Research Designs for Complex, Multi-level Health Interventions and Programs

May 4-5, 2004

Building 31C, 6th Floor, Room 10
National Institutes of Health
Bethesda, Maryland

Borrowed from many

- Barbara Devinney
- Shawna Mercer
- Larry Green
- Denise Dougherty
- Henry Feldman
- Steve Gortmaker
- David Murray
- Marshall Chin
- Wilt Shaddish
- Ten other presenters and others team members

Outline of Presentation

- Background and definitions
- Key Issues – independent of specific designs
- Two research designs
  - Time series
  - Regression discontinuity
- Closing comments on Trade offs

Background and Definitions

Experimental and Quasi-Experimental Designs for Generalized Causal Inference

By Shaddish, Cook, and Campbell

Definitions

- Validity – the correctness of or degree of support for an inference
  - **Internal** the validity of inferences about whether the relationship two variables is causal
  - **External** – does validity holds over variations in persons, settings, and perhaps other variables
- Bias – systematic error in an inference

More interest in alternative designs for different types of study goals

- **Efficacy** – How well an intervention works when implemented under ideal conditions
- **Effectiveness** – How well an intervention works when implemented under conditions of actual application especially in communities or health care systems -
- **Translational research** – What variables influence the rate of implementation of an effective intervention and host of other important questions – there is another definition of translational: lab to academic bedside

Characteristics of situations where classical RCTs are more difficult

- The intervention will vary from one subject-organization-community to another
- The efficacy trial's intervention is difficult to carryout in non research settings
- Control group is likely to start adopting the treatment before the end of the research
- Two choices: adapt the RCT or consider the alternative study designs

Key issue independent of specific designs

Phases of Research and Choice of Design

Efficacy

Develop Hypo. ⟷ Test of Hypo

Effectiveness

Develop Hypo. ⟷ Test of Hypo

Scenario 1: Prevention of Type 2 Diabetes

Marshall Chin, MD

Sanford A. Garfield, PhD

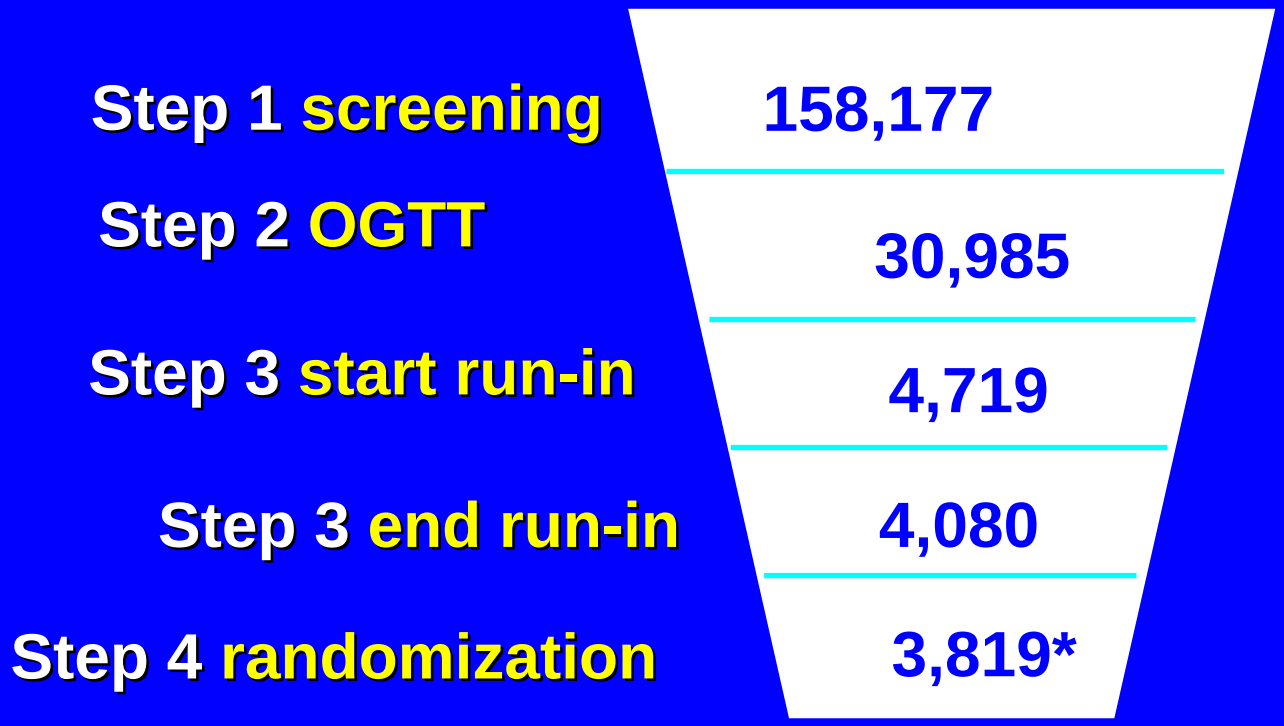
Carol Mangione, MD, MSPH

K.M. Venkat Narayan, MD, MPH, MBA

Myrlene A. Staten, MD

Screening and eligibility

Number of participants



*3,234 in 3 arm study

Lifestyle Intervention

An intensive program with the following specific goals:

- **≥ 7% loss of body weight and maintenance of weight loss**
  - **Calorie intake goal -- 1200-1800 kcal/day**
- **≥ 150 minutes per week of physical activity**

Lifestyle Intervention Realistic in Practice

- 16 session individual core curriculum (over 24 weeks)
- Long-term maintenance program
- Supervised by a case manager
- Access to Lifestyle support staff
  - Dietitian
  - Behaviorist
  - Exercise physiologist

Next question

- Diabetes Prevention Program (DPP): Intensive lifestyle intervention (diet and exercise) reduces relative risk of DM by 58% over 3 years
- *But, can it be translated to real world settings??*

