

---

# Evaluating Evidence

September 6, 2012

Presentation to the OPRE Methodological Advancement Meeting  
Sarah Avellar

**MATHEMATICA**  
Policy Research

# Evidence-Based Decision Making

---

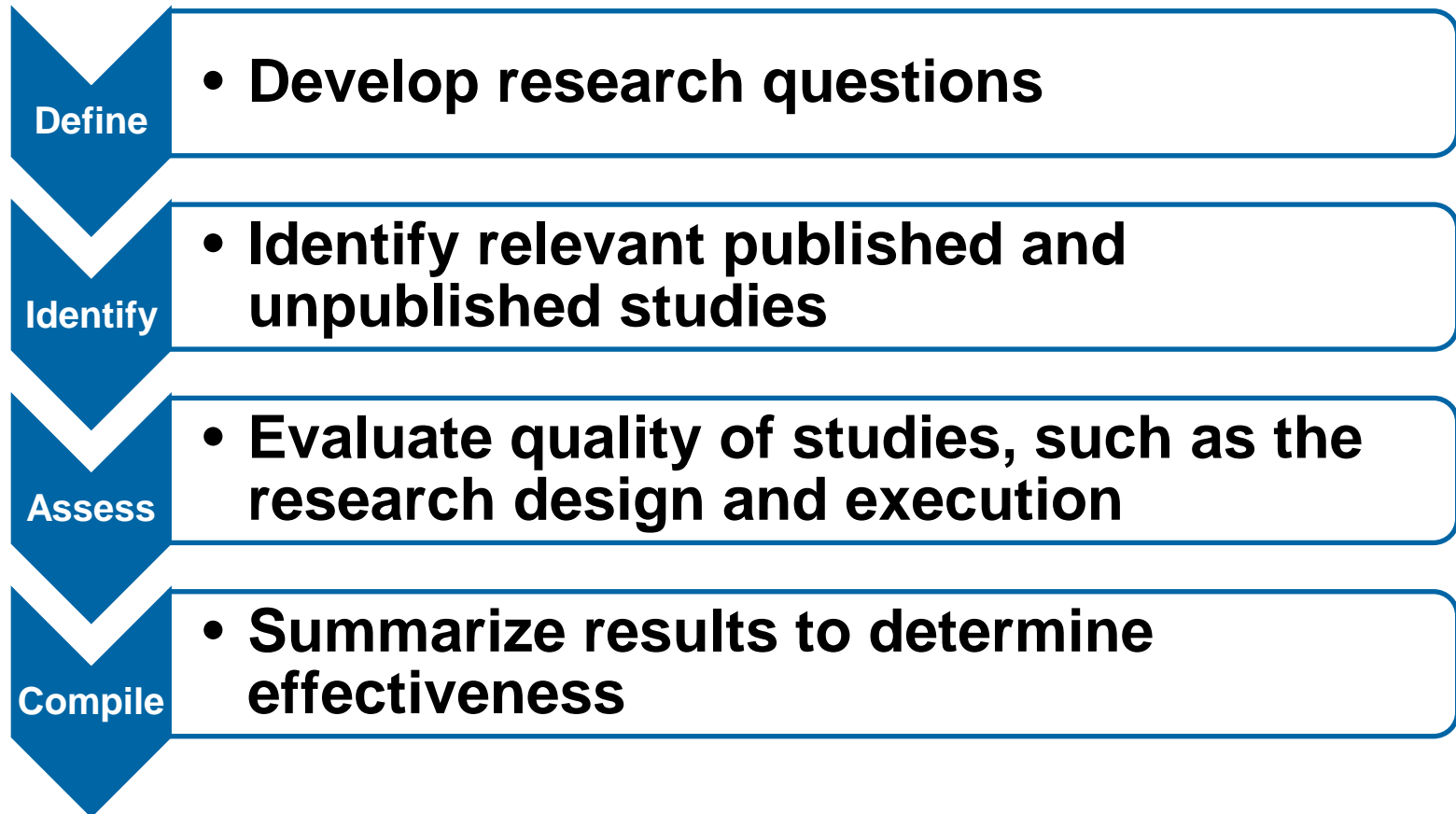
- **Increasing interest in allocating often-scarce resources to programs with established effectiveness**
- **Evidence of effectiveness requires empirical research on impacts**
  - **New research can be tailored to answer questions of interest**
  - **Analyzing existing research may be efficient use of resources**

