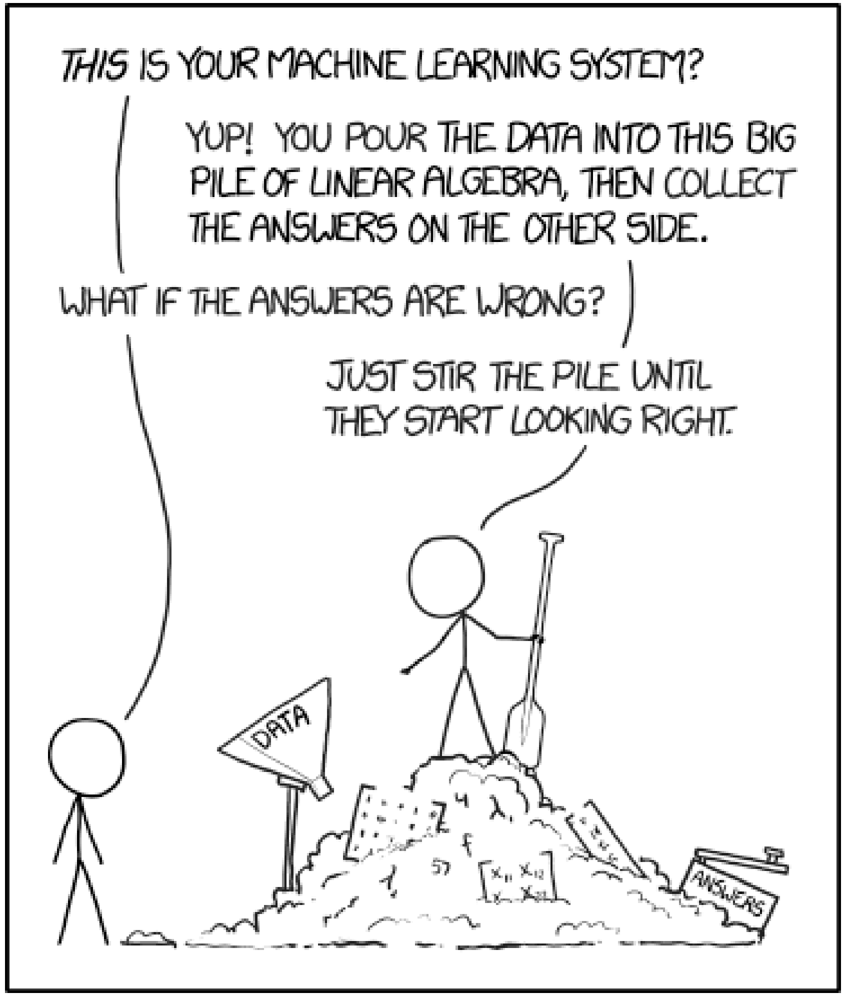

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Source: XKCD
https://xkcd.com/1838/

NC OSBM

OFFICE OF STATE BUDGET
AND MANAGEMENT

Performance Management Academy

Day 3:
Finding and Implementing Evidence

October 26, 2021

PM Academy Roadmap

Title	Date	Key Topics Addressed
Performance Management & Setting Goals	Oct 12	• Academy introduction/overview • Defining performance management; implementation, benefits, etc. • Linking strategic plans to performance management; decision-making, analysis
Deciding What Evidence is Needed	Oct 19	• Overview of evidence 101; impact/process evaluation, etc. •Theory of Change •Generalizability Framework
Finding and Implementing Evidence	Oct 26	•Methods 101; Types of evidence, assessing quality, etc. •Searching for Evidence: Using clearinghouses, Google Scholar •Breakouts: Budget Development, Equity in Implementation, Data & Contracting
Observation and Measurement	Nov 2	• Surveys, psychology of preference, using administrative data, process mapping, etc.
Changing Minds	Nov 9	• Best practices and examples for success • Pre-analysis planning & data visualization • Success stories from NC state government

Today's Agenda

Agenda Items

1. Evaluating Evidence

1. Why do we evaluate?
2. Types of available evidence
3. How do we judge whether the evidence is real evidence/good evidence/lacking?
4. If evidence is limited, how do we build evidence?

2. Small Group Exercise: Critiquing Study Designs

3. Where to Find Strong Evidence: Tips, Tricks & Resources

4. Breakout Groups:

1. Using Evidence in Funding Requests
2. Considering Equity when Using Evidence in Program Implementation
3. Using Evidence in Contracting

Speakers

Paul Devenish
State Budget Management Analyst
Office of State Budget & Management

Evaluating Evidence

Example: The Original D.A.R.E Program (1983-2009)

"The Drug Abuse Resistance Education (D.A.R.E.) program is the most comprehensive drug prevention curricula in the world." – DARE Website

D.A.R.E.'s original curriculum was not shaped by prevention specialists but by police officers and teachers in Los Angeles. Fueled by word of mouth, the program quickly spread to 75 percent of U.S. schools. – *Scientific American*

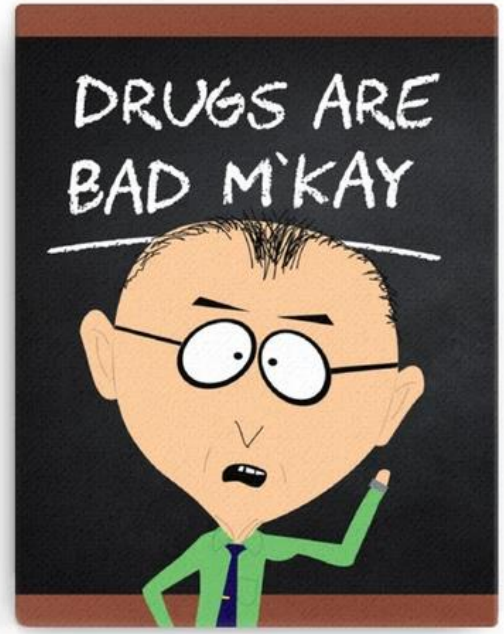
Source: Dare.org

Example: What does the evidence show about D.A.R.E (1983-2009)?

Evidence-Based Justice: DARE proves ineffective

The Sentinel

“The program receives over $200 million in annual funding, despite little or no research evidence that D.A.R.E. has been successful in reducing adolescent drug or alcohol use. – *Rosenbaum, The Center for Evidence Based Crime Policy*



“DARE’s limited influence on adolescent drug use behavior contrasts with the program’s popularity and prevalence” – *Research Triangle Institute, 1994*

Sources: Dare.org, https://www.mycast.io/talent/mr-mackey

Using Evidence to Improve D.A.R.E (2008-Present)

What did D.A.R.E do to incorporate evidence?

- Identified & evaluated 9 potential programs
- Outsourced curriculum development to research scientists for elementary
- Modified training style to evidence based interactive learning

Source: Dare.org

Did it work?

Students who completed keepin’ it REAL indicated that they were less likely to try drugs and alcohol compared to the control group and utilized a variety of tools to stay sober. Students' antidrug attitudes were also more likely to stick over time following the completion of the curriculum. – Hecht et al.

Source: Results First Clearinghouse Database

Recap – Theory of Change Causal Hypothesis

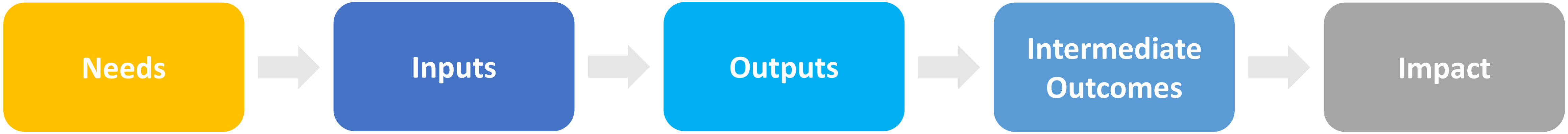
- Answers the question: How do I expect results to be achieved?
- **If** [inputs] and [activities] **produce** [outputs], this should lead to [outcomes] which will ultimately **contribute to** [goal].
- Maps the expected causal pathway between the Inputs and desired outcomes, assumptions, and implementation risks.

This content was developed in consultation with J-PAL North America.

Recap: theory of change

A theory of change is a comprehensive description of how and why desired change is expected to happen in a particular context

Simply put, it is a road-map for your program



Needs → **Inputs** → **Outputs** → **Intermediate Outcomes** → **Impact**

The needs of the population an intervention is designed to serve

What activities and/or inputs are done as part of the intervention

Tangible products or services produced as result of intervention

Changes in behavior, belief, or knowledge, that result from the delivered intervention

The desired change or outcome(s)

Underlying Assumptions

Understanding Evidence: The Full Story

- Why do we evaluate?
- Types of available evidence
- How do we judge whether the evidence is real evidence/good evidence/lacking?
- If evidence is limited, how do we build evidence?

This content was developed in consultation with J-PAL North America.