Phases of Research and Choice of Design

Efficacy

Develop Hypo. ⟷ Test of Hypo

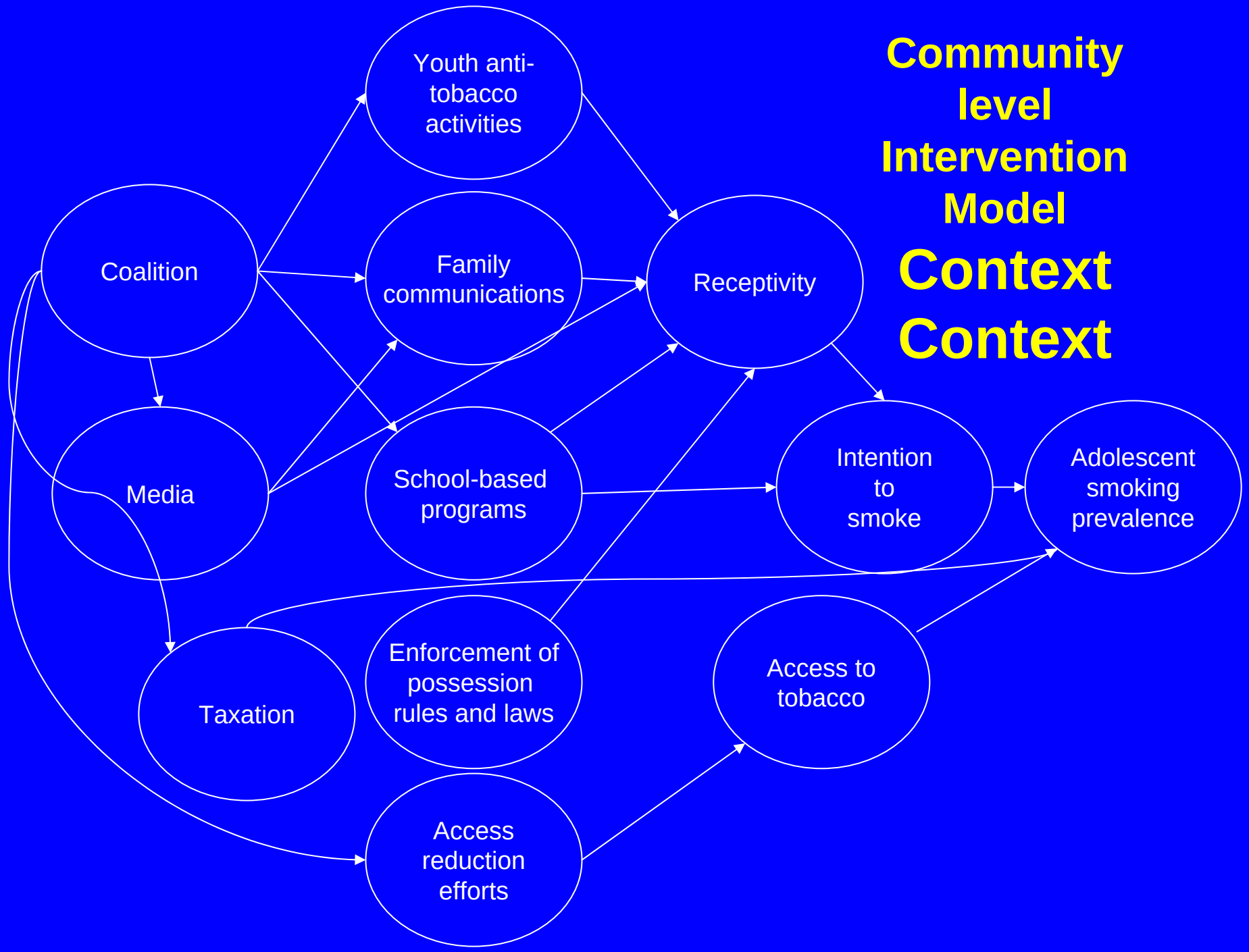
Effectiveness

Develop Hypo. ⟷ Test of Hypo

Good news can use weight loss

Community level Intervention Model

Context Context



Phases of Research and Choice of Design

Efficacy

Develop Hypo. ⟷ Test of Hypo

Effectiveness

Develop Hypo. ⟷ Test of Hypo

- Larry check the time move to slide 29 if necessary

Challenge of Translating DPP to the Community

- Enrolling more generalizable population
  - Funnel of study enrollment for DPP [Effectiveness question hypothesis testing]
- DPP lifestyle intervention intensive – realistic?
  - Training of study personnel and Intensity of intervention [Effectiveness question hypothesis development]
  - Sustainability

Research Designs – RCT's Improving acceptability to organizations and increasing external validity

Scenario 1: Prevention of Type 2 Diabetes

Marshall Chin, MD

Sanford A. Garfield, PhD

Carol Mangione, MD, MSPH

K.M. Venkat Narayan, MD, MPH, MBA

Myrlene A. Staten, MD

Primary Study Question

- Can community interventions designed to increase physical activity and change diet prevent the onset of type 2 diabetes among overweight and obese persons with pre-diabetes?

Design 1: Randomized Encouragement Trial (RET)

- Retains experimental structure but emphasizes a pragmatic public health perspective
- Combines strengths of the RCT and Observational studies
- Instead of mandating treatment assignment, randomizes participants to *encouragement* for the target intervention
- Promotes a more equitable relationship between the researcher and the participant/community

Duan N, et al. Randomized Encouragement Trial: A Pragmatic, Public Health Oriented Paradigm for Clinical Research. *In preparation 2004*

Study Population

- Study setting: Churches in African American and Latino communities of Chicago and Los Angeles; Aim 50 people per church

- Pre-diabetes: Impaired fasting glucose ($\geq$ 100 to 125 mg/dl), age $\geq$ 50 yrs, BMI $\geq$ 30
  - Fallback: Overweight or obese with DM risk factors?

What is Encouragement?

- Offer of resources, incentives, education, and communication (persuasive messages) designed to increase the probability that a participant will want to adopt the treatment
- Encouragement strategies can be developed collaboratively with communities and the population of interest

Why Encouragement?

- Attempts to influence treatment adoption through participants' autonomous choice, leaving ultimate decisions to the participants
- Choices are voluntary
- Some participants might reject all menu choices in the intervention, some might select some
- Some controls may figure out how to get access to the intervention through other means

Physical Activity Menu

- Self-monitoring, pedometer
- Buddy system walking program
- Group walking sessions
- Collaborate with local Y or public parks
- Exercise classes taught by high school / college congregants
- Other suggested by community participants
- ***Particpants are encouraged to select activities from the list that they feel will work best for them***

RET Analyses

- Assuming that the encouragement increases treatment adoption, it can provide an evaluation of treatment effectiveness using an intent-to-treat analysis
- Provides important findings with regard to adoption and what is desirable and feasible in the community
- May have
  - stronger external validity than the classical RCT
  - stronger internal validity than observational studies

Randomized Encouragement Trial Weaknesses

- Better test of strategy rather an specific element of an intervention strategy
- By its less controlled nature RETs tend to have smaller effect sizes and greater within group variance therefore require bigger sample sizes. [**4 times more sample needed for the same power!**]

