What Works for Whom, When and How? Introduction and Overview

Howard S. Bloom
MDRC

This presentation was given at a joint meeting of federal research agencies sponsored by the Office of Planning, Research and Evaluation (OPRE) in the Agency for Children and Families of the U.S. Department of Health and Human Services and held on September 3 and 4 in Washington DC. The presentation is based in part on research sponsored by the W.T. Grant Foundation, OPRE and the Institute for Education Sciences of the U.S. Department of Education. All opinions expressed in the presentation and any errors that it might contain are solely the responsibility of the speaker.

Introduction

The primary goals of program evaluation

- Improving existing programs
- Developing new and better programs

The main focus of past program evaluation

- Assessing program implementation
- Estimating average program effects
- Estimating subgroup program effects
- Fussing over internal validity

Recent changes in program evaluation

- Rapidly growing use of RCTs and QEDs
- Real accumulation of knowledge about average and subgroup effects
- Less fussing over some aspects of internal validity
- More fussing over external validity
- Growing interest in questions about variation in program effects

Core Questions About Variation in Program Effects

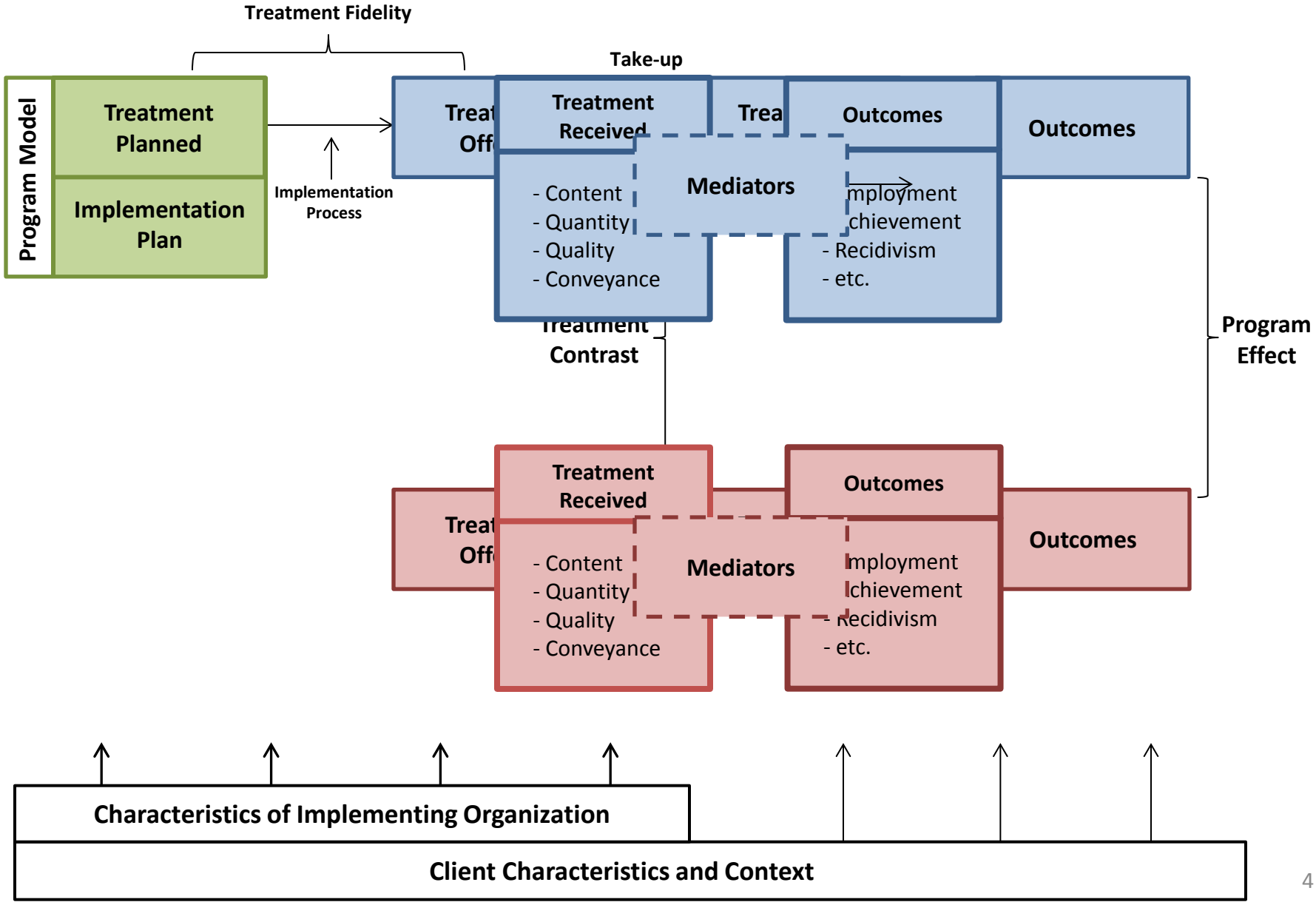
1. By how much do program effects vary across individuals, subgroups and sites?