# What Are Systematic Reviews?

---

- **A systematic review is a comprehensive assessment of existing research to address specified research questions**
  - **Most commonly used to determine whether an intervention or approach is effective**
- **Results may be presented separately (for example, as a range) or aggregated (often using meta-analytic methods)**

# Steps of Systematic Reviews



# Systematic Reviews Are Widely Used

---

- **Federal government is funding systematic reviews on multiple topics**
  - **Examples include education; family strengthening; home visiting; mental health and substance use; medical screening, diagnostics, and therapeutic intervention; and teen pregnancy prevention**
- **Substantial work in other sectors**
  - **Such as health care and policy (Cochrane reviews); education, crime and justice, and social welfare (Campbell Collaboration); and violence prevention (Blueprints)**

# Key Decisions Affect the Results

---

- **Depending on decisions, systematic reviews may lead to different conclusions**
  - Increasing overlap in topics and interventions examined across reviews
- **The overarching questions for reviews and evaluations:**
  - What do we mean by “evidence?”
  - How do we define “effective?”

---

# What Do We Mean by “Evidence?”

# Assessing Causality

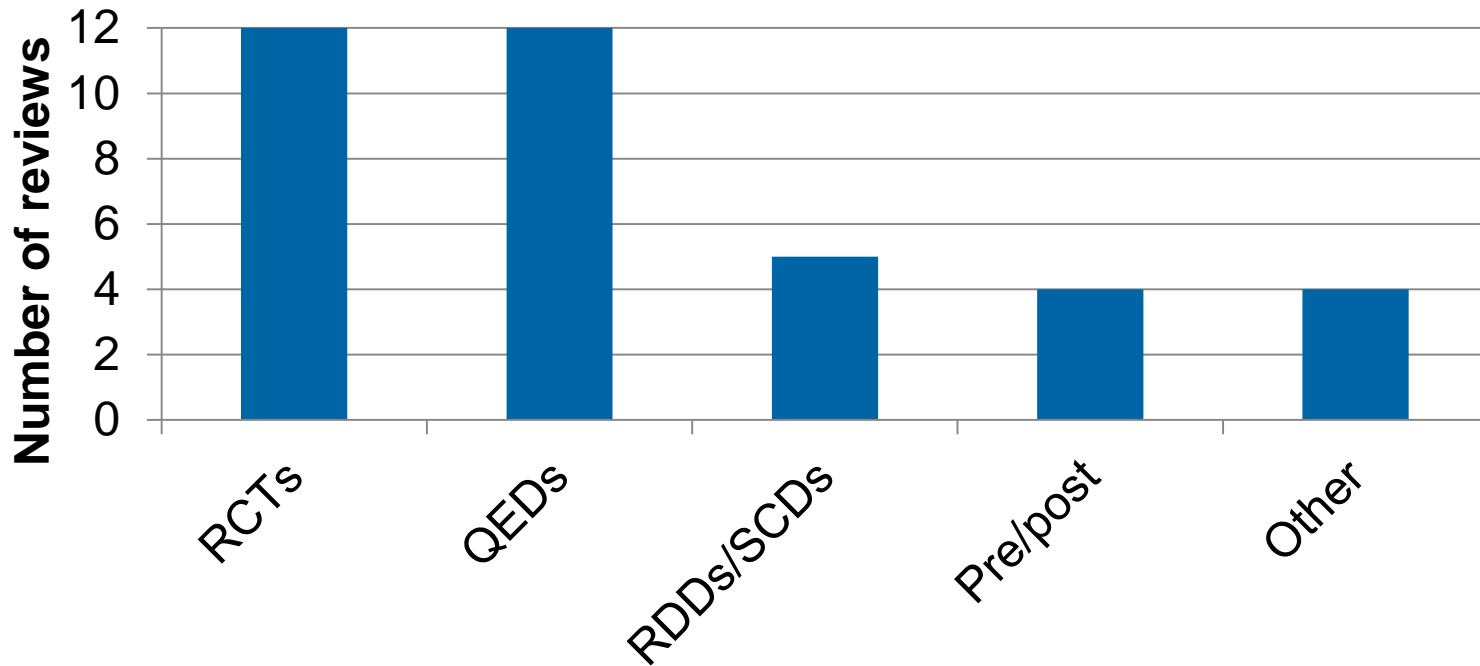
---

- **Most focus on studies' internal validity: the ability to determine whether the effects were caused by the program or intervention**
- **Reviews vary in terms of the study designs they will allow**
  - **Randomized controlled trials (RCTs) always eligible**
  - **Variability in the inclusion of other designs**
    - Quasi-experimental designs (QEDs)
    - Regression-discontinuity designs (RDDs)
    - Single-case designs (SCDs)
    - Pre/post or other designs