Research Designs – Interrupted Time Series

Related designs included repeated time series and multiple baseline studies

Scenario 4: Tobacco Control and Cessation Among Adolescents

Anthony Biglan, PhD

K. Stephen Brown, PhD

Brian R. Flay, D.Phil

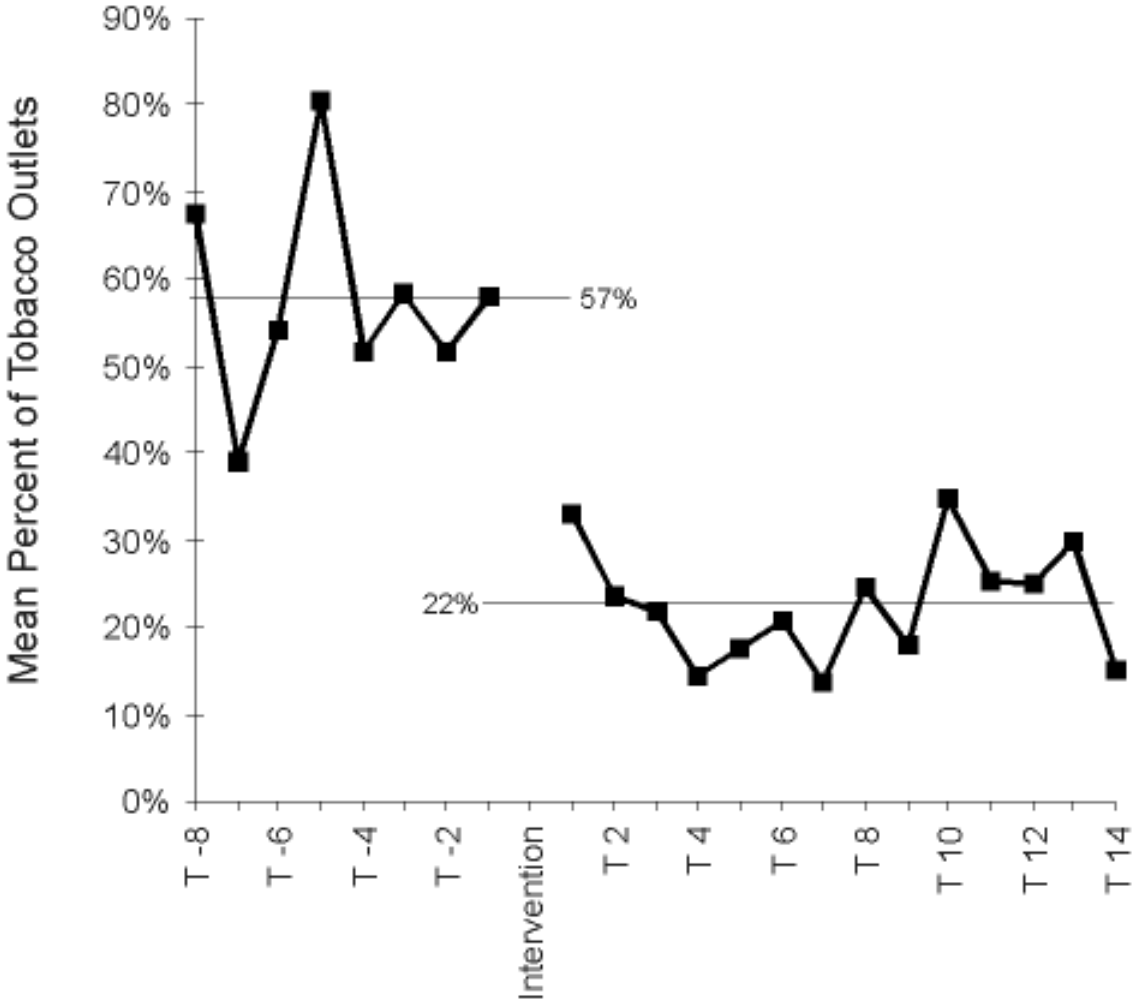
Rachel Nugent, PhD

Terry Pechacek, PhD

Evaluating Access Reduction

- Proclamation
- Reward and Reminder Visits
- Feedback to Outlets

- Enforcement of the Law

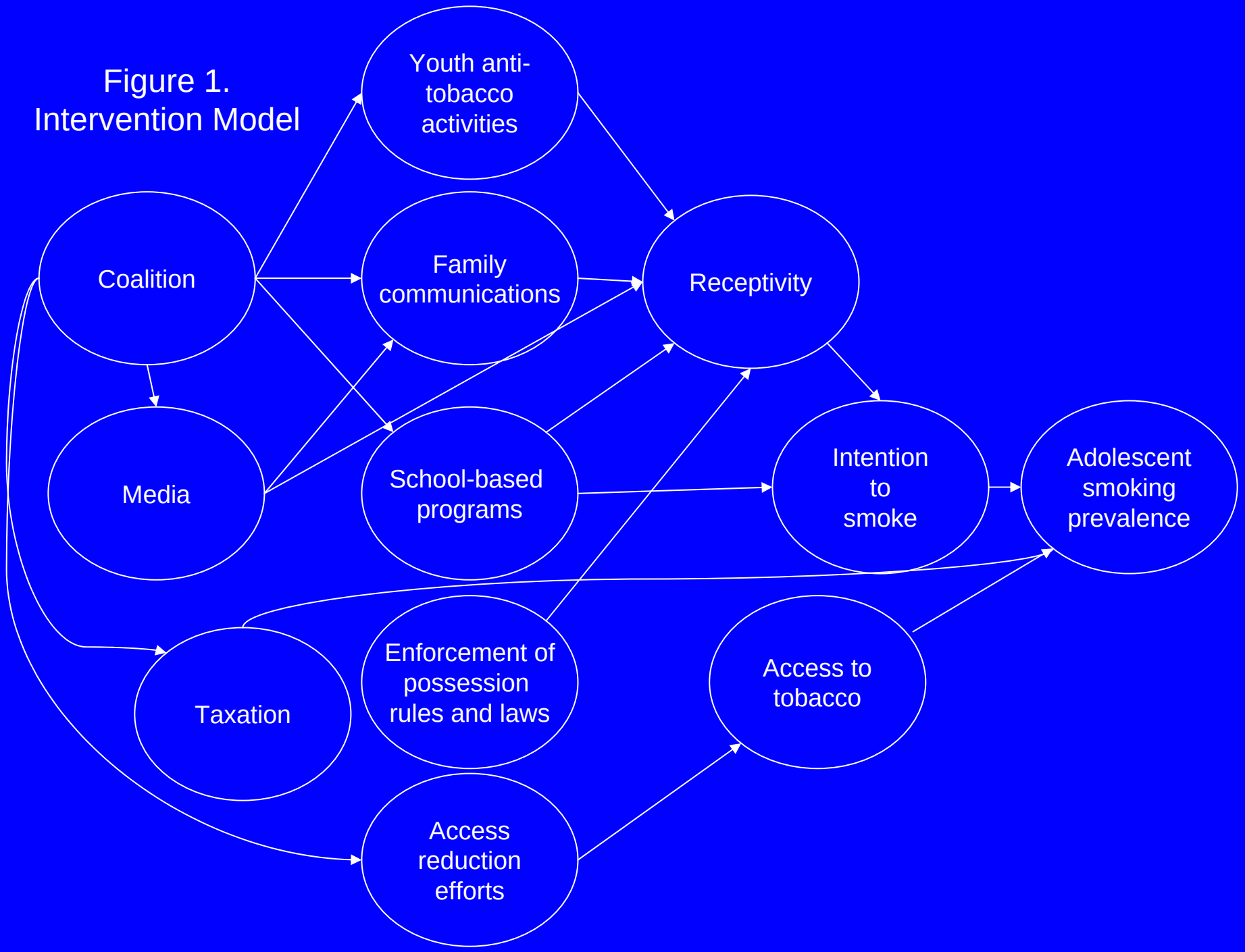


Mean percent of tobacco outlets in eight communities willing to sell tobacco to minors.