2. What characteristics of programs, their participants and their contexts predict (moderate) variation in program effects.

3. What mechanisms cause (mediate) program effects and their variation?

4. In short, *what* works for *whom*, *when* and *how*?

A Simple Conceptual Framework



The Present Meeting

Primary focus: *programs, participants, pathways and context*

- Program characteristics
- Participant characteristics
- Causal pathways
- Program context

Less central focus

- Program implementation and fidelity
- The program treatment contrast (especially services to controls)

Part 1

Studying *Natural* Program Variation

Studying Natural Program Variation

Approaches Considered

- Meta-analysis
- Primary or secondary analysis of multi-site trials
- Analysis of multi-site trials *plus* analysis of systematic variation

Meta-Analysis

(Kaminski and Becker)

Unit of observation = study findings

Applications = e.g. medical treatments, mental health therapies, education programs

Estimation model (fixed or random effects)

$$\boldsymbol{Impact\ Estimate} = \boldsymbol{function}(\boldsymbol{program}, \boldsymbol{participants}, \boldsymbol{context}, \boldsymbol{research\ design})$$

Codification of programs (alternative conceptualizations)

- Existing packages (name brand or generic)
- Theoretical approaches (e.g. cognitive behavioral therapy, small learning communities)
- Modules (specified combinations of components)
- Kernels (the smallest possible indivisible program elements)

Meta-Analysis
(continued)

Strengths

- Can draw on large existing research literatures
- Can be updated periodically
- Has a large institutional base (the Cochran and Campbell collaborations)
- Has been influential in medical research and practice

Limitations

- Input limited to reported study findings (not study data)
- Often there is no information about services to controls
- Difficult to account for all missing unpublished studies
- Currently limited mainly to effects of program assignment (ITT)
- Based on a non-experimental model of variation in study findings

Multi-Site Trials

Approach = use multi-sites trials as a "fleet" of RCTs (or QEDs)

Unit of Observation = individual sample members

Applications = welfare-to work, Head Start, charter schools, small high schools in New York

Estimation Model (random effects illustration)

Level One: Sample Members

Outcome = function(site, treatment assignment, baseline covariates, baseline covariates interacted with treatment assignment)

Level Two: Sites

Mean Impact = function(program characteristics, site characteristics)

Multi-Site Trials

(continued)

Strengths

- Greater flexibility in model specification given individual data
- Greater access to measures that were created but not reported
- An ability to amass a large fleet of RCTs (or QEDs) across studies
- Primary data collection can be tailored to the theory of the program
- The approach can be used to study effects of program assignment (ITT) and program participation (LATE)
- A growing number of multi-site trials are being conducted

Limitations

- Need a relatively large number of sites
- Methods of analysis are currently evolving and not widely known
- Findings are based on non-experimental models of variation in site findings

PS:

Joint Spencer Foundation, W.T. Grant Foundation and IES project on learning *about* and *from* variation in program effects using data from existing multi-site trials.

Using Natural Variation from a Multi-site Trial With Planned Variation from a Multi-arm Trial

(Harvill)

Basic Approach

- Imbed a multi-arm trial of specific combinations of program components (*to induce planned variation*) in a subset of sites for a two-arm RCT that uses *natural variation* across sites to study variation in the effects of different program components.
- Use results from the planned variation analysis to assess and adjust for bias in the larger analysis of natural variation.

Application

- The Health Professions Opportunity Grant (HPOG) evaluation
- Planned variation across arms of a randomized trial with: (1) a no-program control group, (2) a standard program treatment group and (3) an enhanced program treatment group.

