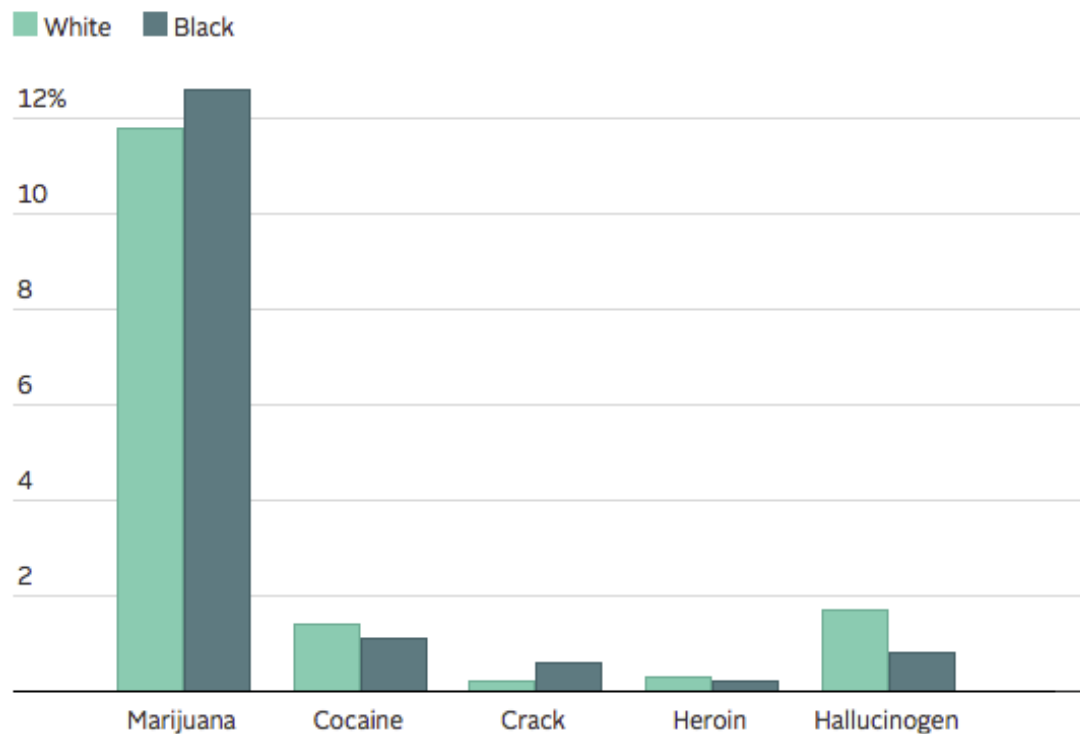
- We achieved a 20.0% reduction in truancy rates in the past three years.
  - In year 1 they divided by total school days.
  - In year 2 they divided by total teacher days.
  - In year 3 they divided by total administrator days.

# Use Appropriate Graph Labels

- Always denote if you use truncated charts.
- Use a Y-axis that measures the full extent of the scale.
- Using smaller scales can inadvertently magnify small differences.
- Avoid merging X-axis categories.

