The Basics of Community Based Program Evaluation

Janet A. Welsh, Ph.D.

Bennett-Pierce Prevention Research Center

Pennsylvania State University

Workshop Outline

- Why Evaluate?
- Types of Community Based Evaluations
- Evaluation Goals & Targets
- Selection of Measures
- Interpretation of Findings
- Sharing Results
- Practice: Designing and Evaluation

Why Evaluate?

WE ARE ACCOUNTABLE!

Accountability Means…..

Are we using resources in responsible ways?

Is the prevention initiative having the intended impact?	Are we implementing the intervention correctly?	Is the intervention a good fit with our target population?	Can we explain why or why not the intervention had the intended impact?

We are often required to provide information to funders and other stakeholders (school boards, County Commissioners, etc) to document that what we are doing is positive and impactful

Accountability Means We Don’t Cause Harm

Remove Cellulite Naturally

Accountability Also Means We Don’t Waste Time and Money

- When we systematically evaluate the impact of our interventions, we are less likely to inadvertently do things that are: ineffective (BAD!) or cause harm (WORSE!)

Types of Evaluations: Program **Outcomes**

IS IT WORKING??

Outcome Evaluations

Outcome evaluations address questions of program impact

- Did the intervention work?
- Did some people benefit more from the intervention than others?
- Were some parts of the intervention more effective than others?
- Were some outcomes impacted and not others?

Outcome Evaluation Designs

- **Randomized Controlled Trial** (RCT)
  - Involves comparisons of outcomes between an intervention group and a control group, randomly assigned to condition
  - Usually includes some period of follow up (a few weeks or months to many years)
  - Provides the strongest causal inference that interventions work, because other factors are accounted for
  - Is almost never done in community settings, but we should look for it in research studies when we select a program
  - Usually requires a university-community partnership

“We have a great intervention that we have been using in our community for more than ten years. We are now being told that we should use evidence based programs. How can we make our intervention evidence based?”

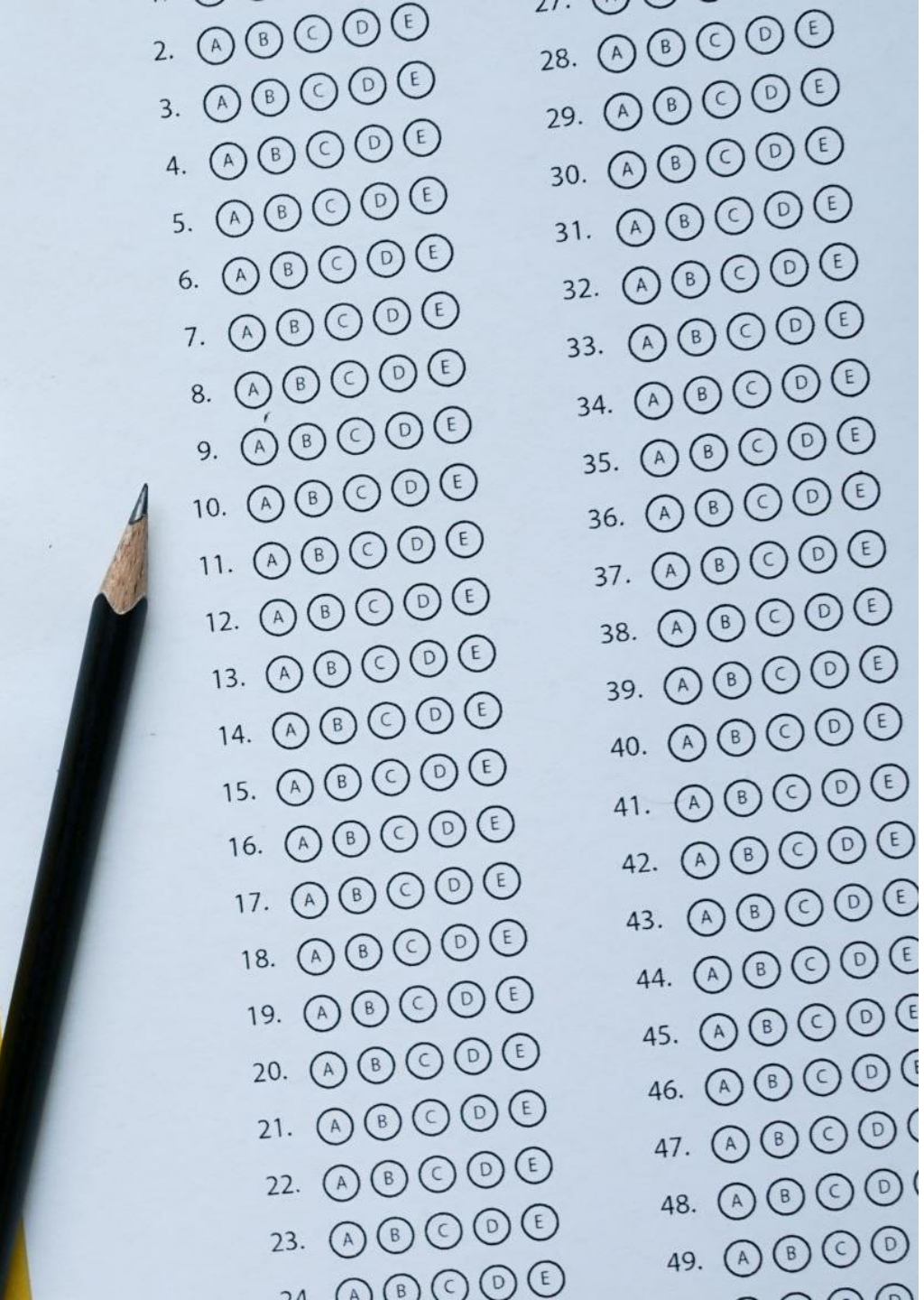
- Lengthy process (many years, especially if there is a follow up)
- Expensive evaluation

Quasi-Experimental Designs

- **Convenience Samples:**
  - Enroll people who are easy or want to participate
  - May or may not have a comparison group
  - “Wait list” comparison
  - Convenient, but……

Example: After-school enrichment programs

The San Antonio, TX school district wants to evaluate the impact of its after school homework support program for high school students. The program was offered to all low-achieving students in the district and parents could opt to have their students participate, or not. Attendance data showed that about 20% of the eligible students (those with GPA's below 2.0) participated. The school district used a quasi-experimental design to compare the grades and achievement test scores of students who participated in the program to those in the same grades who did not participate.