Part 2

Studying *Planned* Program Variation

Studying Planned Program Variation

Approaches Considered

- The Multi-Phase Optimization Strategy (MOST)
- Sequential Multiple Assignment Randomized Trials (SMART)
- Rapid Cycle Evaluations

The Multi-Phase Optimization Strategy (MOST)

(Collins)

Goal

- To *develop* a cost-effective *combination* of program components
- Focuses on *relative* effects of different combinations of program components

Applications

- Smoking cessation program
- Staff training for an HIV prevention program
- A coaching program for Head Start staff members

Approach

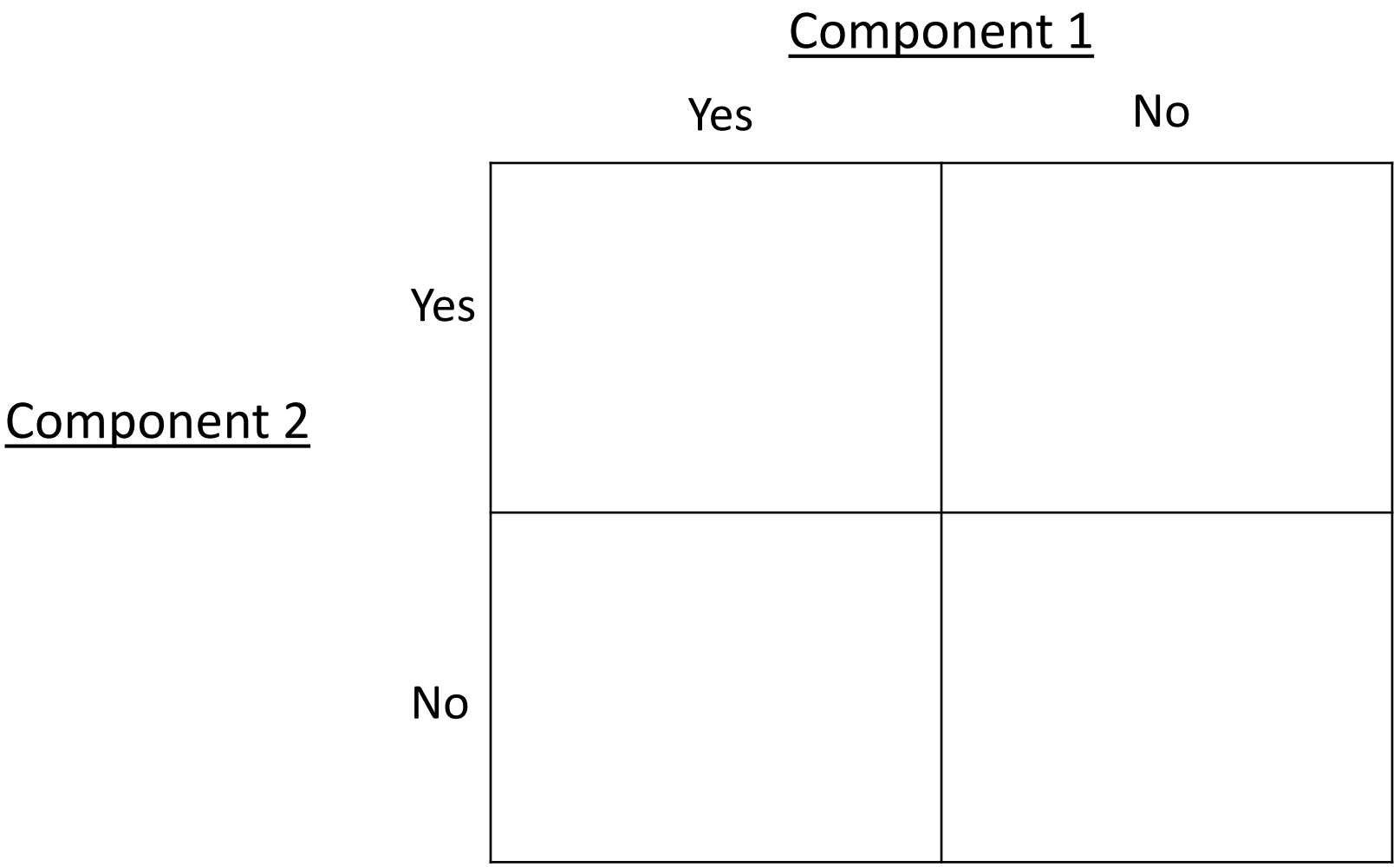
- Systematically *test* alternative combinations of program components to develop the most effective combination given existing constraints
- Can use many different evaluation designs
- In some cases factorial experiments can be a natural for this

The Multi-Phase Optimization Strategy (MOST)
(continued)

Factorial Designs

- Randomize sample members (or clusters) to specific combinations of program components
- This enables one to estimate main effects for all components and interactions for all or many combinations of components (depending on the design)
- A full factorial design enables one to estimate all interactions
- A fractional factorial design enables one to estimate some interactions; the others are conflated with main effects and other interactions.