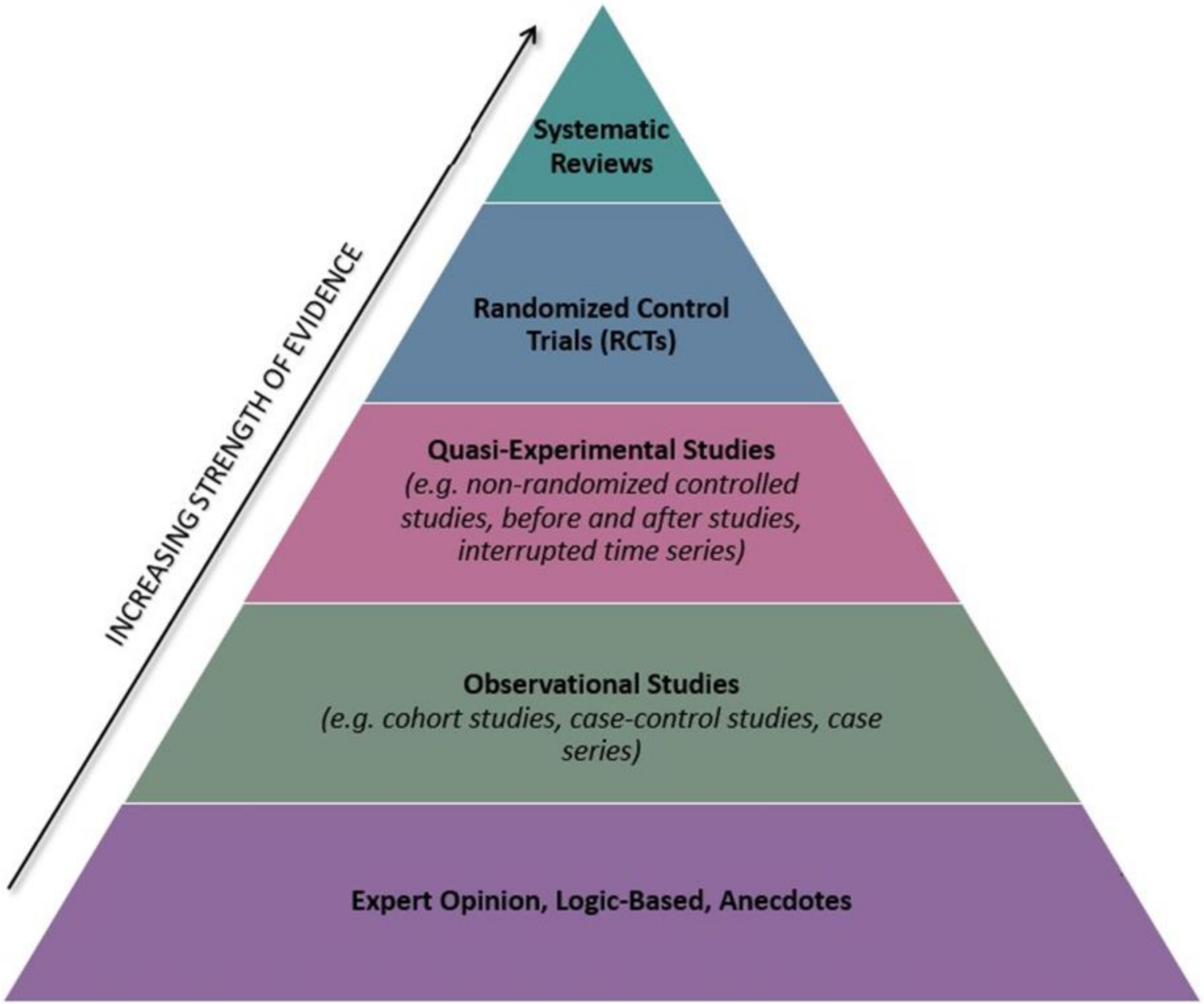
Why We Evaluate

- **What approach does the evidence point to?**
- **Did / does the program work as planned?**
- Measuring outcomes
- Program Integrity

This content was developed in consultation with J-PAL North America.

What Types of Evidence are Available?

Recap: Evidence Pyramid



INCREASING STRENGTH OF EVIDENCE

Systematic Reviews

Randomized Control Trials (RCTs)

Quasi-Experimental Studies
(e.g. non-randomized controlled studies, before and after studies, interrupted time series)

Observational Studies
(e.g. cohort studies, case-control studies, case series)

Expert Opinion, Logic-Based, Anecdotes

Research Evidence

Terms

Treatment: The group that receives the intervention

Control/Comparison: A group that is as similar as possible to the treatment group but is not intended to receive the intervention. Also called the control group or counterfactual.

Quasi-Experimental Studies

- Treatment and comparison are not randomly assigned
- Utilization of statistical methods decrease differences and account for bias

Randomized Control Trials

- Treatment and comparison are randomly assigned
- Both groups have identical characteristics, reducing bias
- Difficult to complete in the real world

Systematic Reviews

- Research on available research
- Need high numbers of rigorous study in the area of interest
- Qualitative summary

This content was developed in consultation with the Pew Charitable Trusts' Results First initiative.

How Do We Assess Evidence?

Bottom Line Questions:

- Does it work? Is the program proven to improve outcomes?
- How confident can we be in what the evidence says?

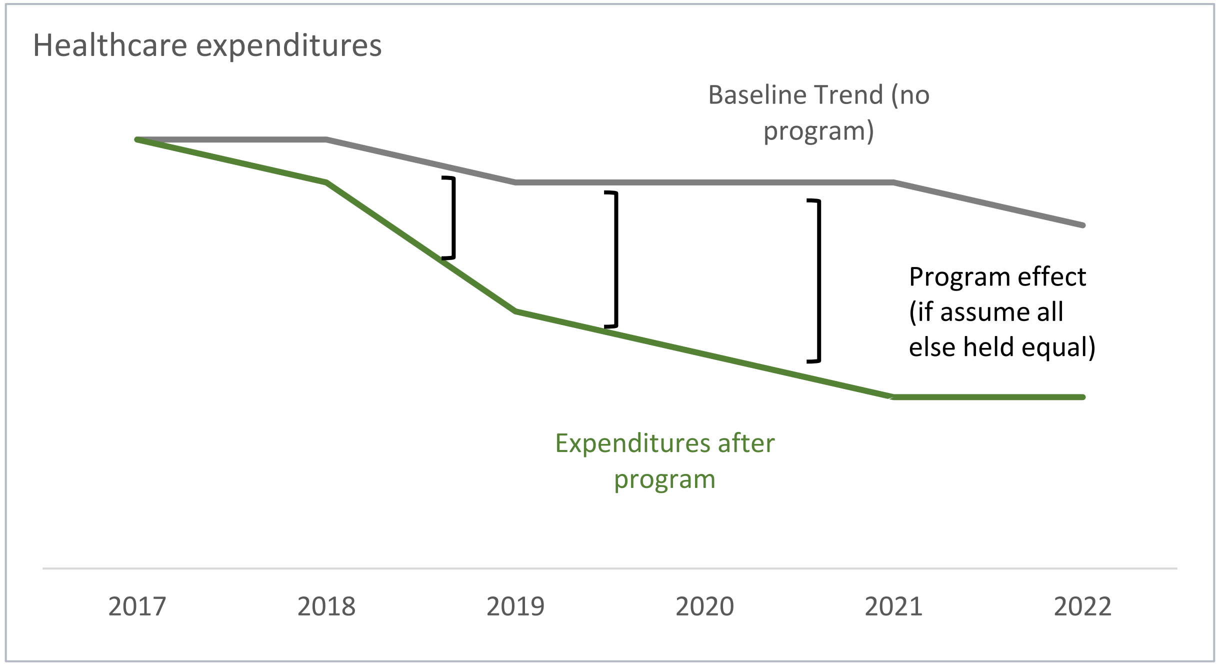
Image source: https://marksmanhealthcare.com/category/social-media/

Identifying Evidence of Program Effectiveness (I)

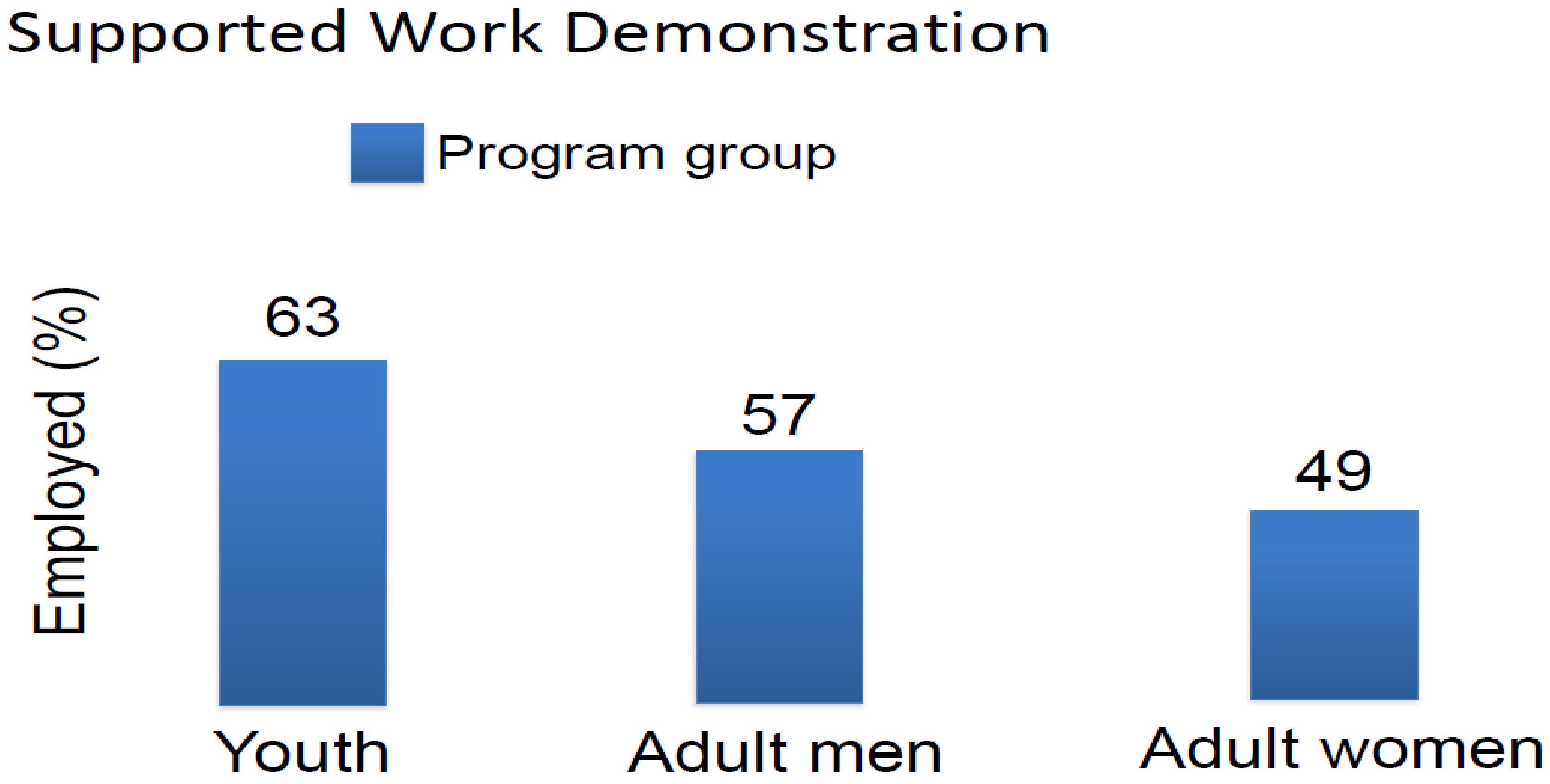
Bottom Line Questions:

Does it work? Is the program proven to improve outcomes?