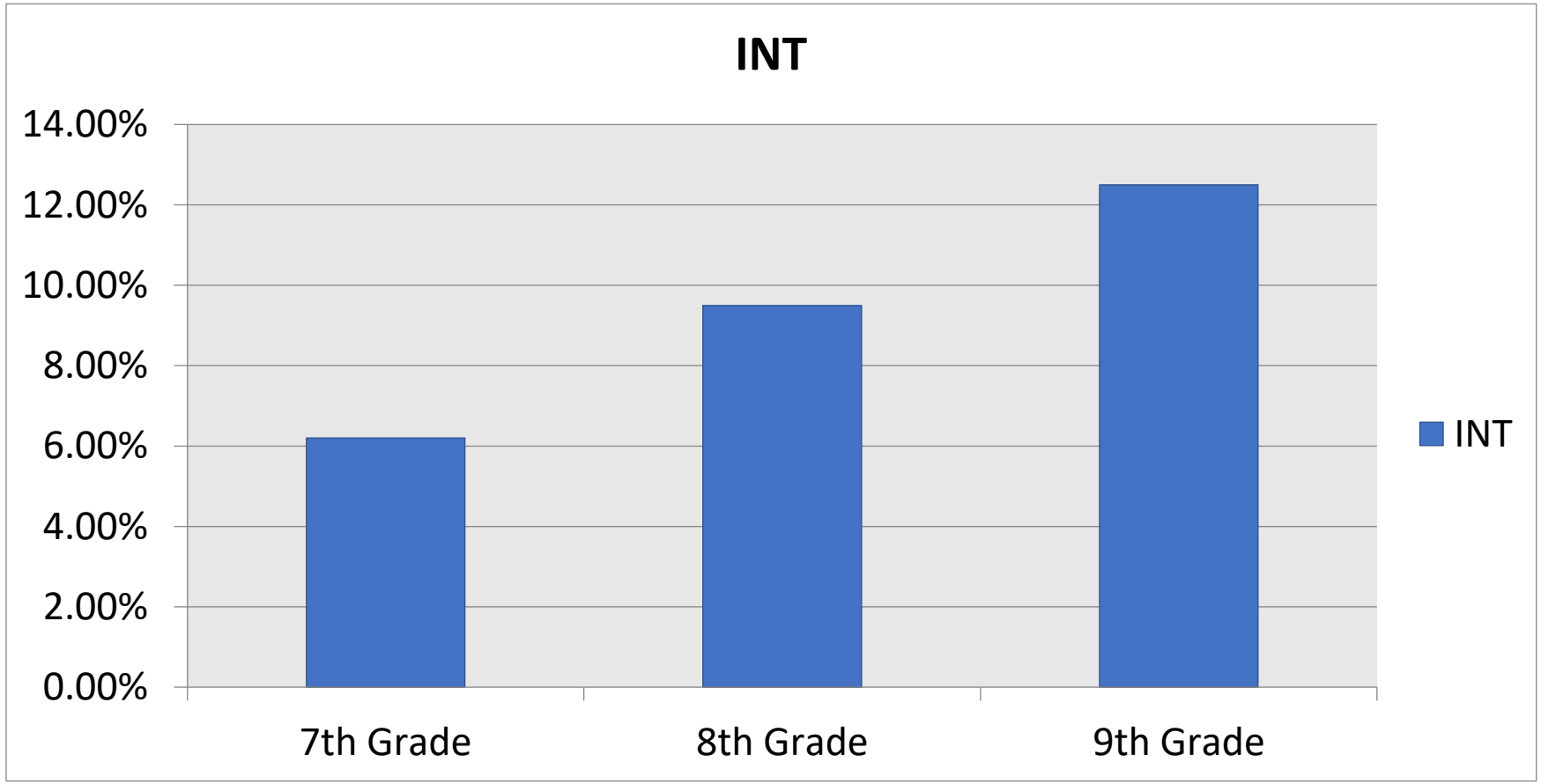


- **Pre/Post designs:**
  - Assessment – intervention – Assessment
  - Very commonly used in community settings
  - Easy to administer, BUT…..

Example: Substance abuse prevention in teens

The town of Verona, New Jersey decides to implement a school based intervention for middle school students designed to reduce use of alcohol and other substances. They administer a survey of substance use patterns before the program begins in 7th grade and after they have completed it, at the end of 9th grade.