The Target

- Reduce the Prevalence of Adolescent Tobacco Use

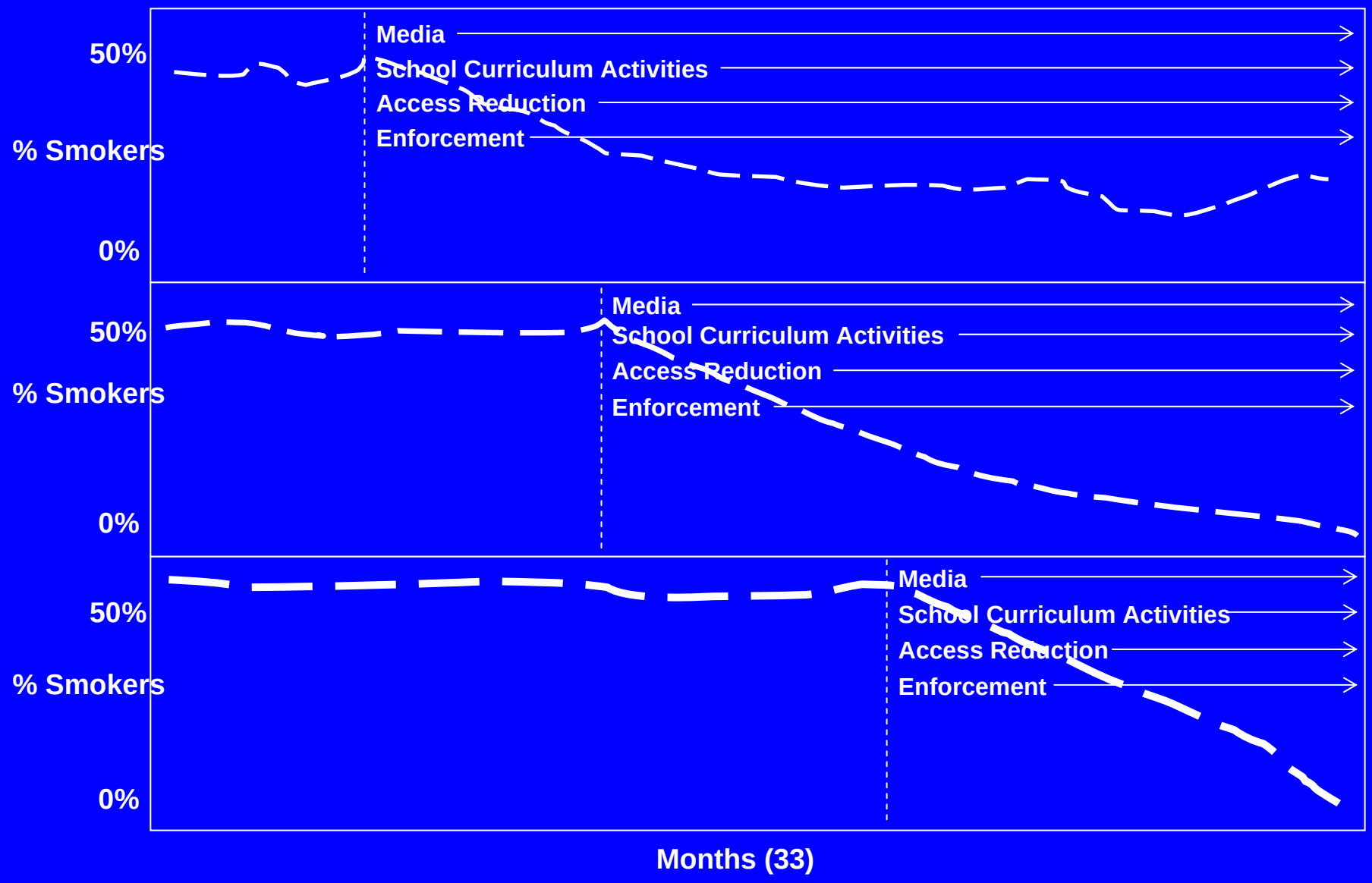
Figure 1.
Intervention Model



Components of the Intervention and Their Proximal Targets

- Mass media messages—reach and frequency; susceptibility
- School curriculum/activities to promote non-smoking—reach and frequency; susceptibility
- Access reduction—percent of outlets willing to sell.
- Enactment and enforcement of laws prohibiting tobacco possession—enactment, citations, perception of likelihood of consequences for possession
- Family communications prohibiting tobacco use—parent-child interactions; youth perception of parental disapproval of smoking.

Effects of Intervention on Adolescent Smoking Prevalence



Two Possible Purposes of the Research

- To test the Efficacy or Effectiveness of a comprehensive intervention.
  - Assumes that we have a set of potent components that reliably affect smoking or the factors that strongly influence smoking.
- To develop intervention components that strongly affect smoking or potent influences on smoking.