# Study Eligibility in Reviews

## Study Eligibility in 12 Federally Funded Reviews



# Standards Within Study Designs

---

- **Reviews also vary in terms of assessments within each type of study design**
- **Across reviews, we know there are differences in areas such as:**
  - **Attrition cutoffs**
  - **Determination of baseline equivalence**
  - **Evaluation of the authors' analyses**

# Differences in Standards: Attrition

---

- **Most reviews account for attrition within RCTs, but cutoffs vary**
  - **Some use a low attrition cutoff with face validity, such as 80%**
    - Empirical basis for the cutoff is not clear
  - **The What Works Clearinghouse (WWC) defined cutoffs based on models of potential bias under different assumptions**
    - Cutoffs also used in other reviews
    - Requires decision about acceptable level of bias

# Differences in Standards: Equivalence

---

- **For QEDs, reviews often require evidence of baseline equivalence**
  - Treatment and comparison groups should be similar at onset
- **What are the important variables?**
  - Many reviews require pre-test measure of outcome
  - In the Home-Visiting Evidence of Effectiveness (HomVEE) review, some programs start prenatally but are interested in child development
    - Equivalence on other moderating variables

# Differences in Standards: Analytic Issues

---

- **The appropriateness of the analysis may be assessed by reviewers**
  - For example, use of control variables
- **Most reviews do not apply post-hoc corrections (e.g., for multiple comparisons)**
  - **Can be unclear how to apply corrections**
    - In HomVEE, for example, follow-ups may extend for years. Should corrections be made within or across follow-ups?

---

# Combining the Evidence

# Challenges of Combining the Evidence

---

- **If there are multiple studies on an intervention, the review must address issues such as:**
  - **Study design and strength of internal validity**
  - **Variations in treatment (components and fidelity)**
  - **Differences in the counterfactual**
  - **Dissimilar sample characteristics and settings**
  - **Varying provider characteristics**
- **Nevertheless, confidence in findings generally increases with the extent of evidence**

# All Reviews Assess Strength of Evidence

---

- **Categorizing the evidence may involve tiers, selected models deemed “evidence-based,” or other options**
  - Decisions are tied to funding in some reviews
- **But the reviews vary in their definitions of categories**
  - Some require an intervention to have an RCT with favorable results
  - Others accept evidence from well-designed QEDs
  - Some require no contradictory findings