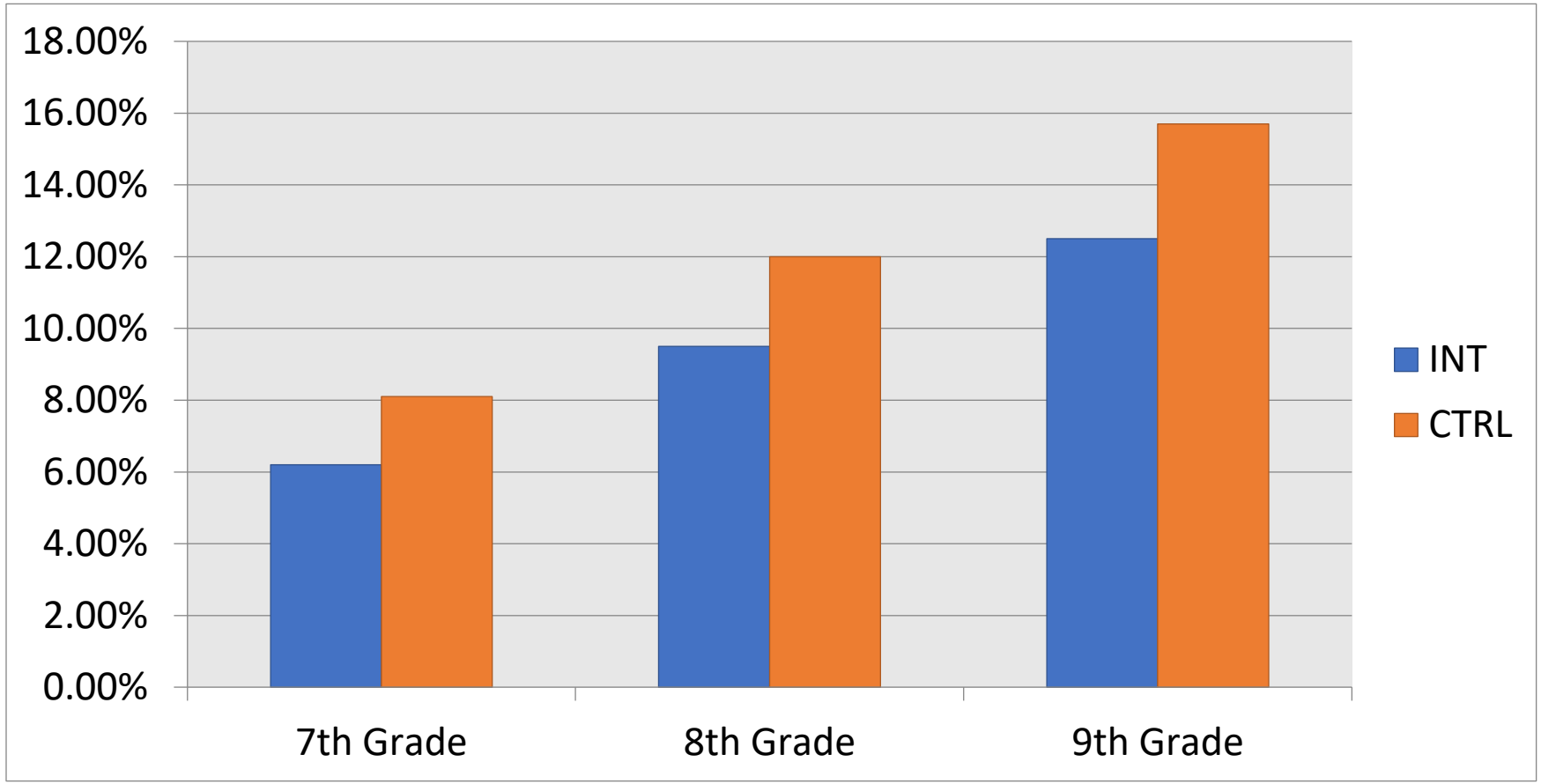
New User Inhalant Rates



All sig at p<.05

PROSPER Partnerships
Communities | Schools | Universities

New User Inhalant Rates



All sig at p<.05

PROSPER Partnerships
Communities | Schools | Universities

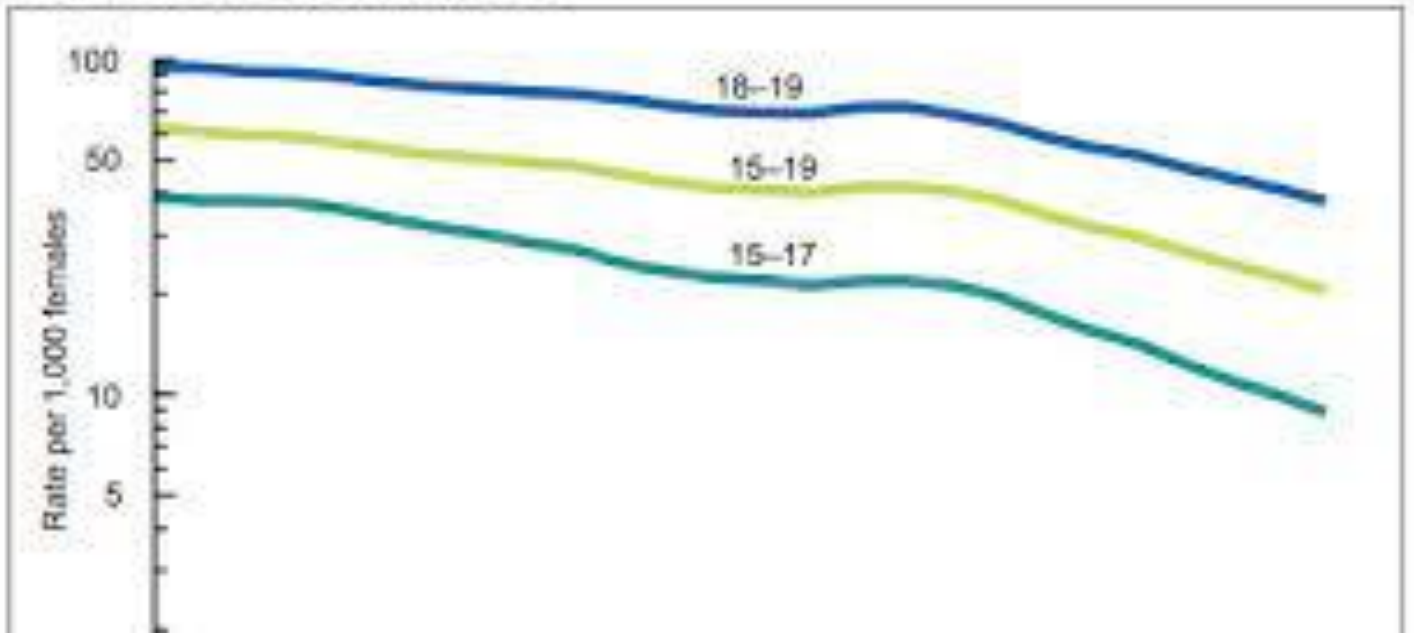
- Pre/post designs are not appropriate for all kinds of data and if not used carefully they can give the wrong message