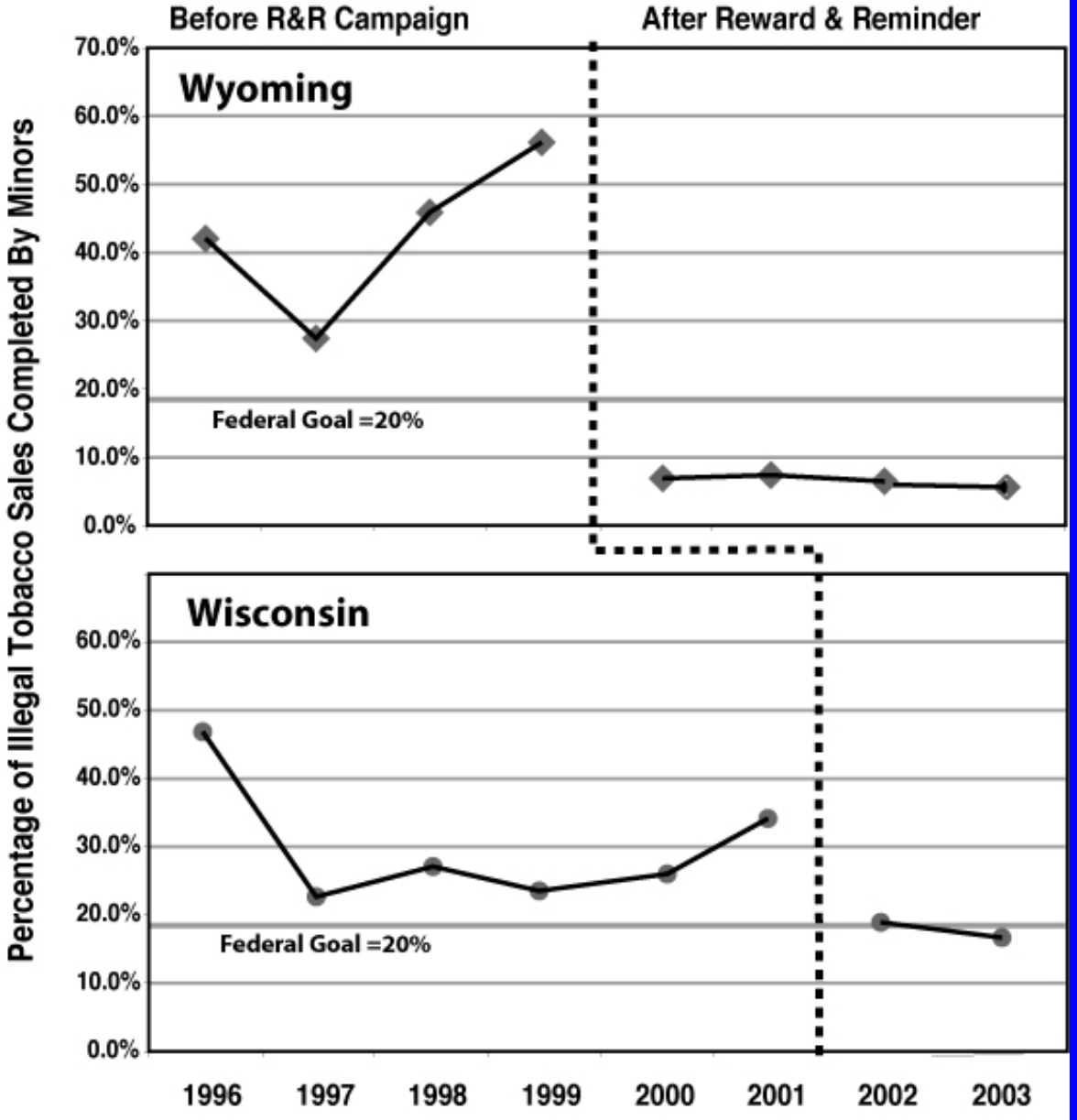
Community or health care system

- Adaptation will be common
- Our ability to predict or select successful potent intervention is limited
- Even with potent interventions we are limited in our ability to predict their potency across communities, time periods, cultures, etc.
- Complex nature of the problem and relative weak potency of our interventions and need to secure buy in – no easy solution

Randomized Controlled Trials

- A Good Way to Test Efficacy or Effectiveness
- A poor way to develop or adapt potent intervention components.
- Easy to suggest hard to get funded

Interrupted Time-Series Designs



Note: Tobacco sales data are independently collected by universities in the respective states as official federal reports for the Synar Amendment, unconnected to the intervention strategy or R&R teams in the states.

Time-series Analysis

- **Synonyms: repeated measures, longitudinal analysis**
- **Origin in economics, insurance, weather, social statistics**
- **Epidemiological applications:**
  - **Surveillance data**

  Repeated visits or observation periods

  Useful for possible changes provided there relevant tracking data

Hallmarks of Time-series

- **Long string of regular data > 100**
- **Aggregated social or physical units**
- **Several related variables – [nonequivalent data]**
- **Noisy data, serial correlation**
- **Advantage over discrete observation: *detection of trend***

Scenario 3: Promotion of Physical Activity

Ross C. Brownson, PhD

David Buchner, MD, MPH

Deborah A. Cohen, MD, MPH

Henry A. Feldman, PhD

Denise Simons-Morton, MD, PhD

TEA-3 Transportation Enhancements & Physical Activity:

Deborah A. Cohen
Ross Brownson
Henry Feldman

What designs might be used to study the relationship between transportation enhancements and physical activity?

Why a Q-E Design?

- Random assignment is not possible
  - Lack of control over the "intervention"
- Not enough units of analysis – six new stations opening up

Design 2 – Time-series

- 5-year study
  - Phone interviews during first week of each month [60 points]
- Six locations where light rail being built
  - Ea. site is its own comparison group prior to light rail

Research Questions

- *Primary aim:*

1. What is the relationship between a transportation enhancement, specifically construction of light rail lines, and rates of moderate physical activity?

Key Variables

- Dependent variable
  - Physical activity behavior
    - Self-report, multiple domains [cannot blind]
    - Motion sensors [objective measures]
- Moderators/mediators
  - Distance from transit
  - SES
  - Threats to internal validity ?

Threats to Interval Validity

- Other interventions
- Selective movement into area pre or post station opening

Adding other design features reduce threats to internal validity

- Nonequivalent dependent variables – differential effect by the intervention but not the threat
- Removing the intervention at a known time
- Multiple replications – good for acute effects: ABABABABA -STATINS
- Switching replications – metro stations, states, screening for PKU,