- Rigorously designed research studies attempt to identify the change in outcomes that is attributable to the program



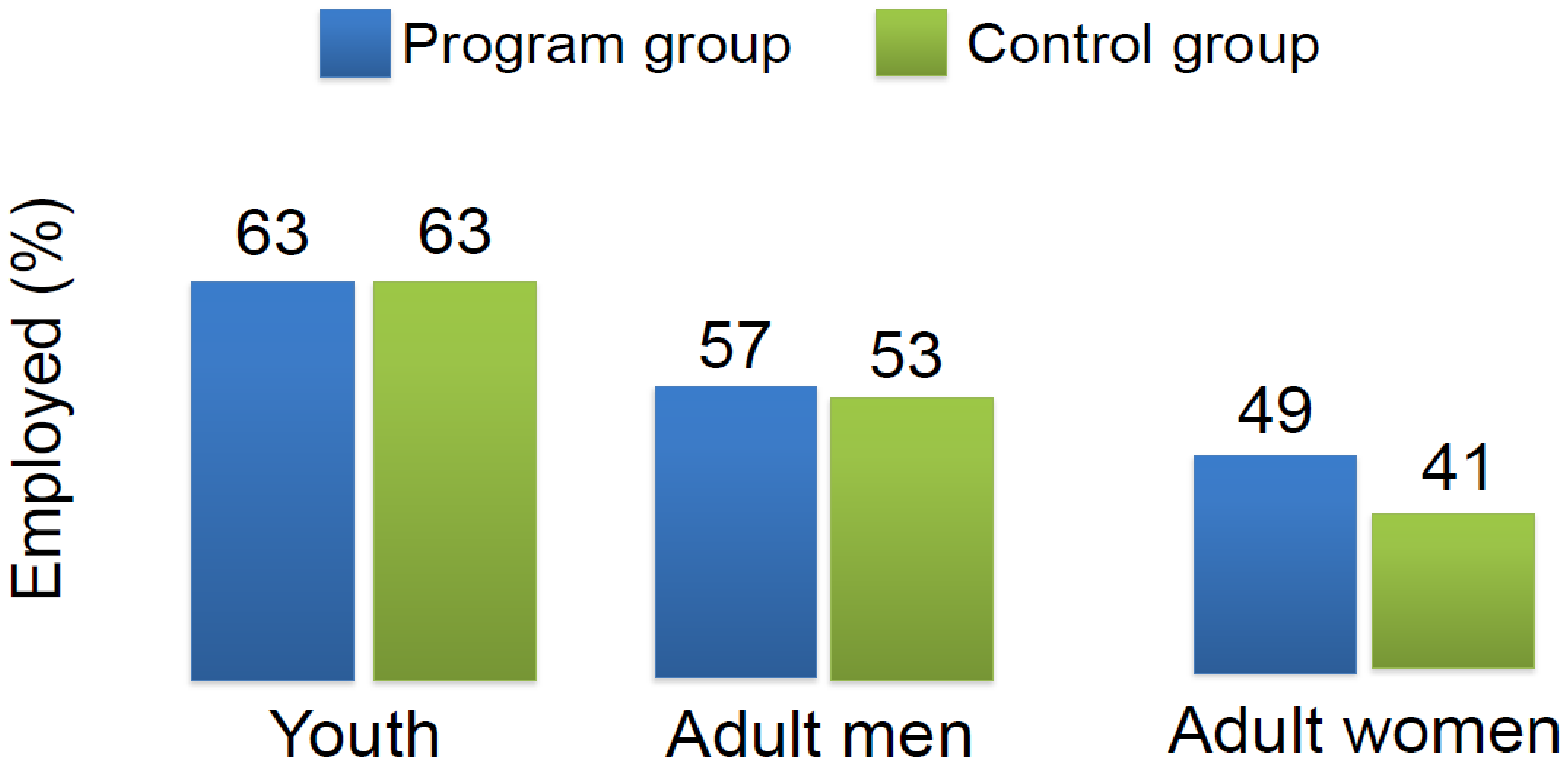
Identifying Evidence of Program Effectiveness (II)



Content developed in collaboration with David Yokum at The Policy Lab, thepolicylab.brown.edu

Identifying Evidence of Program Effectiveness (III)

Supported Work Demonstration



Content developed in collaboration with David Yokum at The Policy Lab, thepolicylab.brown.edu

Tiered Levels of Evidence

- OSBM defined tiered levels of evidence that consider the program's demonstrated effect on targeted outcomes *and* the strength of the evidence to support those conclusions.

Uses

- Report the research findings for existing or proposed programs in a policy area.
- Determine what proportion of expenditures devoted to proven programs.
- Evaluate and inform certain budget proposals.
- Identify priorities for future evaluation.

Tiered Levels of Evidence

Proven Effective

Promising

Theory-based

Mixed Effects

No Effect

Proven Harmful

Tools for Assessing Evidence

Key Tool I: Literature Review

What is it?

Background research that expands or clarifies a problem or issue and summarizes what others have done or learned about the problem or issue. Other methods include systematic reviews and evidence/gap maps.

Appropriate Uses	Issues to Consider
Deepening understanding of the issue or problem.	Need to distinguish solid evidence of program effectiveness from self-serving or poorly-supported claims of effectiveness.
Developing options for addressing the problem by learning from what others have done.	Need to identify research that applies to the relevant program context (i.e. findings about effective distance learning programs won't work well in areas where broadband is lacking)
Gathering evidence about programs that are effective (and those that are not).	Research on the topic may not exist.
Can also help identify sources of advice and assistance.	Could overlook "gray literature."

Key Tool II: Comparative Analysis

What is it?

Comparison of program or activity goals, design, and outcomes, often with states thought to be “the best” or to states with similar characteristics (such as neighboring states). The review may include looking at practices that are widely accepted as promising or best-in-class due to their superior results.

Appropriate Uses	Issues to Consider
Understanding alternate methods of service delivery and program design.	Important to evaluate the evidence base for “best” or “promising practices,” which may gain prominence and become discredited later.
Understanding the context and where you sit relative to those in your comparative analysis.	Good practices often depend on the situation and context.
Developing performance targets.	Requires careful review of research and policy literature and appropriate selection of comparison jurisdictions.
Generating policy, program, and budget options.	Important to think about the criteria you use to select your comparison group.

Assessing Statistical Evidence

Be careful when using research results, esp. statistical evidence!

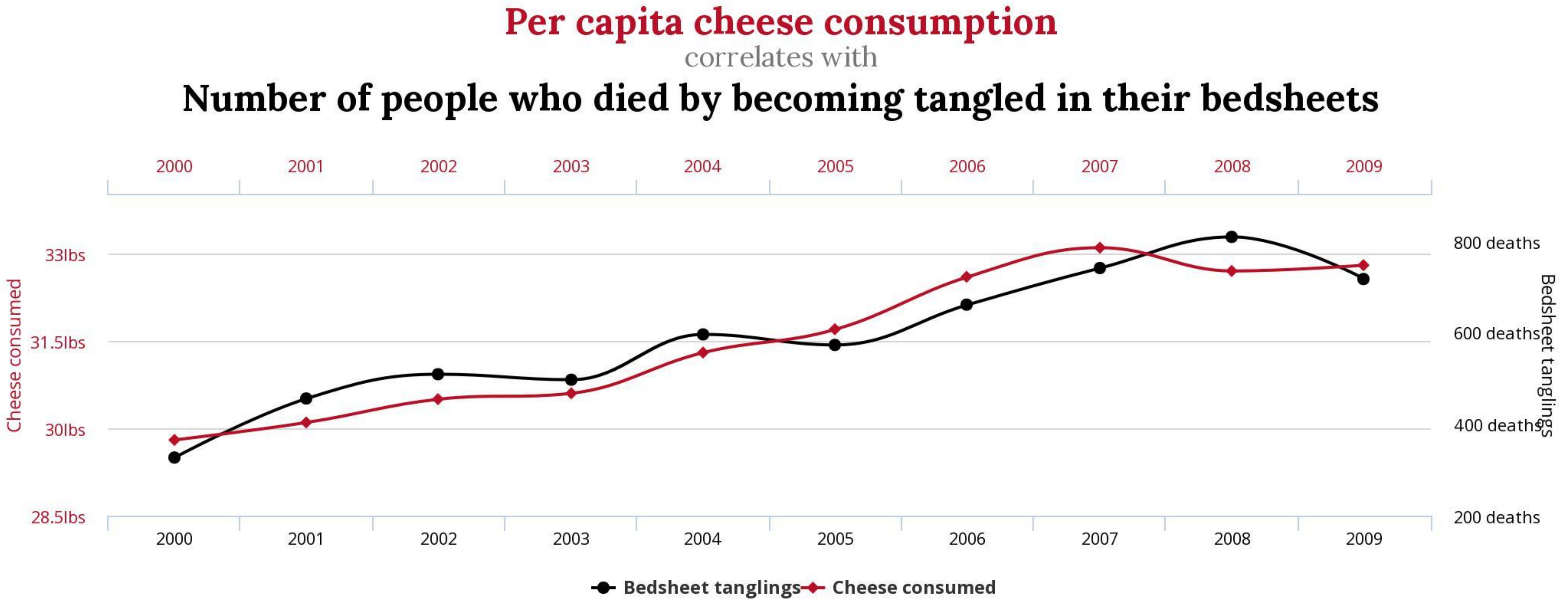
Evidence ≠ Effective

Publication ≠ Quality

Correlation ≠ Causation

This content was developed in consultation with the Pew Charitable Trusts' Results First initiative.

Correlation is not the same as causation!