- OK to use them if:
  - Change in the participants from pre to post was unlikely to have happened for other reasons (i.e., specialized knowledge)

- NOT OK to use them if:
  - The target behavior is likely to increase (or decrease) due to factors such as age, maturity, etc.

Teen Pregnancy

Figure 1. Birth rates for females aged 15–19, by age group: United States, final 1991–2015 and provisional 2016



Weak Designs: Don’t Bother

- **Post-then-pre**

Deliver the intervention, then ask the participants to rate:

1) Knowledge or behavior since participating in the intervention
2) Knowledge or behavior before participating in the intervention

- There are many factors that can bias the results of data collected in this way and cloud the actual impact of the intervention

- The benefit is reduced time spent in evaluation, but results can be hard to interpret and counterintuitive

Types of Evaluations: Implementation

Did we do it right?

Implementation Questions

- How many people came to the program?
- Did the right people receive the intervention?
- How often did they come?
- Were the deliverers trained?

- Were all of the program components fully delivered?
  - Dosage
  - Duration
  - Quality
  - Correct conditions

- Implementation and program outcome evaluations can and should be conducted together
- If programs don’t get their intended outcomes, it may be that implementation factors are to blame
- In rigorous research, low-quality implementation is related to diminished program effectiveness
- Studies of evidence based programs used in real world settings show that implementation often deviates from what is recommended, in both planned and unplanned ways.

Types of Evaluation: Feasibility

- Feasibility evaluations measure whether an intervention can actually be used in a new setting.

CAN WE DO IT?

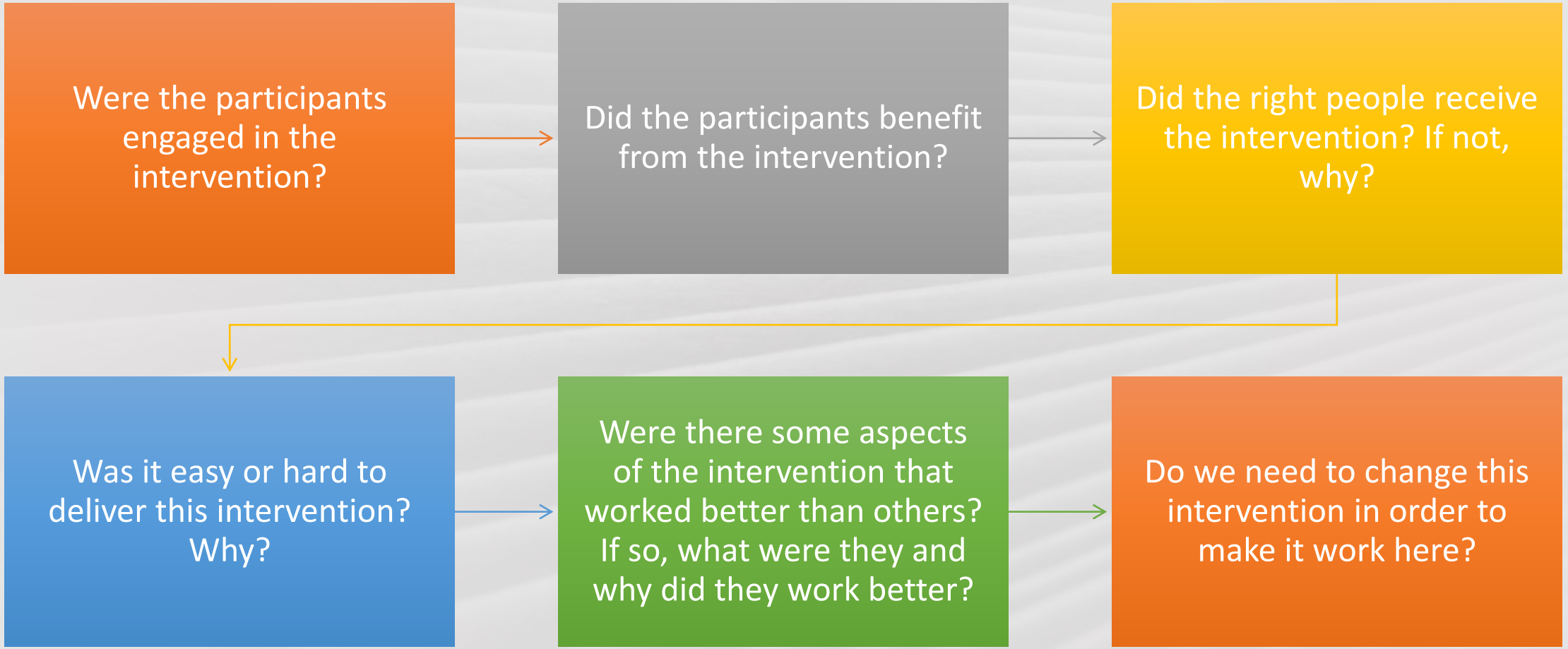
Feasibility evaluations can sometimes be referred to as “pilot” studies because we are trying something out for the first time, usually on a small scale.