# Does this make a difference?

## Percent of population who used a drug in the past year

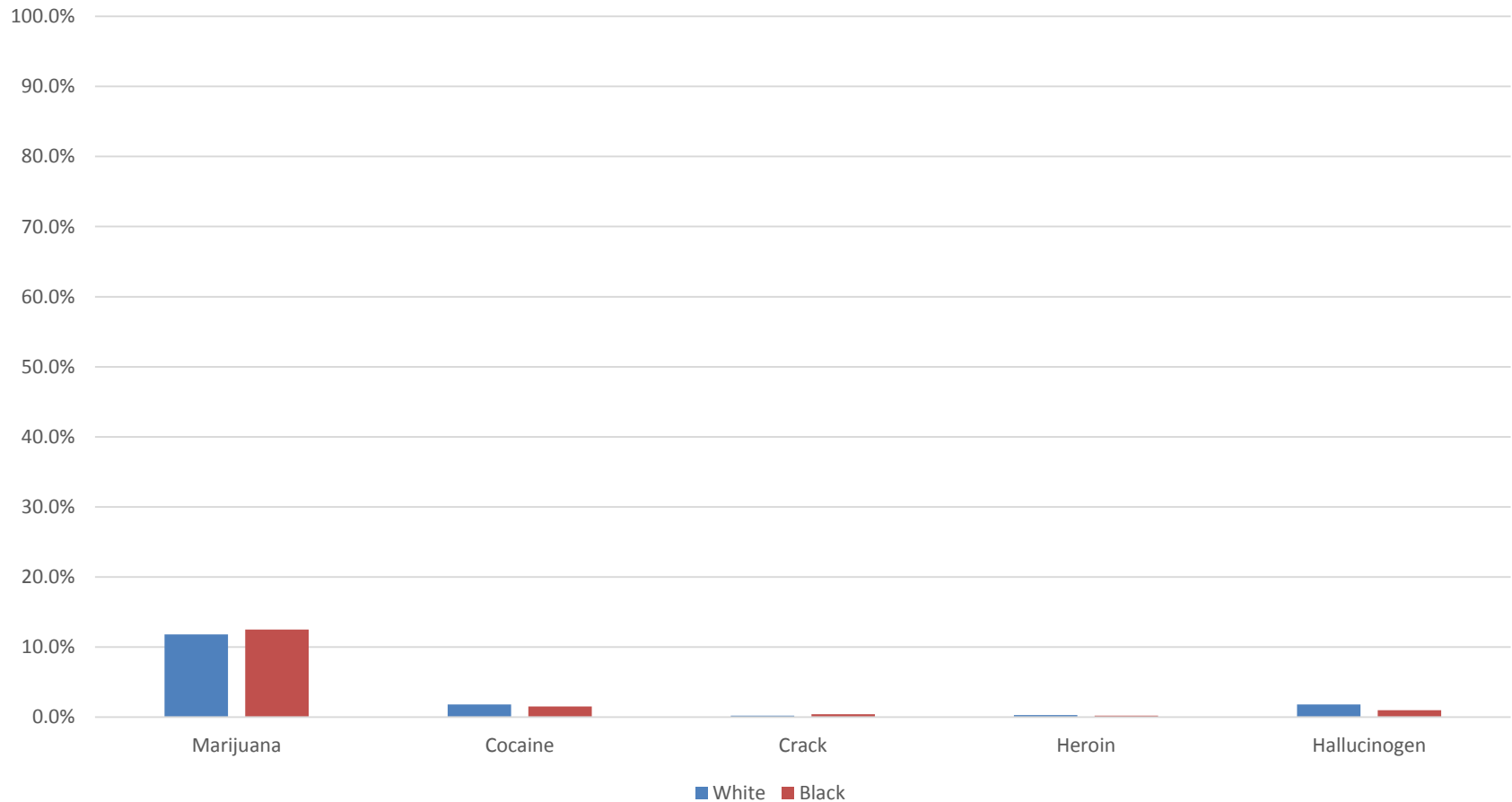


Source: SAMHSA, 2011 data

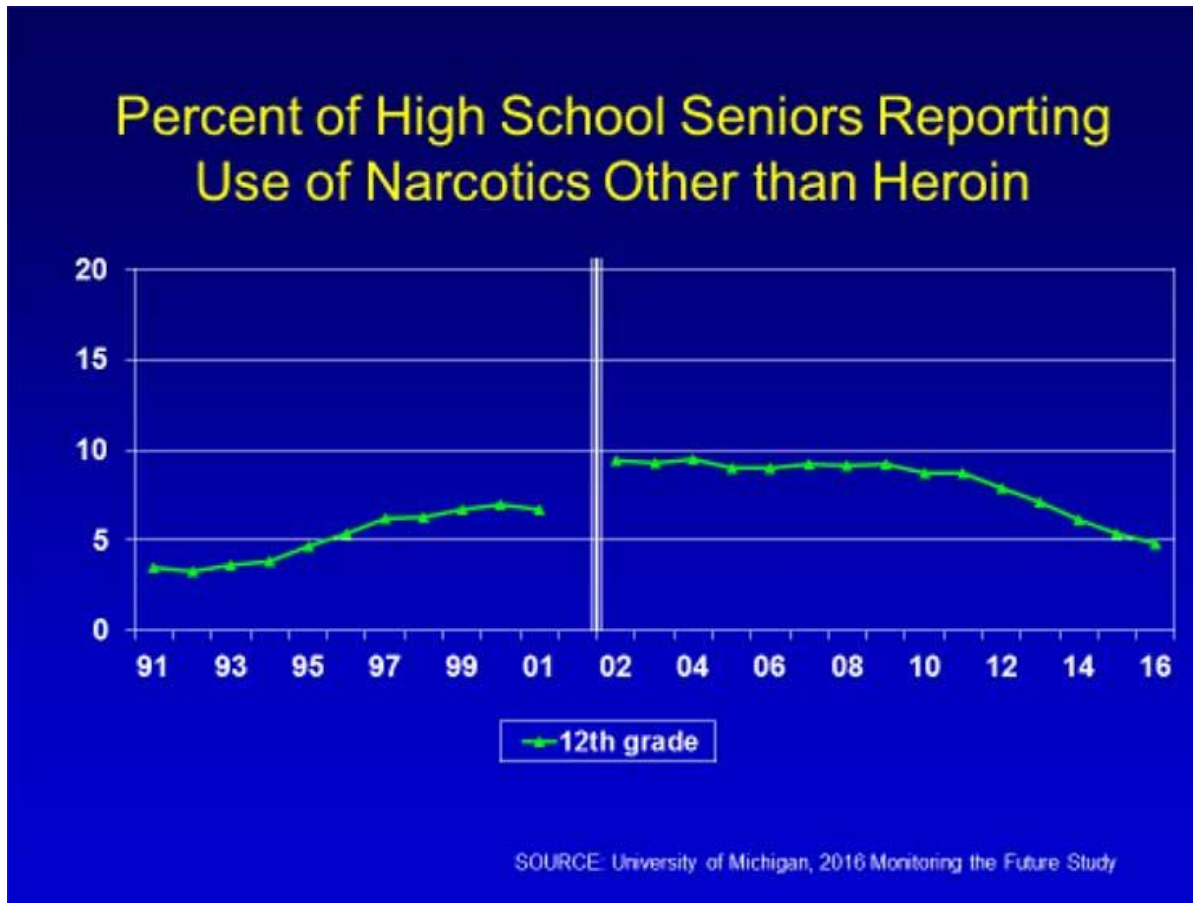


# With Full Y-Axis View

Percentage of population who used a drug in past year