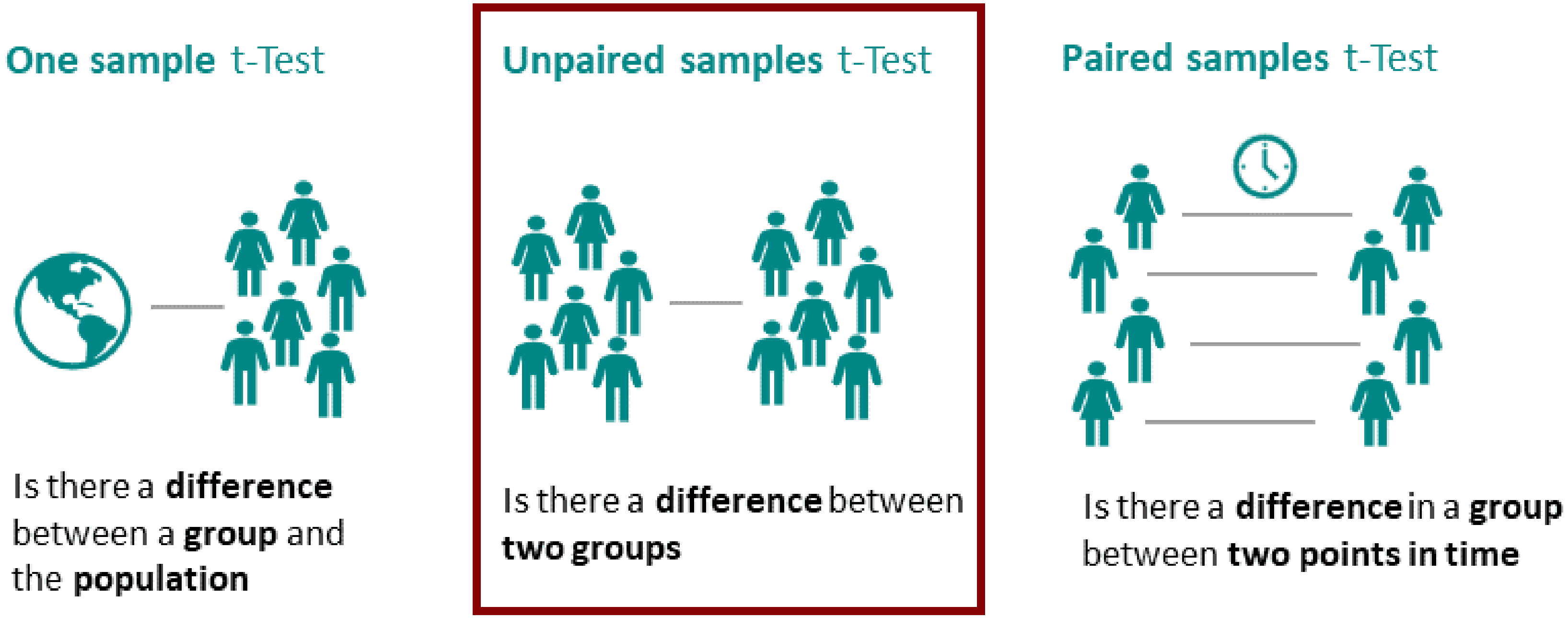


Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

Source: Spurious Correlations, Tylervigen.com

Statistical Tool I: T-tests

- **What is it?** T-tests allow you to compare the means of 2 sets of data to see if they are statistically different.



One sample t-Test

Is there a **difference** between a **group** and the **population**

Unpaired samples t-Test

Is there a **difference** between **two groups**

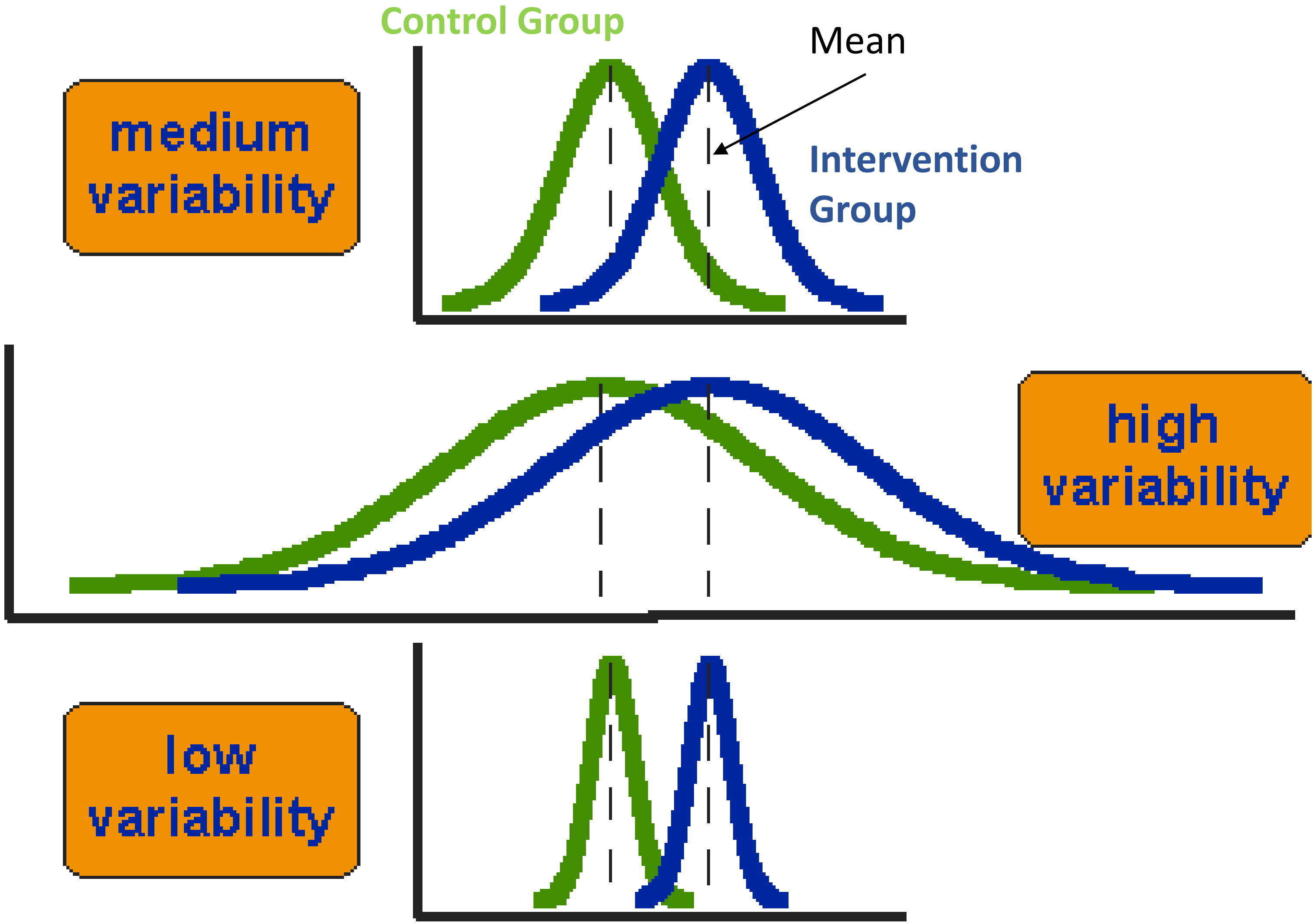
Paired samples t-Test

Is there a **difference** in a **group** between **two points in time**

- **How is it used?** Examples: whether the exam scores of a group of students are statistically different from those of their peers. Or impact of a particular treatment vs a placebo; or can compare some aspect of patients' health before and after a treatment.
- **Measurement**: Probability that groups are different: generally use 5% likelihood that they are the same ('p=0.05') as boundary at which they are determined to be statistically different.

Diagram from: https://datatab.net/tutorial/unpaired-t-test.

T-test: statistical differences illustrated

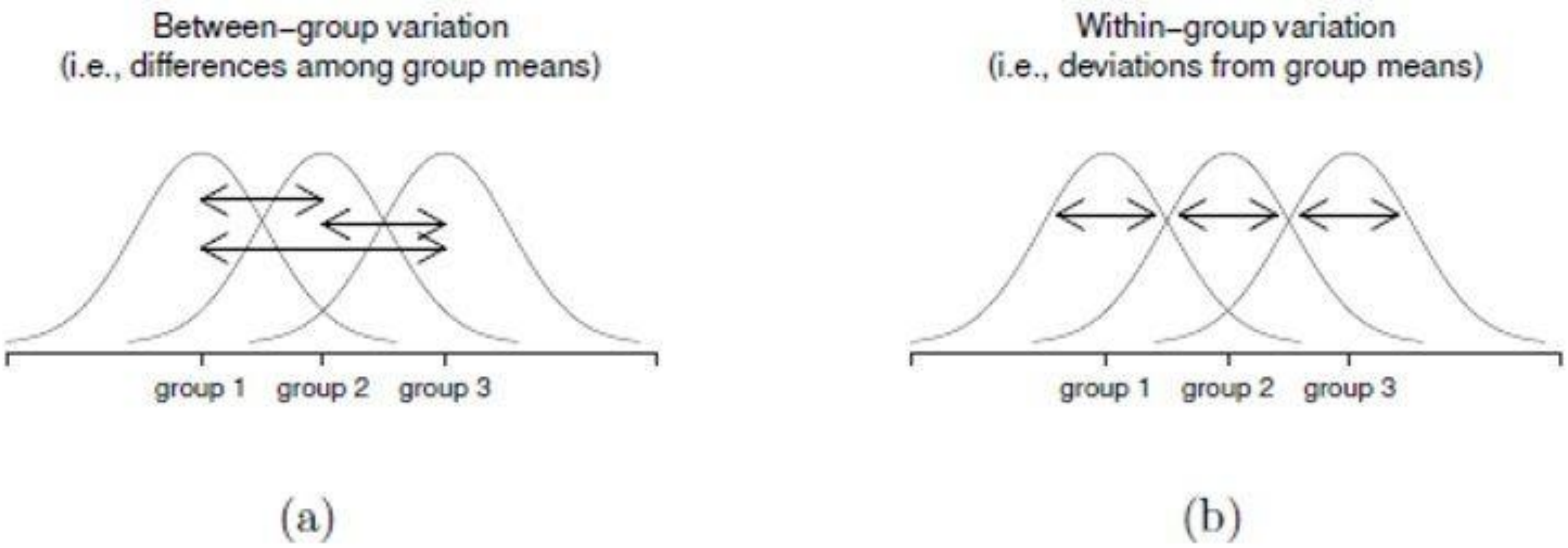


- Recall: T-tests look at whether the **means of a control / baseline group and an intervention group are statistically different** (>95% confidence).
- Where there is high variability in the datasets, they are less likely to be statistically different.
- Conversely, low variability makes it more likely that they are statistically different.
- Key issue to watch for: 1-tailed or 2-tailed test?

Source: https://conjointly.com/kb/statistical-student-t-test/

Statistical Tool II: ANOVA

- **What is it?** ANOVA allows comparison of the variances between 3 or more sets of data.



Source: https://estamatica.net/anova-table-with-spss/.

- **How is it used?** Example: where want to compare the impact not only of taking one drug vs another to treat an illness, but also to compare different doses.
- **Measurement**: compare variance between groups with variance within groups. The greater the multiple, the more likely they're different ('F-statistic')
- **Also** Chi-squared test and various others, e.g. where assumptions that variables are normally distributed don't hold.

Statistical tool III: regression analysis

- A method for establishing a statistical relationship between a set of variables (‘independent’ or ‘explanatory’ variables) and another variable which depends on them (the ‘dependent’ or ‘outcome’ variable).
- Which matters most? Which can we ignore? How do these variables interact? And what’s the strength of the relationship between the dependent variable and the independent variables?