The Simplest Factorial Design



The Multi-Phase Optimization Strategy (MOST)
(continued)

Strengths

- Efficient use of sample (e.g. you can use the *full sample* to estimate the main effect of every component)
- Conventional ANOVA and ANCOVA methods exist for analyzing ITT effects

Limitations

- Implementing many versions of a program is difficult.
- The number of program configurations grows exponentially with the number of components. (and the number of levels for those components).
- To reduce the number of program configurations through a fractional factorial design one must assume that certain interactions are negligible.

Question

- Have methods been developed to deal with treatment assignment non-compliance?

Sequential Multiple Assignment Randomized Trials (SMART)

(Kidwell)

Purpose

To design an adaptive intervention which has the best *combination* and *temporal ordering* of program components

Adaptive Interventions

- Respond to changes in client outcomes and
- Comprise different combinations and time-orderings of program components

Example

- ADHD therapy comprising medicine and behavior modification

Rapid Cycle Evaluation
(Cody)

Purpose

- Real time feedback for program *improvement*
- More about comparing outcomes for different forms of a program than about estimating the net effect of an existing program.

Approach

- Predictive analytics to assess who is most likely to benefit from a program modification
- Real-time testing of a new program or modifications to an existing program
- Can use RCTs or QEDs
- It is like the preceding approaches only it occurs more quickly

Requirements

- Quick response to treatment (e.g. for studies of program intake, participation and persistence)
- Ready access to quality data for large samples

Segue

What was the focus to this point?

- Program components (the “what”)

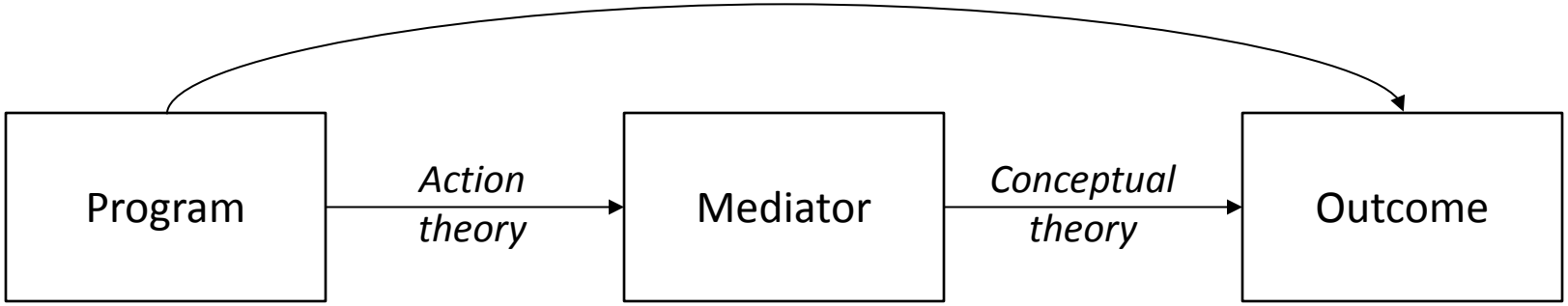
What comes next?

- Moderators (the “whom” and “when”)
- Mediators (the “how”)
- Combinations of mediators and moderators