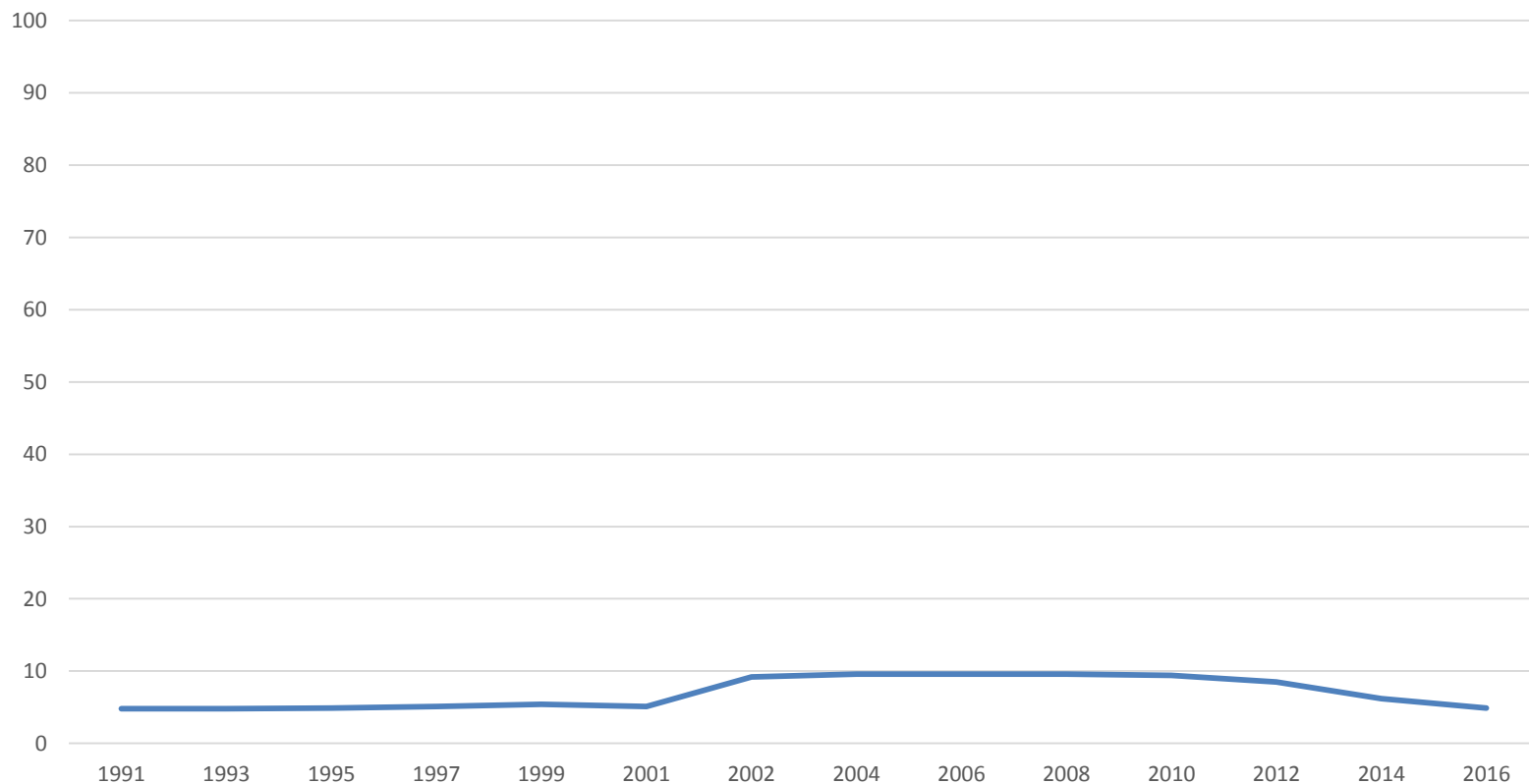


# Does this make a difference?



# With Full Y-Axis View

## Percentage of High School Seniors Reporting Use of Narcotics Other than Heroin



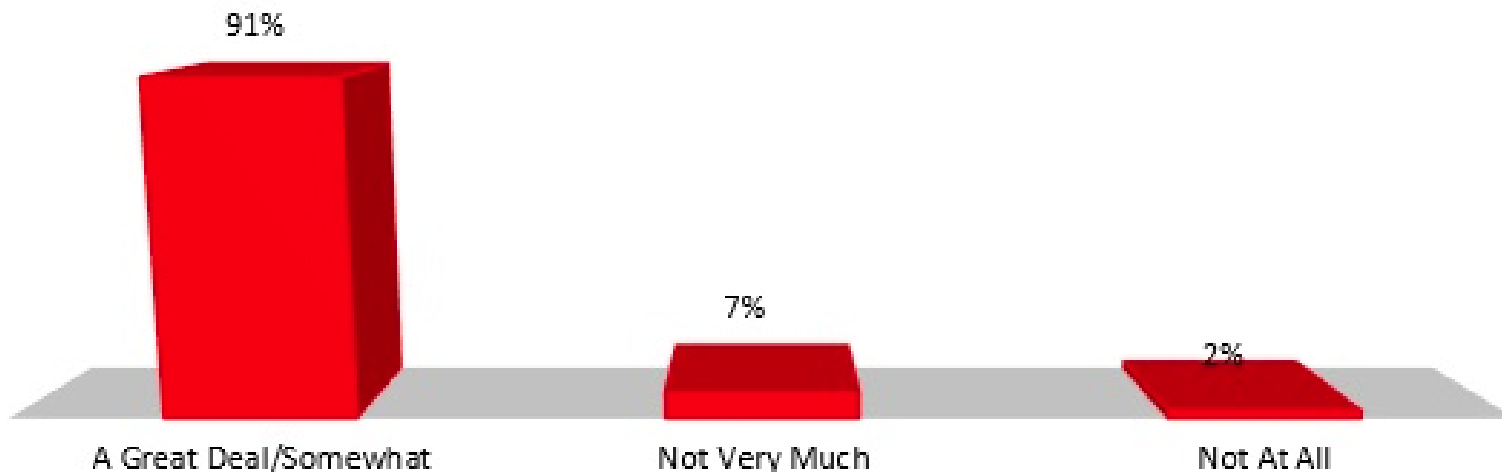
# Use Full X-Axis Categories

- If you ask participants to report by a specific category, you should report by specific categories.
- Merging categories can lead to false conclusions.
- Once each category is fully reported, then you can merge.