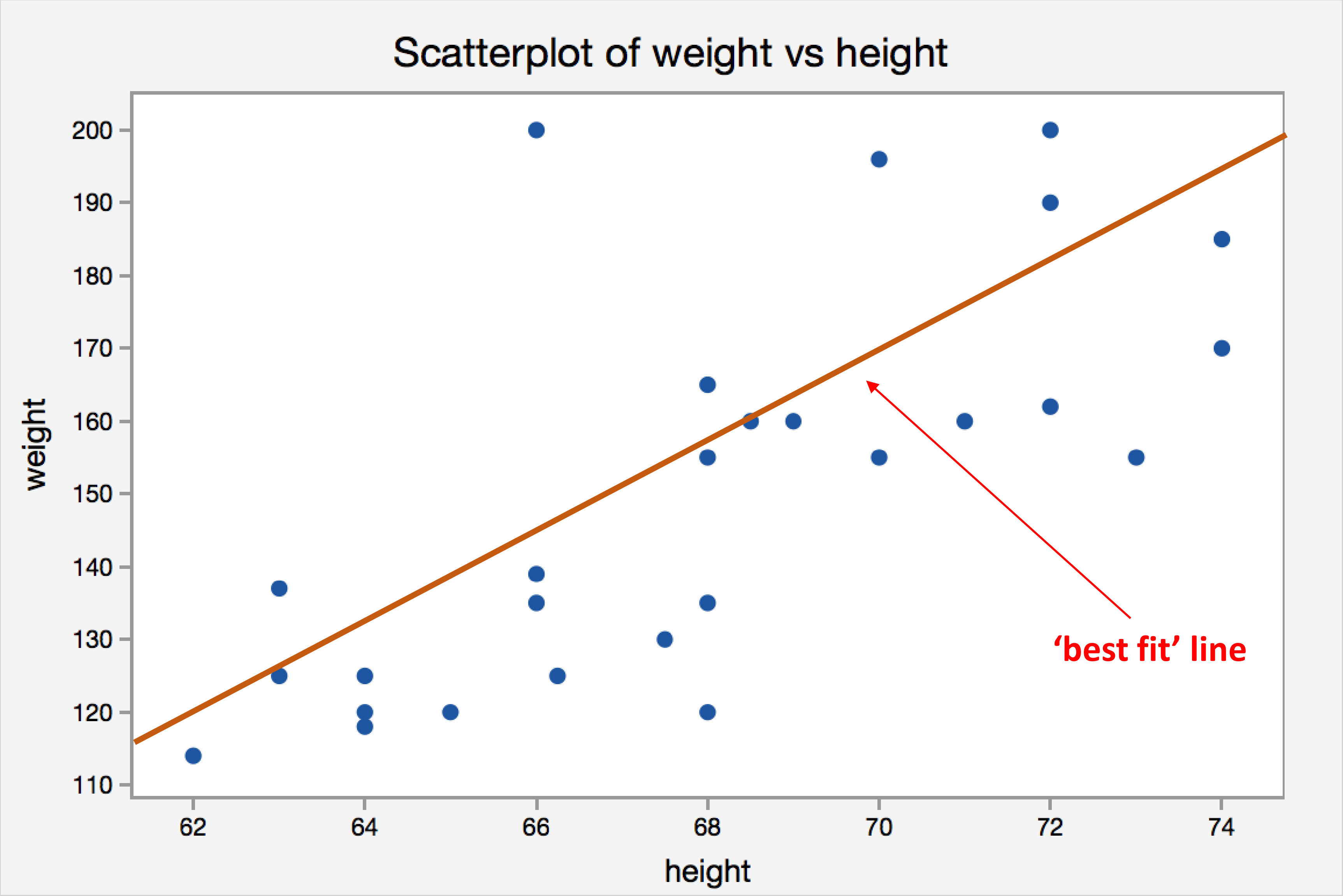
$$\underset{\text{Dependent variable}}{\hat{y}_i} = \underset{\uparrow\ \text{Constant}}{a} + \underbrace{b_1x_1 + b_2x_2 + b_3x_3}_{\text{Independent variables}} \ldots$$

<u>For example:</u>

- Weight = 115 + 8.6(Height_Inches)
- All else equal, each additional inch of height predicts an additional 8.6 lb of weight

Content developed in collaboration with David Yokum at the Policy Lab, thepolicylab.brown.edu

Regression aim: find ‘best fit’ line across variables



Content developed in collaboration with David Yokum at the Policy Lab, thepolicylab.brown.edu

Caution: Study quality can be poor even from “reputable” sources

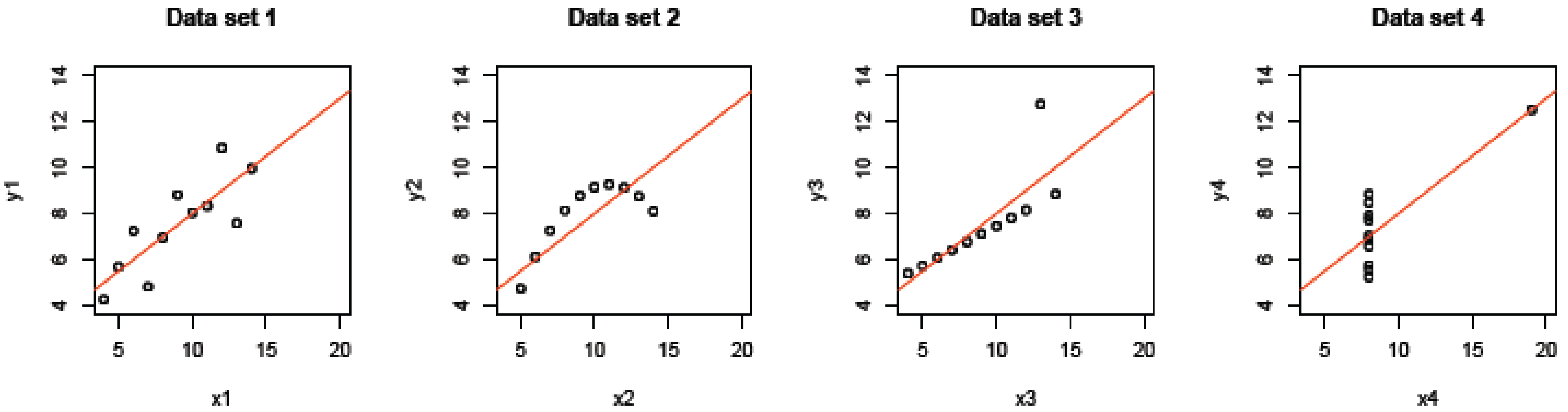
- Too few people were studied
- The people studied are not representative of the population you care about
- Causal claims made, but no counterfactual
- Effect sizes are not talked about clearly
- Studies with ‘wrong’ outcome not published
- Fishy handling of the data
  - Find measures to suit a preconceived hypothesis
  - ‘P-hacking’[1] – playing with data to try to find a statistically-significant result
  - Sometimes results from not really understanding limitations / potential complexities of statistical techniques

[1] For a discussion of p-hacking, see e.g. https://www.psychologicalscience.org/observer/research-preregistration-101#.WMIWj_krLic.

Content developed in collaboration with David Yokum at the Policy Lab, thepolicylab.brown.edu

Common Problems with Regressions

1. **Look** at the data.

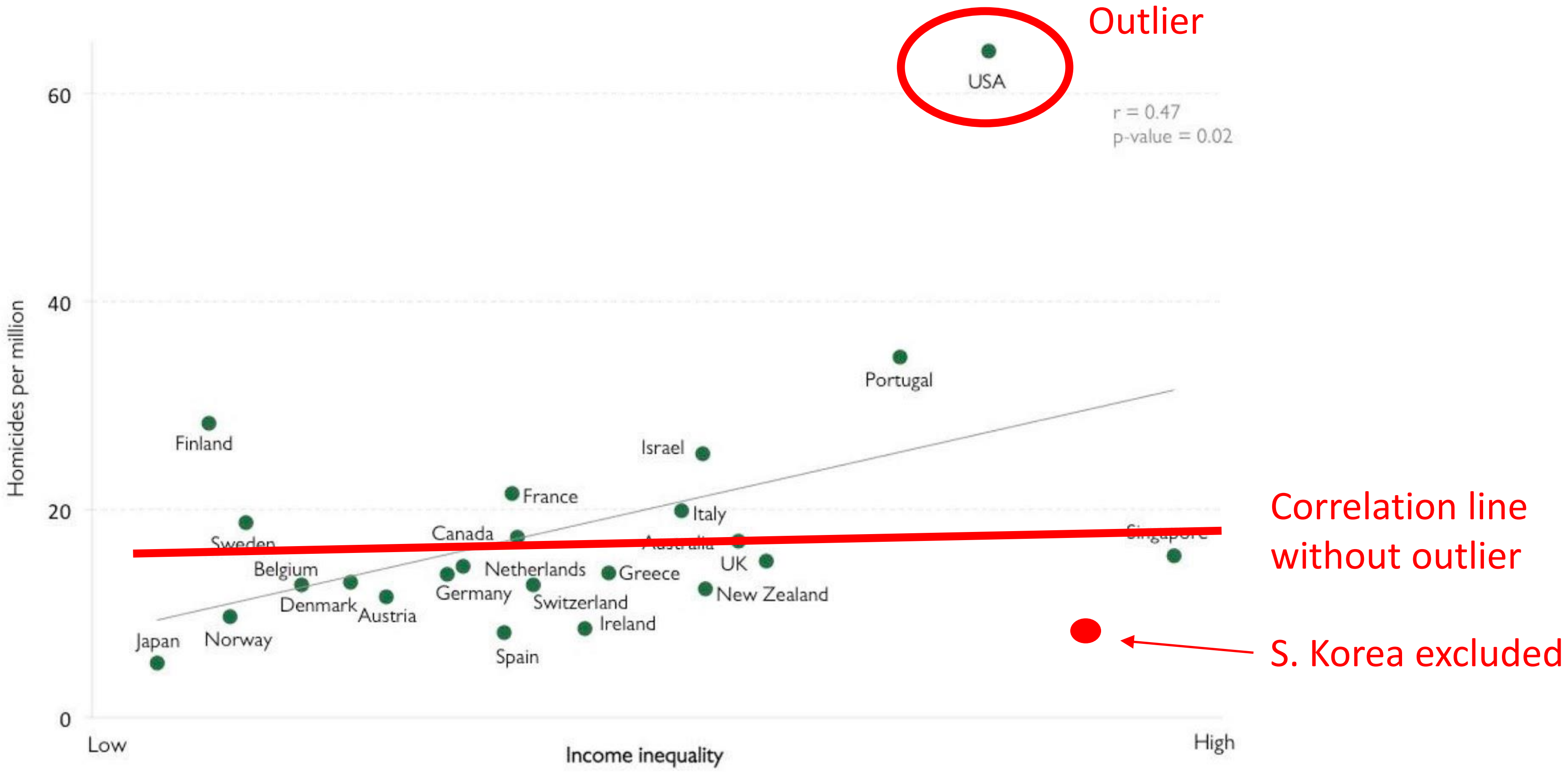


2. Ask yourself whether **all** the relevant predictors are included in the model.
3. Have any data points been excluded; and if so, why (what were the criteria for excluding them and are they reasonable?)? What happens if they are included?
4. What happens over time – are results consistent across years, or if different measures of the variables are used?
5. Is there open and clear discussion of the statistical tests applied to test robustness of regression with **this dataset**?