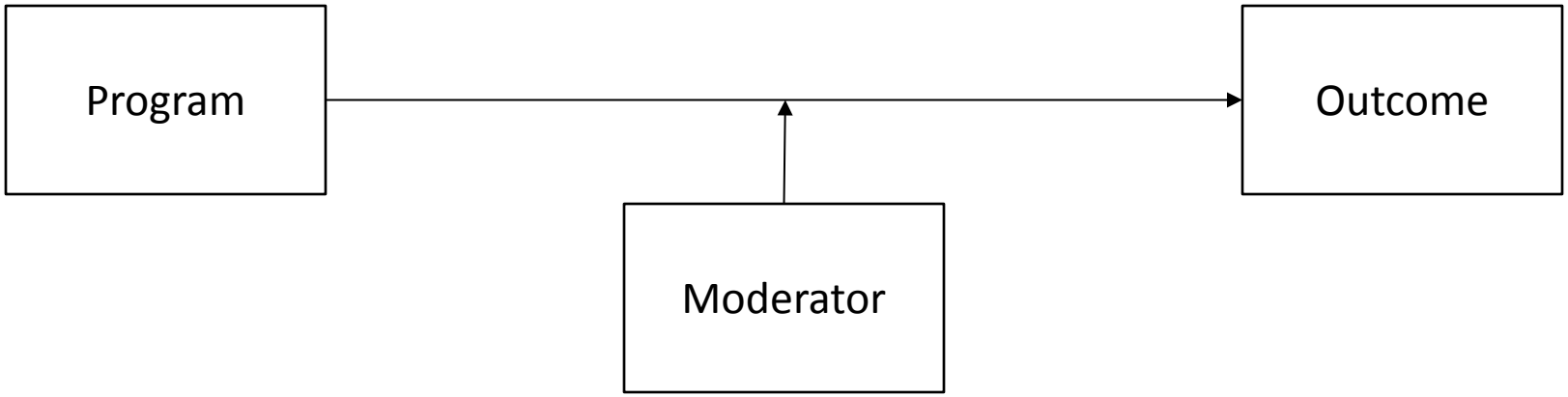
Part 3

Studying Variation in People and Context

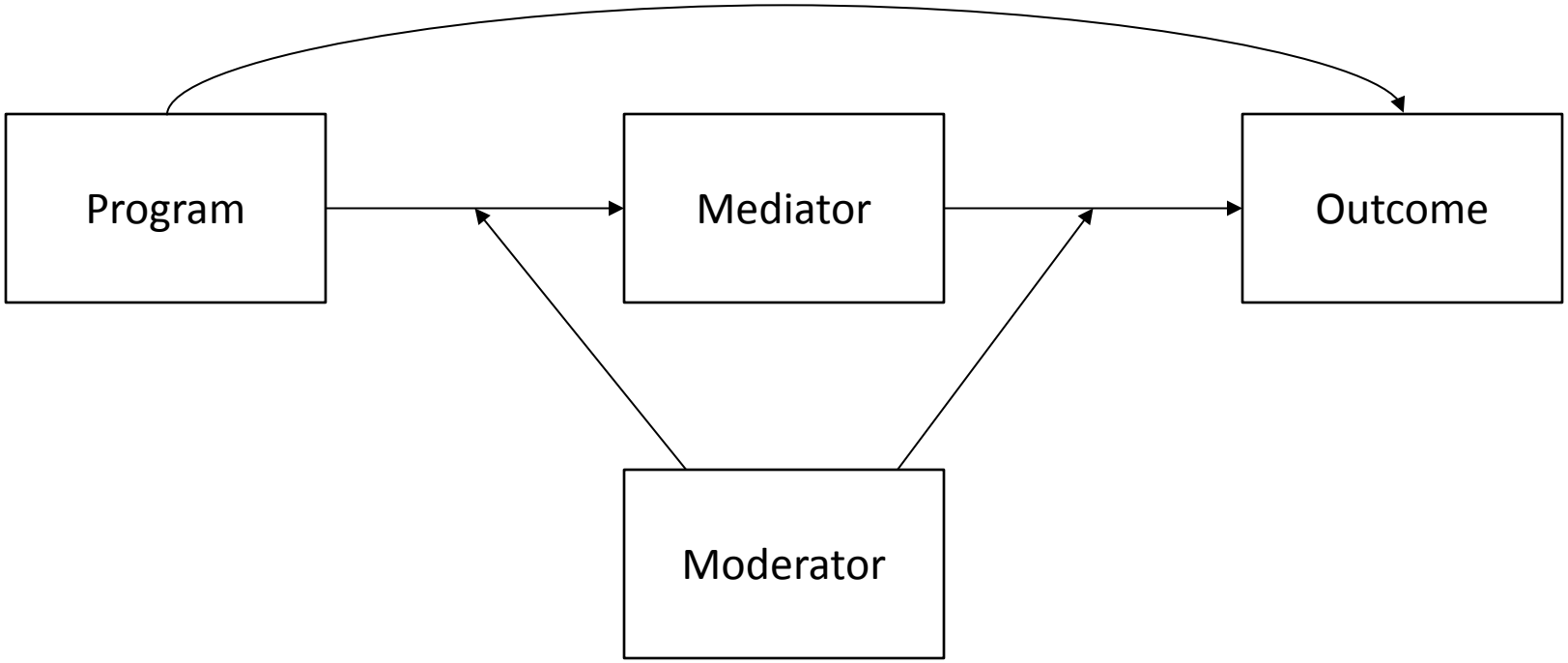
Mediator Model



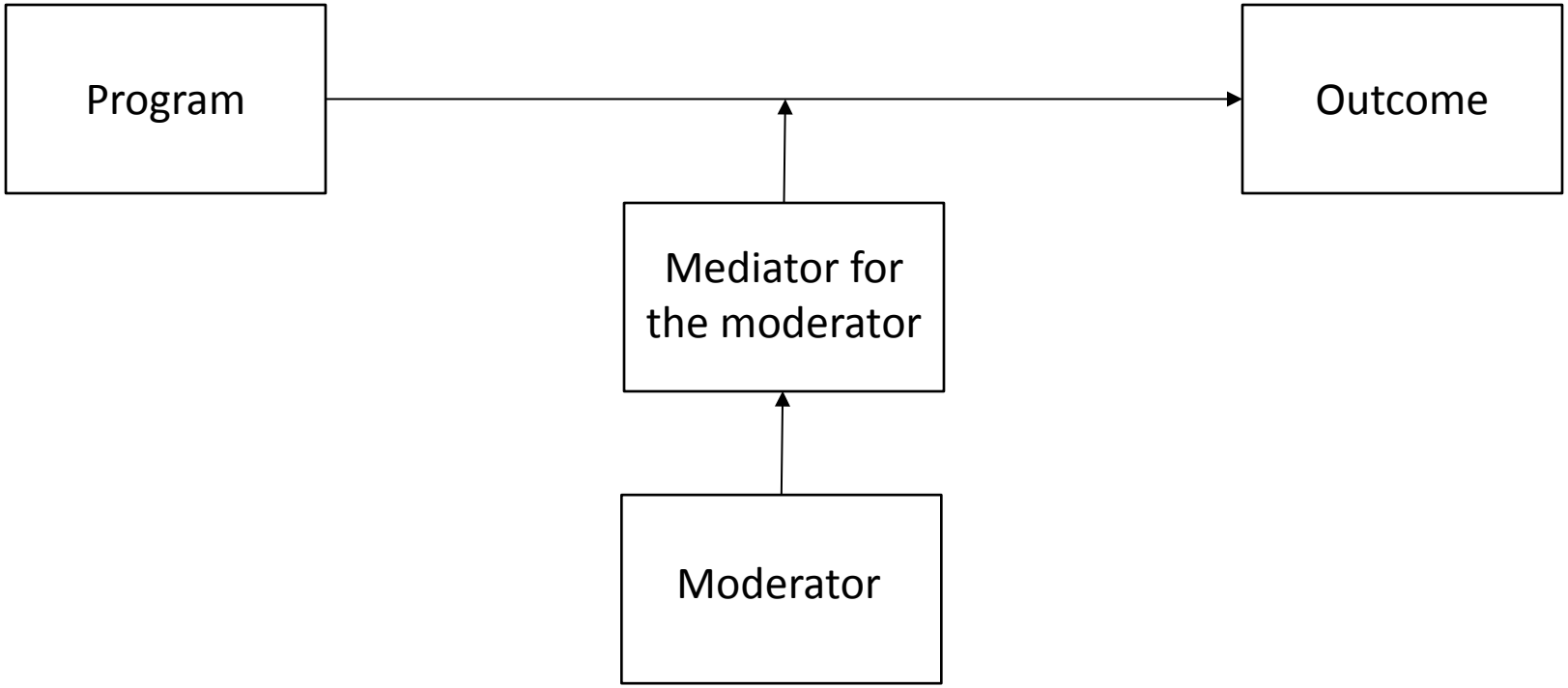
Moderator Model



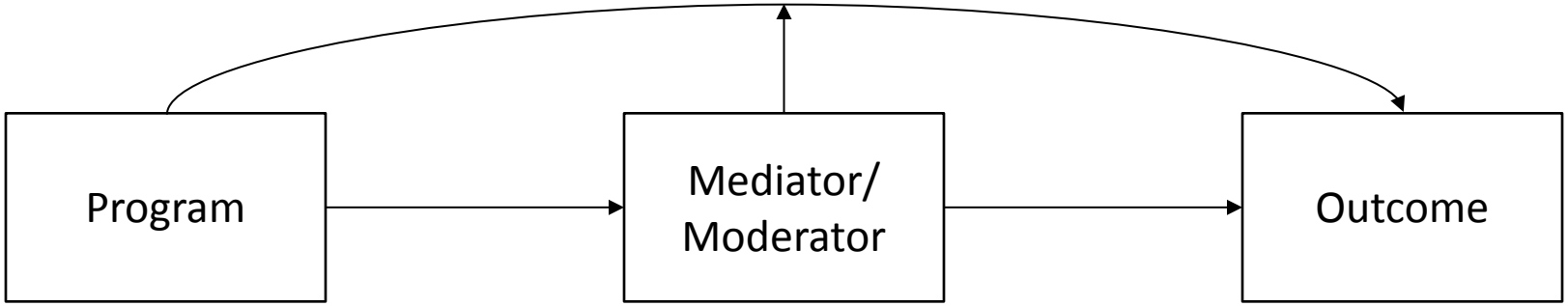
Moderated Mediation Model



Mediated Moderation Model



Mediator/Moderator Model



Studying Variation in People and Context

Exogenous subgroups

- Defined by characteristics of clients or settings that cannot be affected by treatment assignment
- Subgroups should be specified *a priori* based on theory and/or policy relevance
- One should test the statistical significance of *differences* between subgroup effects
- One should pay attention to multiple hypothesis testing

Endogenous subgroups

- Defined by characteristics of clients or settings that can be affected by treatment assignment

- These characteristics represent counterfactual intermediate or final outcomes under:
  - the control condition (e.g. students who without program assignment would drop out of school)

  - the treatment condition (e.g. students who with program assignment would reach a specified milestone)

  - both conditions (e.g. children who if assigned to Head Start would participate in it and if not assigned to Head Start would be cared for at home)

Studying Variation in People and Context