---

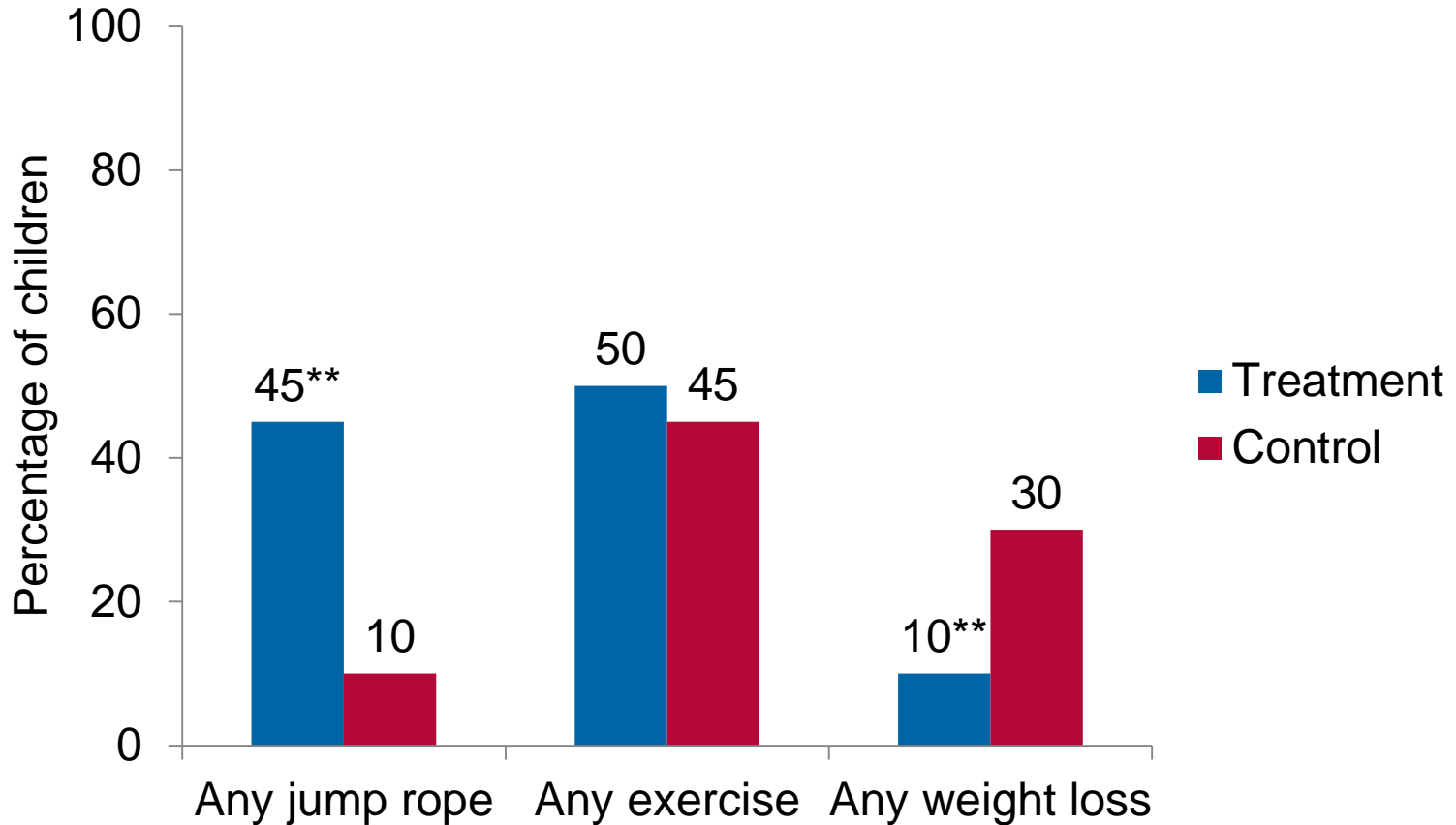
# How Do We Define “Effective?”

# A Hypothetical Example

---

- **The *Jump Start to Health* program is designed to decrease obesity in children living in communities with limited access to fresh food or safe public space for activities. The program organizes group jump-rope events.**
- **The systematic review included the results from three random-assignment evaluations.**

# Is *Jump Start to Health* Effective?



\*\*  $p < 0.05$

# What Findings Are Required?

---

- **Any favorable finding? A favorable finding on any key outcome?**
- **A pattern of favorable findings on similar outcomes? Breadth of findings?**
- **No unfavorable findings?**
- **Do the findings have to be statistically significant or of a certain magnitude?**
- **Impacts observed after the program has ended? For what period of time?**

# What About External Validity?

---

- **Many reviews focus solely on internal validity, but this does not provide information on effectiveness with other populations or in other situations (external validity)**
  - **Criticism is that this information is too narrow to define effectiveness of programs**
- **A challenge is that there are no widely used or accepted standards for assessing external validity**

---

# Moving Forward

# Understanding the Evidence

---

- **Systematic reviews have the advantage and challenge of drawing on multiple evaluations**
- **General agreement on larger issues, such as need for strong internal validity, but variation in detailed standards**
  - Need for transparency in decision making
- **As systematic reviews increase in prevalence, coordination of standards would be beneficial**
  - HomVEE is organizing calls with stakeholders of federally funded systematic reviews

# Refining the Effectiveness Question

---

- **One program is unlikely to fit all**
- **Consider what works, when, and for whom**
  - **Under what circumstances is an intervention effective?**
    - Population or subgroup
    - Setting
    - Organization and staff
    - Treatment variations
- **This nuanced approach is sometimes deemed too complicated, however**



# For More Information

---

- **Please contact:**
  - **Sarah Avellar**
    - [savellar@mathematica-mpr.com](mailto:savellar@mathematica-mpr.com)