Content developed in collaboration with David Yokum at the Policy Lab, thepolicylab.brown.edu

Outliers / Skewed Data

Homicide rates are higher in more unequal rich countries

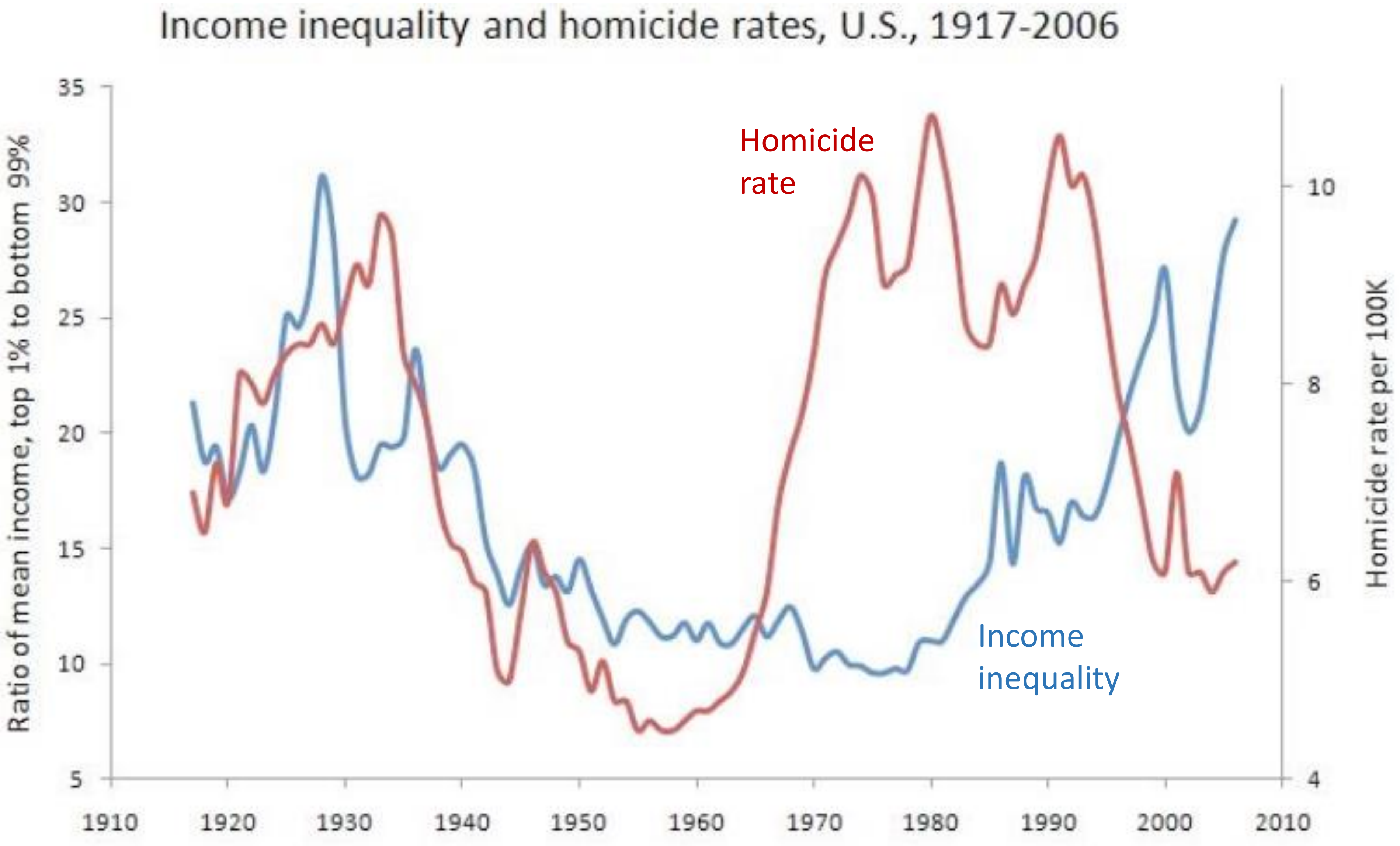


Source: Wilkinson & Pickett, The Spirit Level (2009)

THE EQUALITY TRUST

Source: Christopher
Snowdon

Detailed analysis shows no correlation within the outlier country (USA)



Source: Christopher Snowdon

Resources

- Tools for evaluation:
  - Clearinghouses (more to come this afternoon!)
  - Technical resources – R (open-source, free!), SAS, Stata
- Universities
- Office for Strategic Partnerships (OSP) – more on Day 4!
- Learning Agendas
- Evaluation.gov

References

- *D.A.R.E. America*, dare.org/.
- Ennett, S., Tobler, NS., Ringwalt, C., & Flewelling, R. (1994). How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health*, *84*(9), 1394-1401.
- Hecht ML, Marsiglia FF, Elek E, Wagstaff DA, Kulis S, Dustman P, Miller-Day M. Culturally grounded substance use prevention: an evaluation of the keepin' it R.E.A.L. curriculum. Prev Sci. 2003 Dec;4(4):233-48. doi: 10.1023/a:1026016131401. PMID: 14598996.
- “Keepin'it REAL .” *The Pew Charitable Trusts Results First Clearinghouse Database*, The Pew Charitable Trusts, www.pewtrusts.org/en/research-and-analysis/data-visualizations/2015/results-first-clearinghouse-database.
- Kopper, Sarah, et al. “Introduction to Measurement and Indicators.” *The Abdul Latif Jameel Poverty Action Lab (J-PAL)*, The Abdul Latif Jameel Poverty Action Lab (J-PAL), www.povertyactionlab.org/resource/introduction-measurement-and-indicators.
- Nordrum, Amy. “The New D.A.R.E. Program-This One Works.” *Scientific American*, Scientific American, 10 Sept. 2014, www.scientificamerican.com/article/the-new-d-a-r-e-program-this-one-works/.
- “Program Details: Drug Abuse Resistance Education (DARE) (1983-2009).” *CrimeSolutions, National Institute of Justice*, National Institute of Justice, crimesolutions.ojp.gov/programdetails?id=99.
- Snowden, Christopher, “The Spirit Level 10 Years On”, available at http://spiritleveldelusion.blogspot.com/2019/03/the-spirit-level-ten-years-on.html.
- Spurious Correlations, www.tylervigen.com
- UCLA Institute for Digital Research and Education, "What are the differences between one-tailed and two-tailed tests?", available at: https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/
- Vaughn, Joshua. “Evidence-Based Justice: DARE Proves Ineffective.” *The Sentinel*, 23 Mar. 2018, cumberlink.com/news/local/closer_look/evidence-based-justice-dare-proves-ineffective/article_b85d88bf-ccfd-5c9a-b958-2da91dbc8abe.html.
- “Where to Search for Evidence of Effective Programs.” *The Pew Charitable Trusts*, The Pew Charitable Trusts, Apr. 2020, www.pewtrusts.org/-/media/assets/2020/04/evidenceresources_resultsfirst_fs.pdf.
- Wilkinson, R and Pickett, K (2009) “The Spirit Level: Why Greater Equality Makes Societies Stronger”, Bloomsbury Press

QUESTIONS?

Take a Break

Welcome Back!

Breakout Group Activity:

- Look at a series of write-ups of study results. Assess whether it is a good example of evidence, discuss.

Breakout Group Activity

Instructions & Disclaimer!

- The following slides each have a write up of study results. Your job is to assess whether it's a good creation and use of evidence or, if not, why not (e.g. what would you do differently; what else would you want to know?).
- You'll think about it yourself first, and then discuss with your group.
- The examples are made up, to simplify for teaching purposes.

Breakout Group Activity

Telework and Job Effectiveness

- When the COVID-19 pandemic disrupted the ability to work in the office, the research firm Schmidtt Analytics partnered with PhoneTech, Inc. (a large call center company) to study whether teleworking employees were equally or more effective at home as in the office. Over the course of four months, they randomly assigned twelve telemarking staff to either work from home or return and work from the office. They found that the six people working from home were able to process, on average, 45% more calls than those working at their cubicle! PhoneTech decided to convert its 4,300 telemarking employees to fully remote work, without delay.

Breakout Group Activity

Telework & Job Effectiveness **SAMPLE SIZE**

- When the COVID-19 pandemic disrupted the ability to work in the office, the research firm Schmidtt Analytics partnered with PhoneTech, Inc. (a large call center company) to study whether teleworking employees were equally or more effective at home as in the office. Over the course of four months, they randomly assigned twelve telemarking staff to either work from home or return and work from the office. They found that the six people working from home were able to process, on average, 45% more calls than those working at their cubicle! PhoneTech decided to convert its 4,300 telemarking employees to fully remote work, without delay.

Breakout Group Activity

Minimum Wage

- The District of Columbia initiated a new employment program in March 2021. Six hundred people without employment have participated. The results have been impressive: of the 600 people, 234 now have jobs. The Director of the program is asking for funding in FY23 to double the size of the program.

Breakout Group Activity

Minimum Wage. **CONTROL GROUP**

- The District of Columbia initiated a new employment program in March 2021. Six hundred people without employment have participated. The results have been impressive: of the 600 people, 234 now have jobs. The Director of the program is asking for funding in FY23 to double the size of the program.

To know whether 234 is good or bad, need a control group – how many would be employed by now even without the program?

Breakout Group Activity

Absenteeism

- In 2020, a large randomized controlled trial (RCT) was conducted to evaluate whether the program Every Kid Counts increases school attendance. 20,000 1st graders participated. The 10,000 1st graders who participated in Every Kids Counts only had a 1.4% absenteeism rate, compared to 9% for the control group. Based on these results, Every Kids Counts is going to be scaled across all high schools in the state.