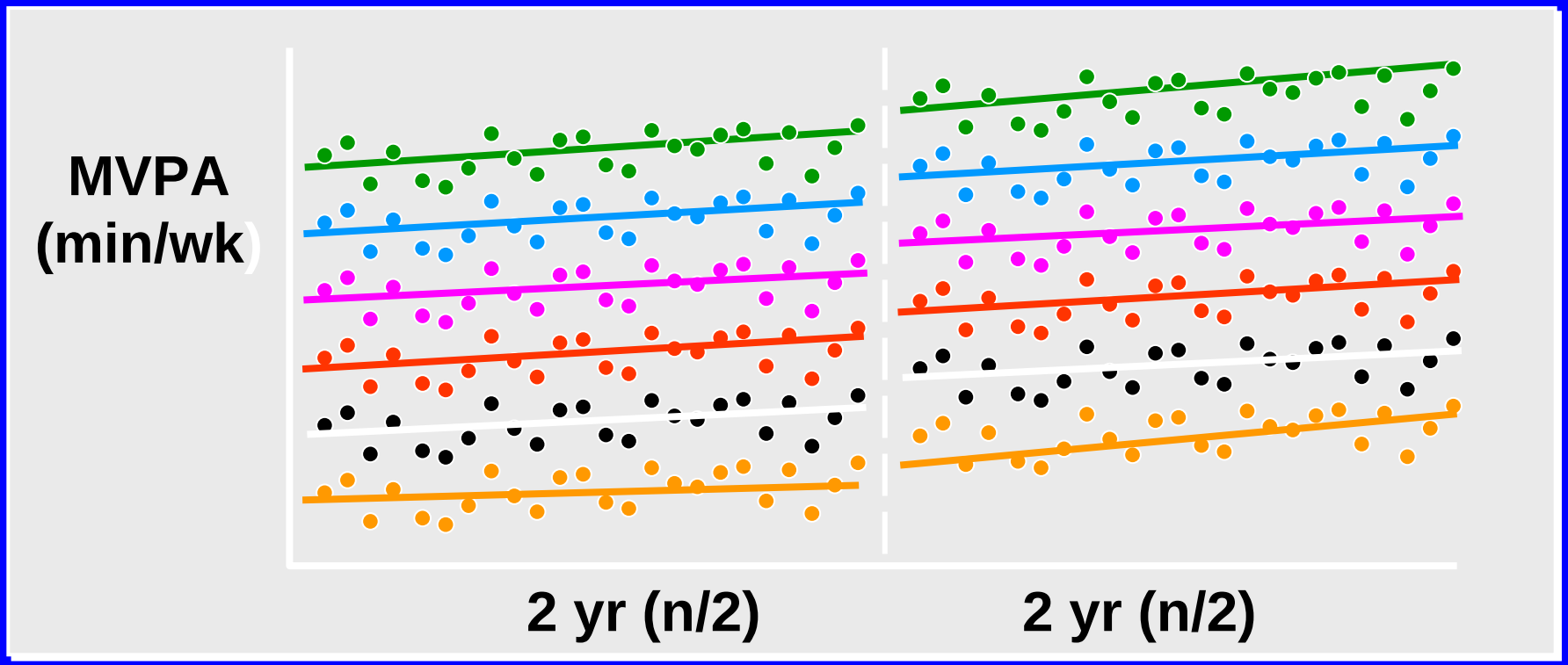
Adding other design features reduce threats to internal validity

- Control group
- Extra post interventions follow-up – persistance

Statistical Issues

- Henry Feldman

Fancier Analysis: Time series (6 v 0)



Community endpoint: *Trend* in mean MVPA

Analysis: ...

Techniques for Time-series

- **Regression (lines, curves, waves)**
- **Autoregressive moving average (ARIMA)**
- **Anything that takes advantage of order and serial correlation to characterize and compare *trends* and patterns**
- ***May* provide the most precise picture pre- and post-intervention, boost power to detect change**

Paired
Pre Post Designed

- Pre- Post- Q-E design
  - 3 intervention sites
  - 3 comparison sites
- Settings
  - Urban
  - Suburban

Design 2

- Time-series design
- Most aspects of design are similar to Design 1
  - 5-year study as in Design 1
- Phone interviews during first week of each month
- Six locations where light rail being built
  - Ea. site is its own comparison group prior to light rail

Detectable Effects: Bottom Line

Design	Analysis	Min/wk 6n=3000	Min/wk 6n=6000
Time series	One-sample t	12.8	9.1
Pre-post	Paired t, r = 0	39.4	27.8
“	“ , r = 0.2	35.2	24.9
“	“ , r = 0.5	27.8	19.7

Trade-offs: All Else Equal

Design feature	Guards against	Time series	Pre-post
Random assignment	Confounding	X	X
Repeated measures	Variability	+	+
Community pairing	“	X	?
Controls	Secular trend	X	+
Detectable effect	Futility	+	X
Analytic refinement	“	+	X

Research Designs – Time Series

Repeated interrupted time series without randomization can be convincing given large effects and long baselines [50 to 100 points] - even with these designs , randomization to times is still preferable, especially if long baselines is not possible –

SPR

Lottery Winner increasing ticket sales

Did we ask you if you want a Lottery ticket? If not, you get one free

Ask for the Sale

- Participation by stores was voluntary
- Control matched on chain, pretest level of tickets sales, zipcode
- Four pretest and four post test weeks of sales
- 3 nonequivalent dependent variables: sales of gas, cigarettes, and food
- Some stores removed and then restated the program

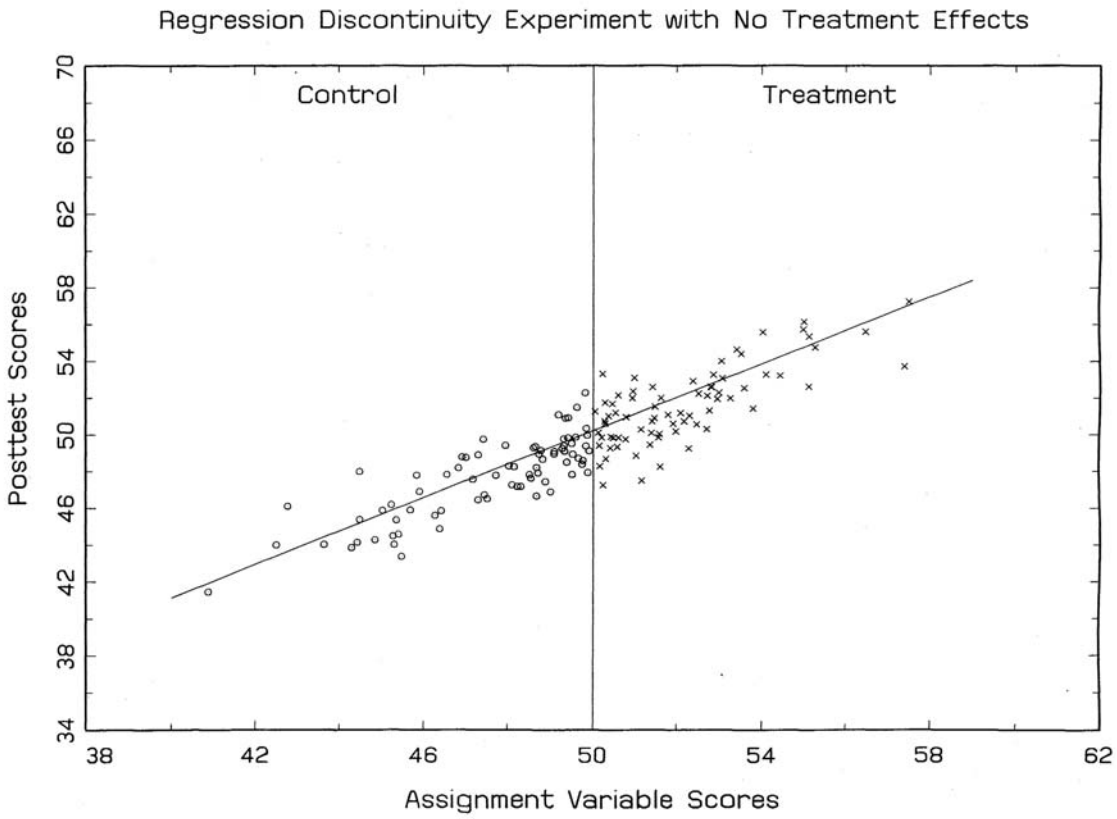
Research Designs – Regression discontinuity