Evaluation questions can include:

- Can we even do the program in this setting with this target population?
- Is this intervention acceptable to stakeholders and will they do it?
- Do we need to change it somehow in order for it to work better?

- In a feasibility evaluation, we are not really trying to see if an intervention “works” and we may not collect data on specific outcomes

- Instead, we might use focus groups or surveys to assess other factors


Were the participants engaged in the intervention?
Did the participants benefit from the intervention?
Did the right people receive the intervention? If not, why?
Was it easy or hard to deliver this intervention? Why?
Were there some aspects of the intervention that worked better than others? If so, what were they and why did they work better?
Do we need to change this intervention in order to make it work here?

Example: New Parent Support Program

The town of Cedar Rapids, IA recently decided to implement a home visiting program for high risk new parents. Obstetricians refer to the program any pregnant women they feel are at elevated risk for child maltreatment because of age, substance abuse history, significant mental illness, significant life stress or a history of IPV. The plan was to implement this on a wide scale and conduct a pre-post intervention evaluation of changes in parents' understanding of child development and self-reported frustration and anger with infant and toddler behaviors. However, a pilot study revealed that very few of the referred mothers actually enrolled and those who did often dropped out after three sessions. No one in Cedar Rapids is sure what to do now.

Understanding the “Why”: Focus Groups & Key Informant Interviews

- Qualitative evaluation that provides insight into specific findings, particularly those that are concerning, disappointing, or don’t make sense:
- *Why did our program work for boys but not for girls?*
- *Why don’t teachers like this program?*
- *Why did youth find session 4 of this program so difficult?*
- *What should we do if parents don’t want to do home visits?*

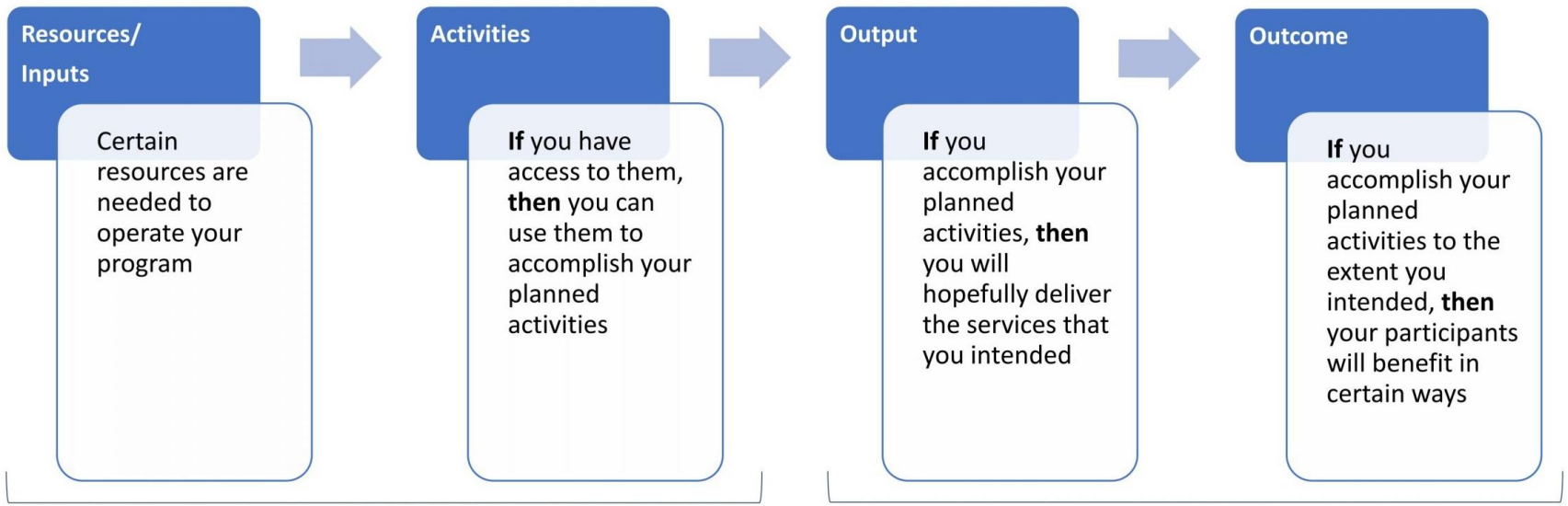
Choosing Evaluation Measures

- There are many considerations when choosing evaluation measures

- Measures should be closely aligned with the intervention's theory of change

- For this reason, all interventions should have a well-articulated logic model

Thinking of Logic Models as a Series of *If . . .Then* Statements



Adapted from: Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division for Heart Disease and Stroke Prevention (https://www.cdc.gov/dhdsp/evaluation_resources/guides/logic_model.htm).

Is It Logical?

- *Providing first graders with tutoring in phonics will improve their word decoding skills by the end of the school year*
- *Providing parents with a self-guided class on positive discipline and monitoring will prevent delinquent behavior in teenagers*
- *Showing youth dramatic reconstructions of accident scenes will reduce their use of alcohol and illicit substances*
- *Delivering a workshop on children's language development to preschool teachers will increase teachers' knowledge of language development in young children*

Use Established Measures When possible

Established measures have usually been checked for psychometric properties such as validity and reliability

Many EBPs have evaluation measures provided by the developer

Even when not using an EBP, established measures exist for many outcomes of interest in the field of prevention