Breakout Group Activity

Absenteeism. **GENERALIZABILITY**

- In 2020, a large randomized controlled trial (RCT) was conducted to evaluate whether the program Every Kid Counts increases school attendance. 20,000 1st graders participated. The 10,000 1st graders who participated in Every Kids Counts only had a 1.4% absenteeism rate, compared to 9% for the control group. Based on these results, Every Kids Counts is going to be scaled across all high schools in the state.

- **It works for 1st graders, but teenagers are very different!**

Breakout Group Activity

Work & Satisfaction

- The plot on the right shows the relationship between number of hours worked during a day and job satisfaction. Based on the regression results (orange line), there is no relationship.

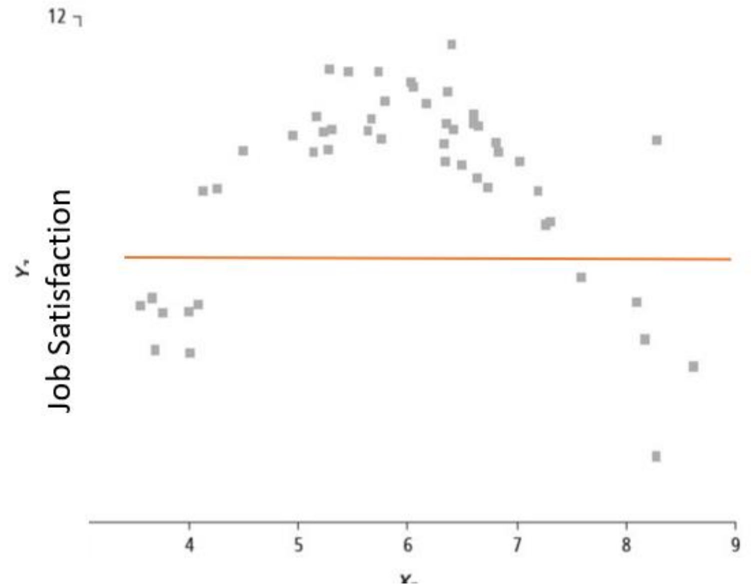


Figure 7.7 ♦ Scatter Plot for Strong Curvilinear Relationship (for These Data, $r = .02$)

Breakout Group Activity

Work & Satisfaction **REGRESSION WOES**

- The plot on the right shows the relationship between number of hours worked during a day and job satisfaction. Based on the regression results (orange line), there is no relationship.
- **The regression line is flat, but *look* at the data - it masks the fact that satisfaction is low, goes up, then drops again.**

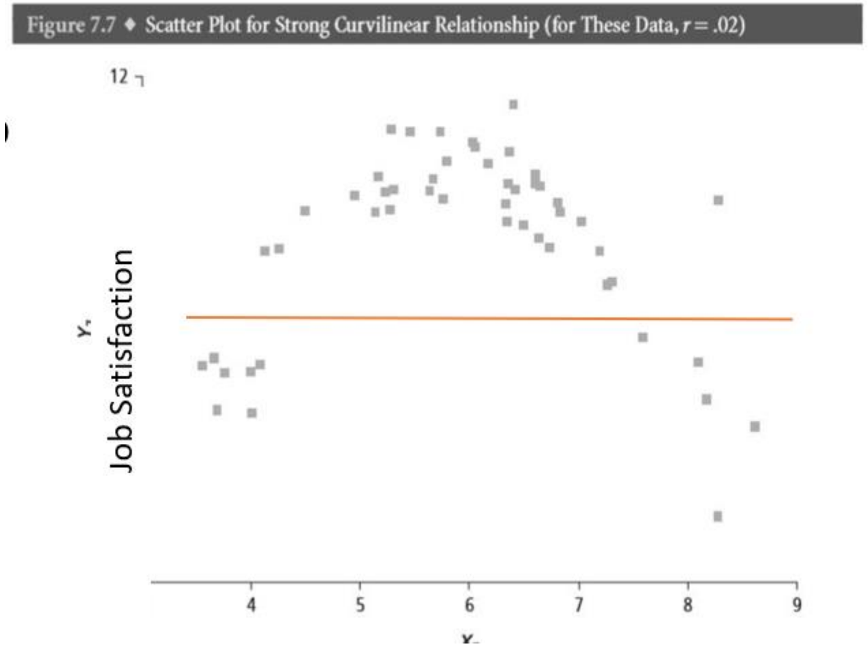


Figure 7.7 ♦ Scatter Plot for Strong Curvilinear Relationship (for These Data, $r = .02$)

Breakout Group Activity

Coffee and Grades

- A 2019 survey of over 23,000 college students found that the more coffee students drank, the higher their grades. The researchers were worried that other factors might also affect grades, so in their regression model they also controlled for how highly ranked was the college by U.S. News & Reports as well as whether a student was taking courses on a letter grade or only pass/fail.

Breakout Group Activity

Coffee and Grades. **CORRELATION ≠ CAUSE**

- A 2019 survey of over 23,000 college students found that the more coffee students drank, the higher their grades. The researchers were worried that other factors might also affect grades, so in their regression model they also controlled for how highly ranked was the college by U.S. News & Reports as well as whether a student was taking courses on a letter grade or only pass/fail.
- **Many, many things influence grades. To isolate the causal effect of coffee specifically, we'd need to control for *all* of those many, many things. Very hard! Really need a randomized controlled trial (RCT) here.**

Breakout Group Activity

Substance Abuse Treatment

- Jasper Naden was a 26-year old male with a history of substance abuse. It was wrecking his life – he had dropped out of college, his relationships were strained, and he was unemployed. In 2016, he entered the Bright Ventures program, which provided counseling, medication, and a paid internship. Jasper has been clean now for three years, and he works full time as a counselor at Bright Ventures. He recently testified before the legislator: "Bright Ventures changed my life. I know for certain that if I hadn't entered the program, I would be homeless and alone right now – or worse. If we could only expand this program across the state, we could seriously reduce the problem of substance abuse."

Breakout Group Activity

Substance abuse treatment **ANECDOTAL**

- Jasper Naden was a 26-year old male with a history of substance abuse. It was wrecking his life – he had dropped out of college, his relationships were strained, and he was unemployed. In 2016, he entered the Bright Ventures program, which provided counseling, medication, and a paid internship. Jasper has been clean now for three years, and he works full time as a counselor at Bright Ventures. He recently testified before the legislator: "Bright Ventures changed my life. I know for certain that if I hadn't entered the program, I would be homeless and alone right now – or worse. If we could only expand this program across the state, we could seriously reduce the problem of substance abuse."

Speakers

Melissa Roark
State Budget Management Analyst
Office of State Budget & Management

David Yokum
Senior Advisor
NC Office of Strategic Partnerships

Searching for Evidence

Approaches to Finding Evidence

Issue Investigation

Strategic Planning

Program Evaluations or Implementation Design Experiments

Data and Analytics

Cost-Benefit Analysis

Process Improvement

Performance Monitoring

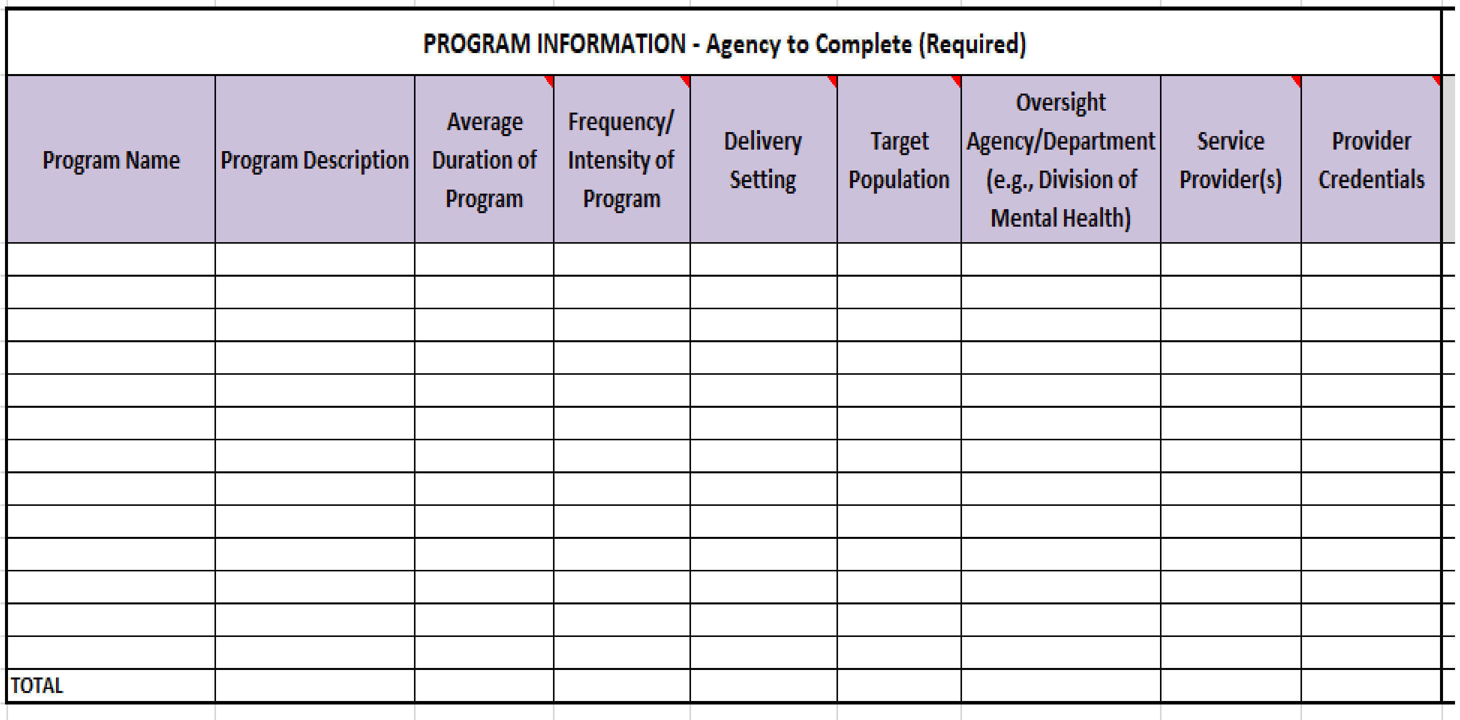
Tool Spotlight: Clearinghouses

A clearinghouse is a “one-stop” resource to find information on the effectiveness of interventions/services. Clearinghouses conduct literature reviews and rate the interventions in a range of policy issues.