# Definition Please

**DID THE MYTI POSITIVELY IMPACT  
YOUR EFFECTIVENESS AT ADDRESSING  
SUBSTANCE ABUSE?**



(Community Anti-Drug Coalitions of America, 2017)

# Practitioner's Notes – Reporting

Most ethical issues related to data collection occurs in the \_\_\_\_\_ of that data.

Using a different \_\_\_\_\_ between pre- and post-test reports is almost always \_\_\_\_\_.

Most ethical concerns in data reporting occur in the use of \_\_\_\_\_.

A practitioner should always \_\_\_\_\_ if a chart is \_\_\_\_\_.



# Practitioner's Notes – Reporting

Most ethical issues related to data collection occurs in the **reporting** of that data.

Using a different **mathematical calculation** between pre- and post-test reports is almost always **unethical**.

Most ethical concerns in data reporting occur in the use of **charts**.

A practitioner should always **denote** if a chart is **truncated**.



# Summary

## Wrapping this all up



# Wrap-Up Question

**Did you learn something today about ethics and data collection that will help you in your prevention efforts?**

A. Yes, more than three things

B. Yes, one or two things

C. No, nothing at all

# Post-Test Question 1

## Why are British Sailors called Limeys?

- A. Because a favorite port of call was Lima.
- B. Because they ate limes.
- C. To honor the first Admiral of the Fleet, William de Limebourne, sailors are called “Limeys” in his honor.

# Post-Test Question 2

**If you collect information about a participants' substance misuse history – you must comply with HIPAA.**

A. True

B. False

# Post-Test Question 3

**The first rule of data collection is...**

- A. Ensure the data are valid
- B. Collect signed, active, consent forms
- C. Do no harm
- D. Seek grantor approval



# Post-Test Question 4

**Which federal law requires consent forms to be used even when collecting anonymous data?**

A. PPRA

B. FERPA

C. HIPAA

D. None, no federal law covers anonymous data

# Post-Test Question 5

**If you survey middle school students regarding alcohol, tobacco, or other drug use you should seek \_\_\_\_\_ from the students:**

- A. Active consent
- B. Passive consent
- C. Assent agreement
- D. All of these
- E. None of these as parents must give consent

# Post-Test Question 6

**Post Office ZIP codes are considered what kind of data?**

- A. Personally Identifiable Information (PII)
- B. Personal Health Information (PHI)
- C. Personally Protected Information (PPI)
- D. None of these

# Post-Test Question 7

**For practitioners, the best practice is to \_\_\_\_\_ the data.**

- A. De-identify
- B. Disaggregate
- C. Blur
- D. Redact

# Questions?



If you have questions or comments, please don't  
hesitate to contact:

Janet Porter

T/TA Specialist, CAPT West Resource Team

[jporter@casat.org](mailto:jporter@casat.org)

530-249-1790

If you have questions or comments, please don't  
hesitate to contact:

Kyle D. Barrington  
CAPT Associate

[Kyle.Barrington@Zajonc-Corp.com](mailto:Kyle.Barrington@Zajonc-Corp.com)

979-696-6373

# References

1. *Ethics*. (2017). In Dictionary.com. Retrieved from <http://www.dictionary.com/browse/ethics>
2. *Vulnerable populations*. (2017). University of North Carolina-Asheville. Institutional Review Board. Retrieved from <https://irb.unca.edu/vulnerable-populations>
3. Lostetter, M.J. (2012, March 14). Moral, ethical, legal: What's the difference? [Blog post]. Retrieved from <https://lostetter.wordpress.com/2012/03/14/moral-ethical-legal-whats-the-difference/>
4. Cdn2. (2012, January 24). *Percentage of population who used a drug in the past year* [digital image]. Retrieved from [http://cdn2.vox-cdn.com/assets/4690865/drug\\_use\\_by\\_race.png](http://cdn2.vox-cdn.com/assets/4690865/drug_use_by_race.png)
5. National Institute on Drug Abuse; National Institutes of Health; U.S. Department of Health and Human Services. (2016). *Monitoring the future survey: High school and youth trends*. Retrieved from <https://www.drugabuse.gov/publications/drugfacts/monitoring-future-survey-high-school-youth-trends>
6. Community Anti-Drug Coalitions of America. [Email]. (2017). Mid-year training institute: Engineering healthy communities. Mid-year makes a difference.