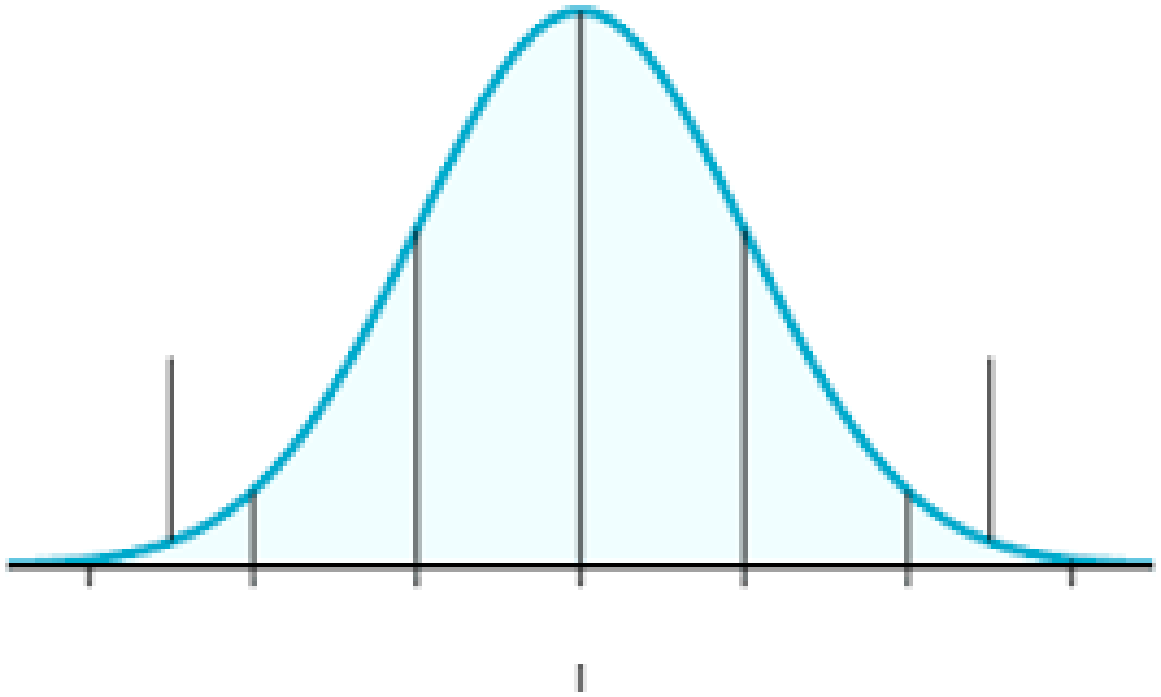
EPIS

Academic research partners

If using made-up measures, be careful!

- Lack of variability
  - **Ceiling effects**: Everyone “maxes out” resulting in a lack of variability in individual scores
  - **Floor effects:** The measure aims “too high” for most participants, and they don’t even achieve the lowest score
- Ethics & legal implications
  - Criminal acts
  - Child maltreatment
  - Domestic violence
  - Self-harm

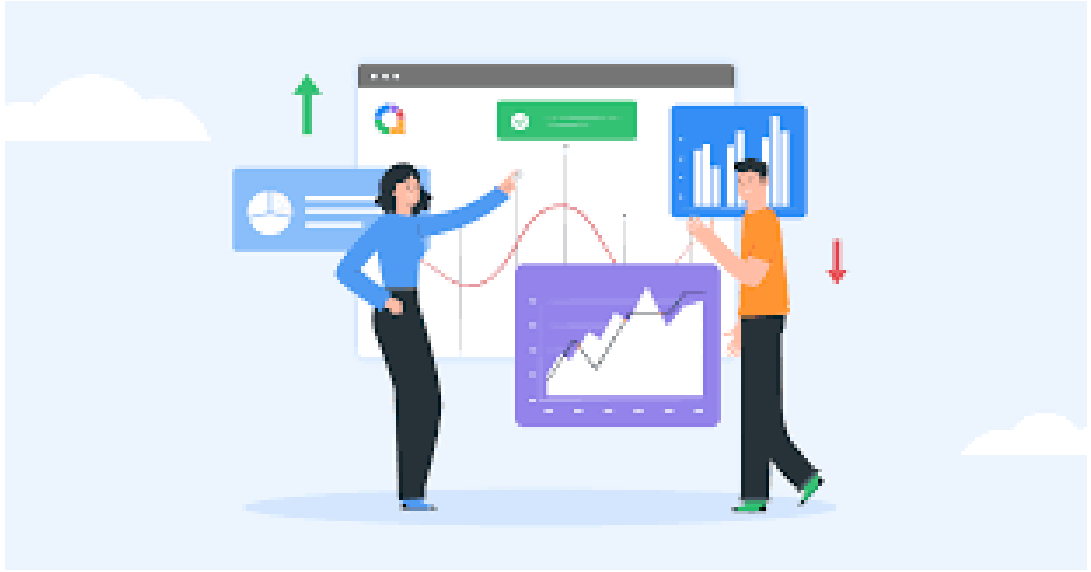
Normal Distribution



Practical Considerations for Measurement

- Language: be mindful of the first language and reading level of the participants
  - Children younger than 3rd grade can't reliably complete surveys without help
  - Most measures for parents or other community members should strive for about a 6th grade reading level
  - Participants should complete measures in their native language whenever possible
- Participant burden: how often are we asking people to complete measures, and how long does it take?