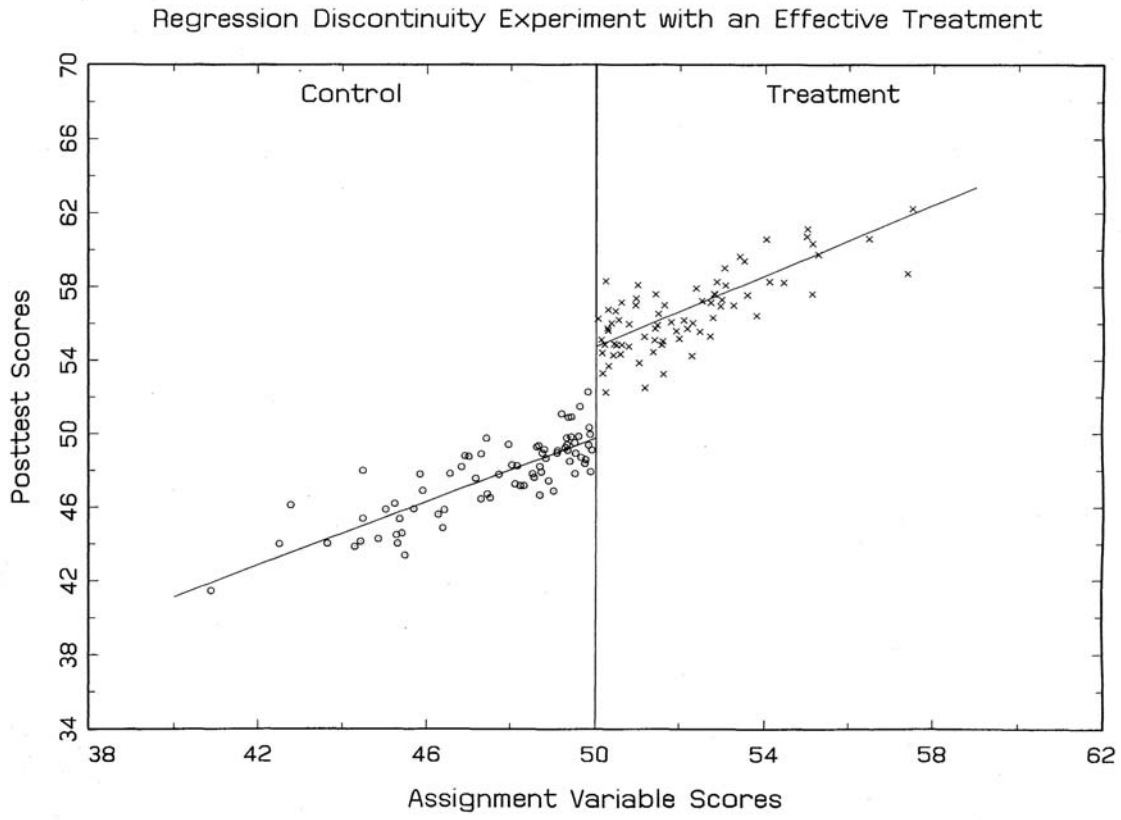
Regression Discontinuity Design

William Shadish
University of California, Merced

Regression Discontinuity Design

- Units are assigned to conditions based on a cutoff score on a measured covariate,
- For example, communities that exceed a certain cutoff on arrests for drunk driving for young drivers per 100,000 receive treatment, and communities below that cutoff are in the comparison condition.
- The effect is measured as the discontinuity between treatment and control regression lines at the cutoff (it is not the group mean difference).


Regression Discontinuity Experiment with No Treatment Effects
Control
Treatment
Posttest Scores
34
38
42
46
50
54
58
62
66
70
38
42
46
50
54
58
62
Assignment Variable Scores


Regression Discontinuity Experiment with an Effective Treatment
Control
Treatment
Posttest Scores
34
38
42
46
50
54
58
62
66
70
38
42
46
50
54
58
62
Assignment Variable Scores