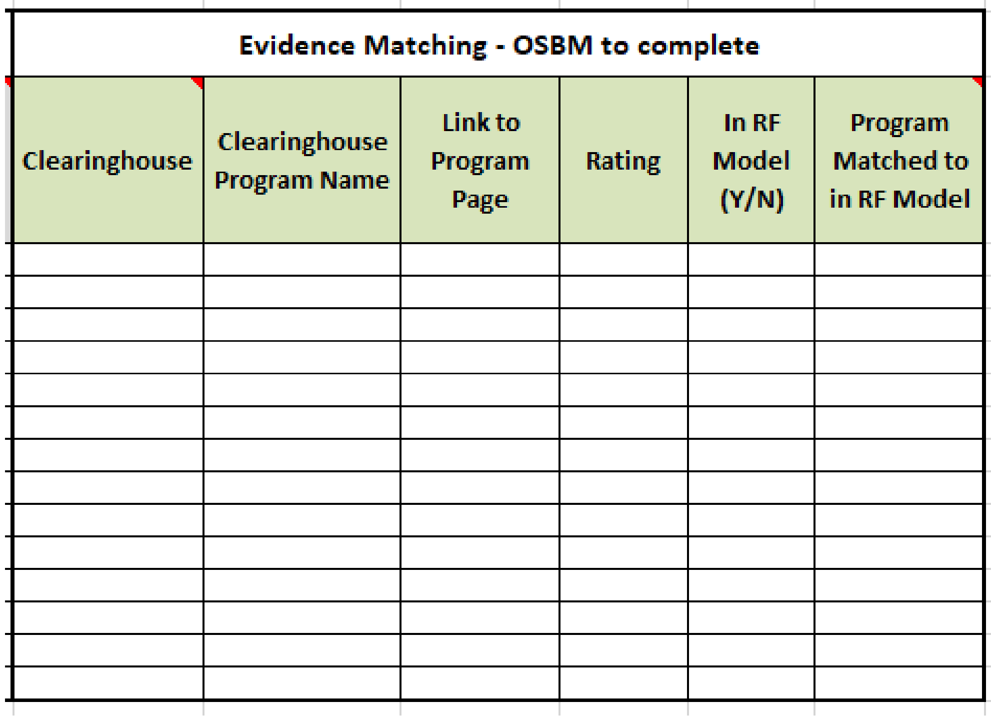
Considerations When Using a Clearinghouse

Appropriate Uses	Issues to Consider
Learning about the effectiveness of currently funded programs.	Clearinghouses do not exist for all policy areas.
Understanding the level of research and/or gaps in research in each area.	Research may be limited within a clearinghouse or may exist outside of the clearinghouse.
Generating policy, program, and budget options for new or existing programs.	Each clearinghouse operates independently and uses somewhat different terminology when reporting results.
Identifying where to prioritize funding for future research.	While the clearinghouses follow similar approaches, each clearinghouse may have slightly different criteria and procedures for rating their programs.

Using the Clearinghouse to Inform Policy



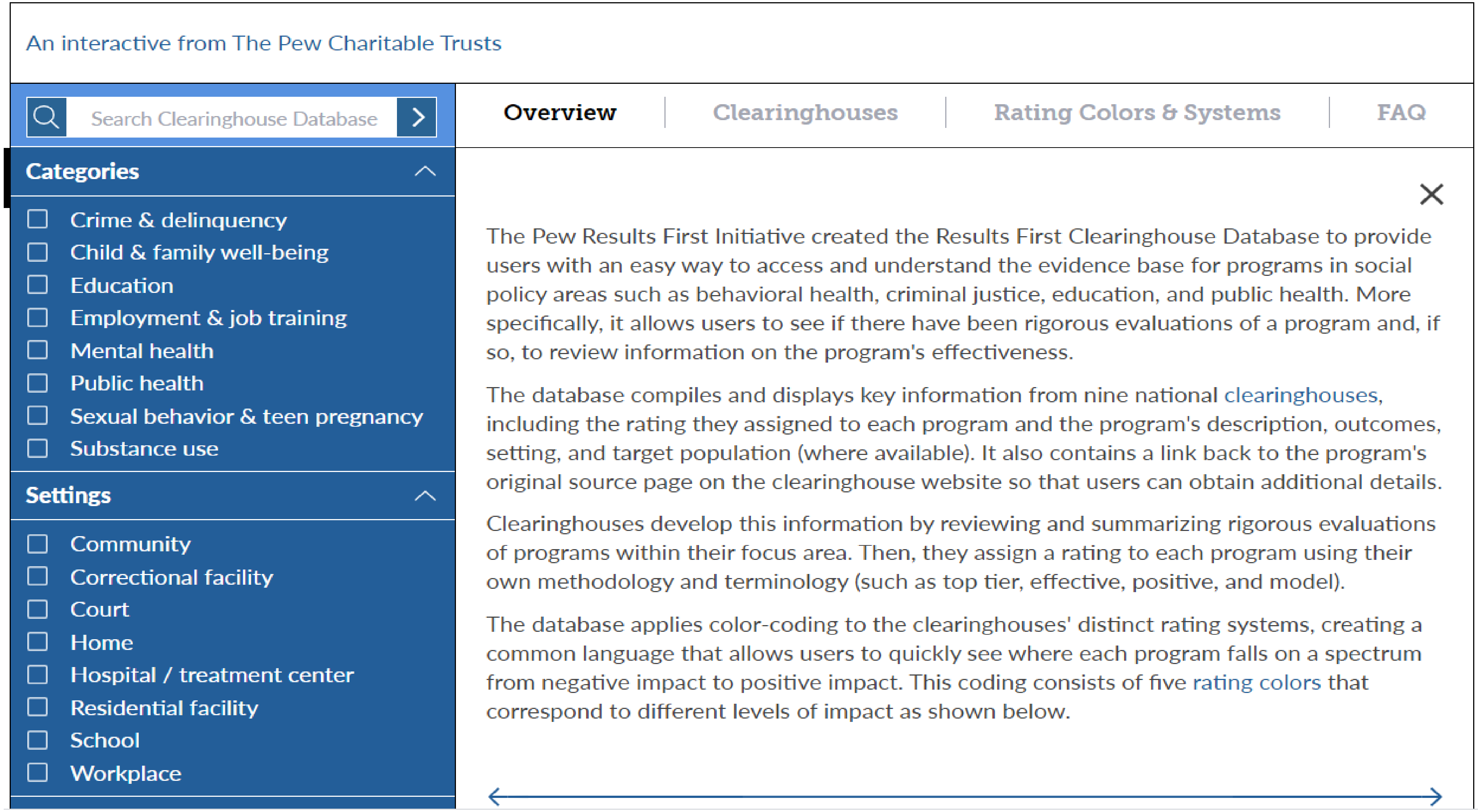
PROGRAM INFORMATION - Agency to Complete (Required)								
Program Name	Program Description	Average Duration of Program	Frequency/ Intensity of Program	Delivery Setting	Target Population	Oversight Agency/Department (e.g., Division of Mental Health)	Service Provider(s)	Provider Credentials
TOTAL								



Evidence Matching - OSBM to complete					
Clearinghouse	Clearinghouse Program Name	Link to Program Page	Rating	In RF Model (Y/N)	Program Matched to in RF Model

This content was developed in consultation with the Pew Charitable Trusts' Results First initiative.

Pew Clearinghouse



An interactive from The Pew Charitable Trusts

Categories

- □ Crime & delinquency
- □ Child & family well-being
- □ Education
- □ Employment & job training
- □ Mental health
- □ Public health
- □ Sexual behavior & teen pregnancy
- □ Substance use

Settings

- □ Community
- □ Correctional facility
- □ Court
- □ Home
- □ Hospital / treatment center
- □ Residential facility
- □ School
- □ Workplace

The Pew Results First Initiative created the Results First Clearinghouse Database to provide users with an easy way to access and understand the evidence base for programs in social policy areas such as behavioral health, criminal justice, education, and public health. More specifically, it allows users to see if there have been rigorous evaluations of a program and, if so, to review information on the program's effectiveness.

The database compiles and displays key information from nine national clearinghouses, including the rating they assigned to each program and the program's description, outcomes, setting, and target population (where available). It also contains a link back to the program's original source page on the clearinghouse website so that users can obtain additional details.

Clearinghouses develop this information by reviewing and summarizing rigorous evaluations of programs within their focus area. Then, they assign a rating to each program using their own methodology and terminology (such as top tier, effective, positive, and model).

The database applies color-coding to the clearinghouses' distinct rating systems, creating a common language that allows users to quickly see where each program falls on a spectrum from negative impact to positive impact. This coding consists of five rating colors that correspond to different levels of impact as shown below.

Clearinghouse Policy Areas

- Aging & Disability
- Behavioral Health & Healthcare
- Criminal & Juvenile Justice
- Economy
- Environment
- Transportation
- Education

This content was developed in consultation with the Pew Charitable Trusts' Results First initiative.

Clearinghouse Example

- Search by a topic term -- Search “obesity” in the search bar.
- Search by program name – type “Big Bro” in the search bar.
- Search by program type - type in Community Kitchen in the search bar.

Tool Spotlight: Google Scholar

Google Scholar provides access to scholarly literature and academic resources that are in Library databases or publicly available databases.

https://scholar.google.com/

QUESTIONS?

Take a Break

Breakout Groups

Recap

- Evidence is defined as "something that supports or challenges a claim, theory or argument".
  - Bottom line questions: Is the program proven to improve outcomes?

- Clearinghouses and Google Scholar are two resources that allow you to search for and assess evidence in a specific policy area.
  - Don’t take them at face value, make sure to dig into a variety of articles and studies.

PM Academy Roadmap

Title	Date	Key Topics Addressed
Performance Management & Setting Goals	**Oct 12**	• Academy introduction/overview • Defining performance management; implementation, benefits, etc. • Linking strategic plans to performance management; decision-making, analysis
Deciding What Evidence is Needed	**Oct 19**	• Overview of evidence 101; impact/process evaluation, etc. •Theory of Change •Generalizability Framework
Finding and Implementing Evidence	**Oct 26**	•Methods 101; Types of evidence, assessing quality, etc. •Searching for Evidence: Using clearinghouses, Google Scholar •Breakouts: Budget Development, Equity in Implementation, Data & Contracting
Observation and Measurement	**Nov 2**	• Surveys, psychology of preference, using administrative data, process mapping, etc.
Changing Minds	**Nov 9**	• Best practices and examples for success • Pre-analysis planning & data visualization • Success stories from NC state government