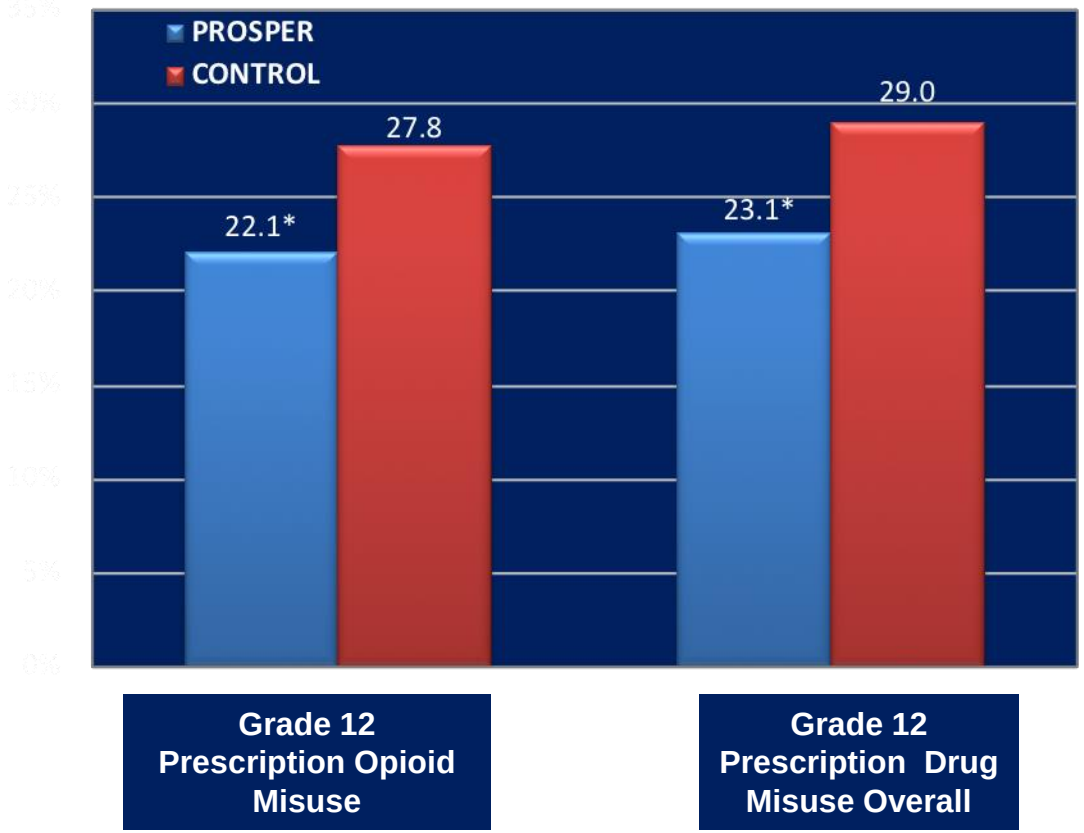
- Cost: Evaluation can be expensive & time intensive
  - Finding and preparing measures
  - Collecting data
  - Preparing data for analysis, checking for errors, etc.
  - Analyzing the data
  - Summarizing and presenting the data to stakeholders



- Ethics: Evaluations must be done carefully

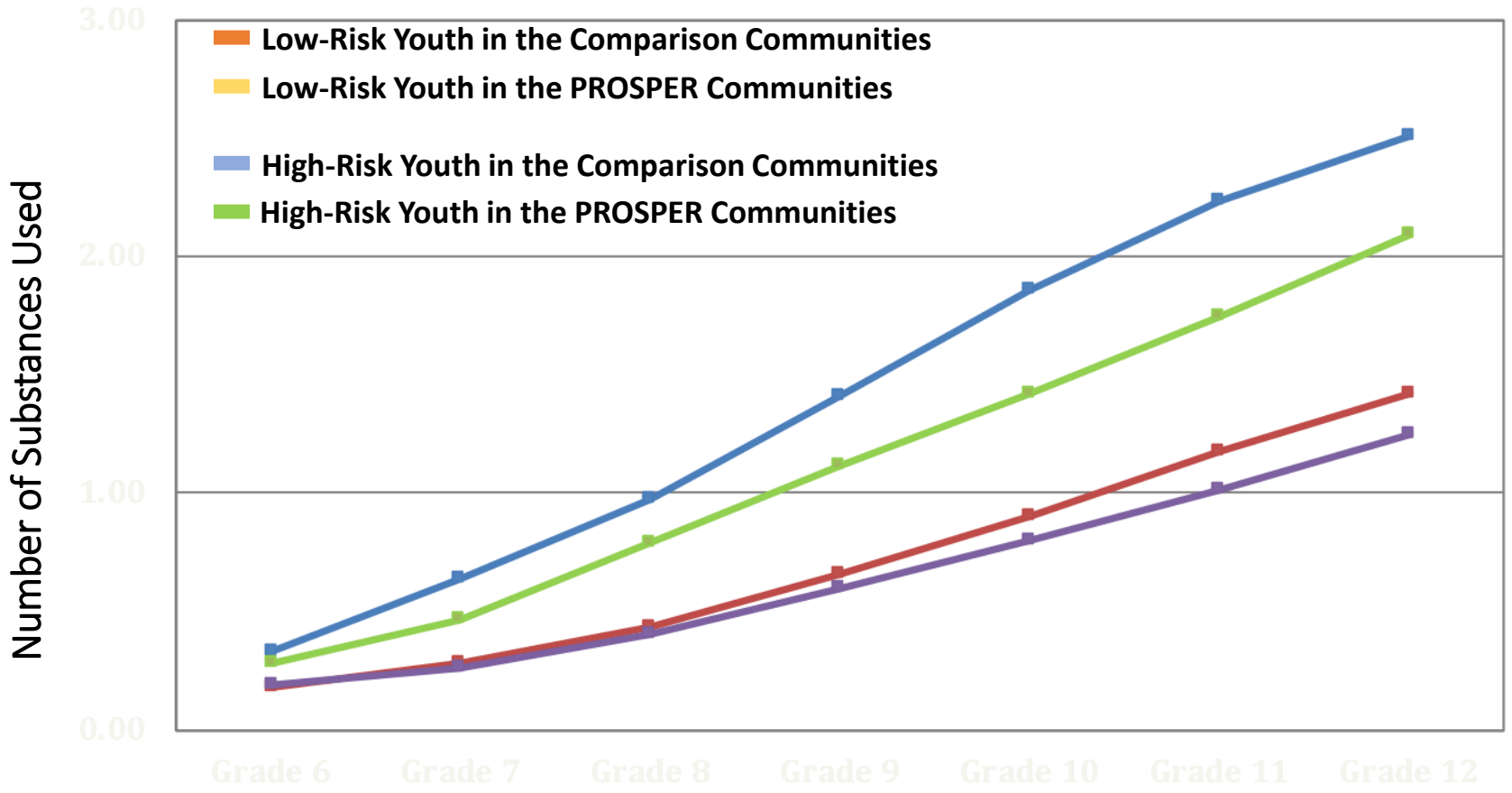
- Participant confidentiality and privacy
  - Where will the data be stored?
  - Who will have access to it?
  - What could happen if confidentiality is accidentally violated?