(continued)

Central problem: Endogenous subgroup members cannot be directly observed in both the treatment group and the control group.

Basic approach:

- Predict members of endogenous subgroups and estimate their program effects experimentally
- Adjust these estimates by assumption and modeling for errors in subgroup prediction

Interpretation of findings:

- Without further assumptions results represent moderation of program effects
- With further assumptions, results can represent mediation of program effects

Studying Variation in People and Context
(continued)

Analysis of symmetrically predicted subgroups (Peck)

- Endogenous subgroups based on one counterfactual outcome
  - Example #1: If a student were not assigned to the program he would drop out of school (control condition)
  - Example #2: If a student were assigned to the program he would reach a specified milestone (treatment condition)
- Estimation = variants of OLS and IV

Principal stratification analysis (Page)

- Endogenous subgroups based on two counterfactual outcomes
  - Example: If a student were assigned to the program (treatment condition) he would reach a specified milestone and if he were not assigned to the program (control condition) he would drop out of school.
- Estimation = Maximum likelihood analysis of Bayesian models

Studying Variation in People and Context
(continued)

Neighborhood effects (Kling)

- This mediator analysis addresses the generic question: *Through what causal paths does a given program produce its observed effects?*
- The analysis uses data from a *multi-site, multi-arm trial* (Moving to Opportunity, MTO) to create multiple instrumental variables
- Two mediators are examined: (1) neighborhood poverty and (2) neighborhood segregation
- These instruments are used to estimate effects of the two mediators on the subjective well-being of sample members

Part 4

Uncovering Steps in the Causal Chain

Uncovering Steps in the Causal Chain

Research question

- Through what causal paths does a given program produce its observed effects?

Key limitation:

- You often cannot randomize the mediators of interest (but you should do so when you can).

A**pproaches**

- IV analysis of multi-site RCTs (Reardon)
  - Defining assumption = exclusion (the specified mediators account for *all* of the program's effect on the outcome of interest)

- Causal mediation analysis of RCTs (Keele)
  - Defining assumption = sequential ignorability (it is *as if* the mediators were randomized within experimental groups)

- Moderated mediation analysis (Fairchild)

Part 5

Implications for Policy and Research:

Where Should We Go Next?