- Results
  - Are they accurate?
  - How will they be presented?
  - Could they be misinterpreted?



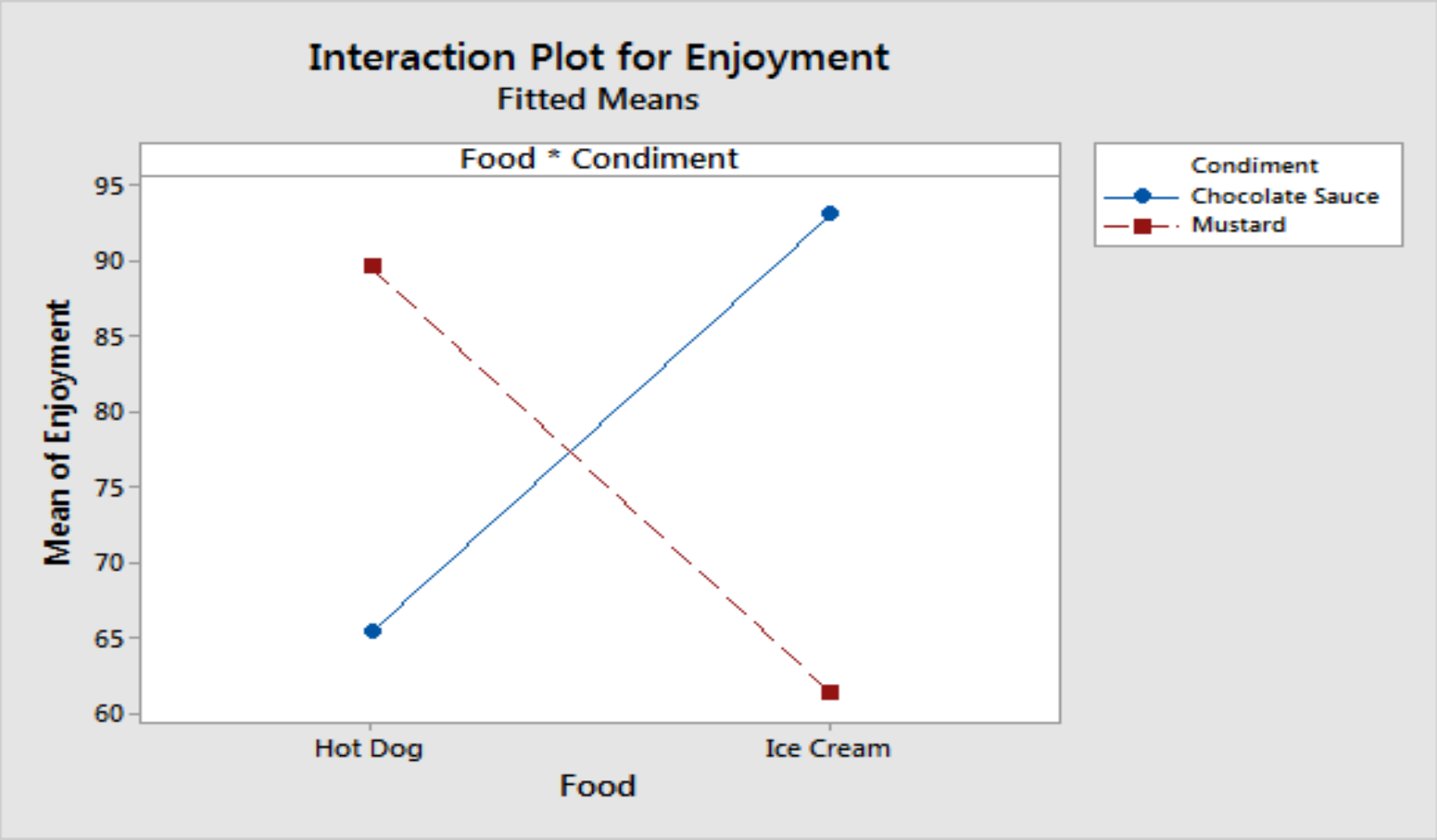
PROSPER vs. control differences are practically significant: For every 100 users in non-PROSPER communities, there are about 20 fewer in PROSPER communities.

Alcohol, Tobacco & Marijuana Use Through End of High School: High-Risk vs. Low-Risk Youth



Source: Spoth, Redmond, Shin, Greenberg, Feinberg, et al. (2012). PROSPER community-university partnerships delivery system outcomes through 6½ years past baseline. Manuscript under internal review.
* Higher Risk = score of 1 or higher on gateway use at baseline (sum of 0, 1 score on lifetime alcohol, cigarette or marijuana use); Lower Risk = score of < 1.

Do you understand your data?



- If your data are not clear, don’t share them with stakeholders
- Take time to really understand your data, even if it means consulting an evaluation expert
- Evaluation findings are often used to make important decisions, so we need to be sure they are right.

Thank You!

jaw900@psu.edu