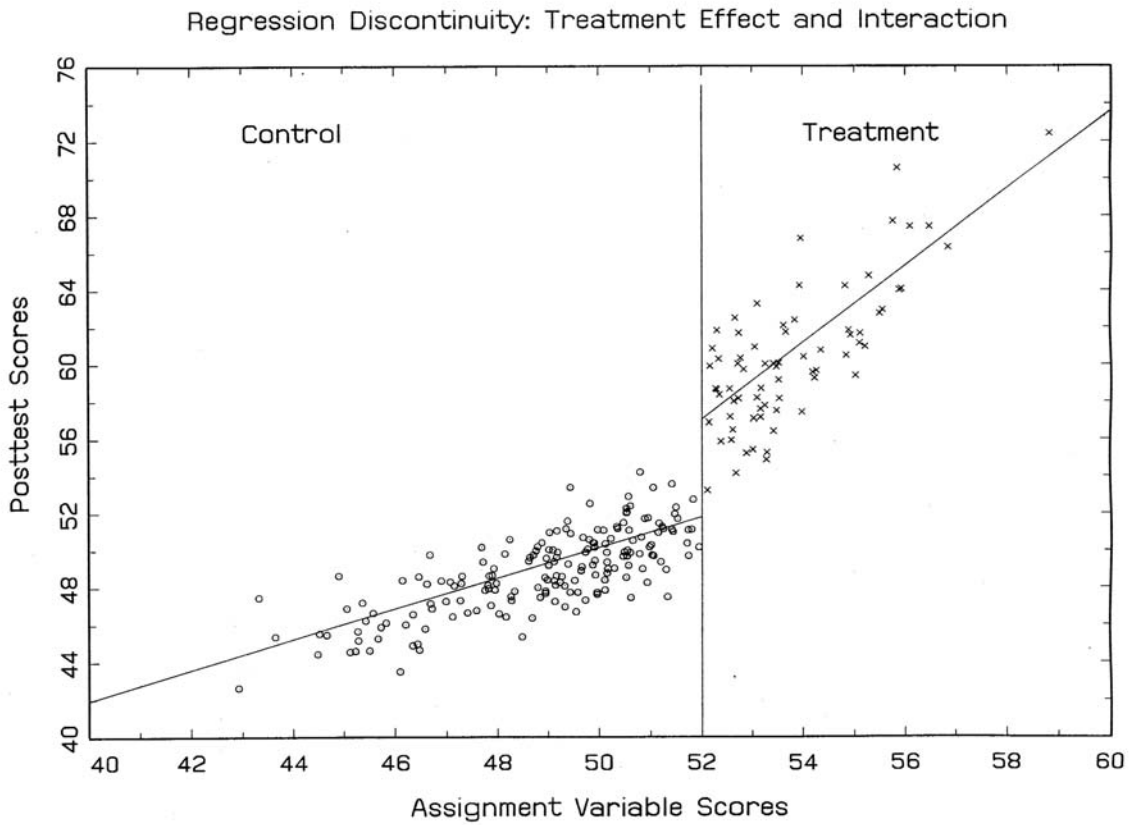
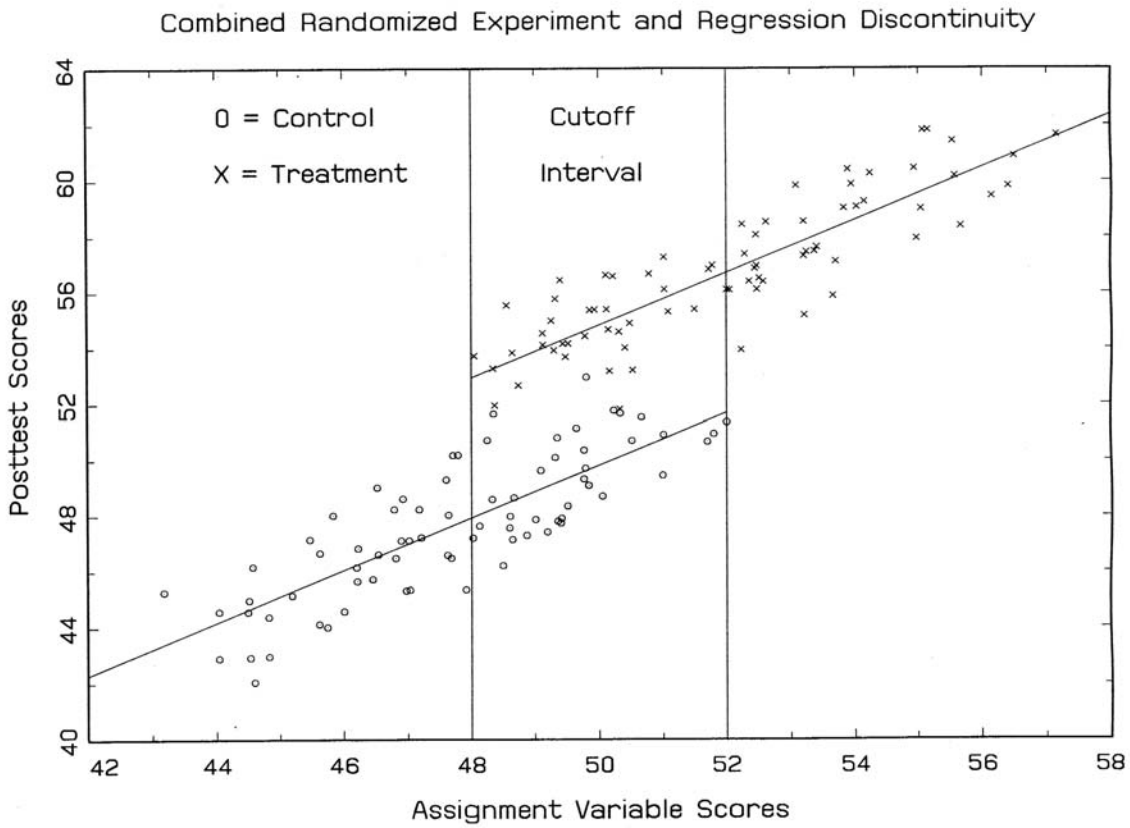

Regression Discontinuity: Treatment Effect and Interaction
Control
Treatment
Posttest Scores
Assignment Variable Scores
40
44
48
52
56
60
64
68
72
76
40
42
44
46
48
50
52
54
56
58
60

Fig 7-11


Combined Randomized Experiment and Regression Discontinuity
O = Control
X = Treatment
Cutoff
Interval
Posttest Scores
Assignment Variable Scores
40
44
48
52
56
60
64
42
44
46
48
50
52
54
56
58

7-13

Advantages

- When properly implemented and analyzed, RD yields an unbiased estimate of treatment effect (see Rubin, 1977).
- Communities are assigned to treatment based on their need for treatment, consistent with how many policies are implemented.

Disadvantages

- Statistical power is considerably less than a randomized experiment of the same size.
- Effects are unbiased only if the functional form of the relationship between the assignment variable and the outcome variable is correctly modeled, including:
  - Nonlinear Relationships
  - Interactions

Summary

- Of the QE designs being considered RD is the only one that yields an unbiased estimate.
- RD can be used with both archival data and original data.
- But there is question about whether it can be implemented with sufficient power in this case.

Closing comments

Trade offs

Causality

Your questions and comments

No Free Lunch -why randomization remains the gold standard for causality questions when feasible

- Sample size other factors constant, the sample size requirements are the same for GRTs, natural experiments, and quasi-experiments.

No Free Lunch Effectiveness and Translational causal studies – bad news and good news

- Intervention in the real world is likely to be less potent and more variably delivered
- Secular trend often a threat
- Maybe able to use more common intermediate outcomes

Judging Causality for Policy

Is not same thing as judging the internal validity of a single study– Bradford Hill Criteria and others – multiple studies, size of effect, plausibility – a judgment call

Clarifying Some Issues About the Choice of Design - Shaddish

- There are many good reasons not to do randomized designs, and not to do experiments at all.
  - When it is not possible (treatment already started, impossibility of getting control)
  - When it is unacceptable to a community?
  - When it is not necessary (the counterfactual is known, PKU screening example)
  - When we are not asking a causal question

Clarifying Some Issues About the Choice of Design - Shaddish

- But let’s not reject experiments for the wrong reasons

Are We Asking Causal Questions?

- The question should drive the method. Both efficacy and effectiveness are causal questions. But many translational research questions are not.
- Quasi-experiments are also helpful (RD, ITS, etc.)

Are We Asking Causal Questions?

- But if the question is not causal, the method should be different. For example
  - What is the need for treatment?
  - How do stakeholders value treatments and outcomes?
  - How are resources distributed?
  - How is treatment implemented?
  - Costs of treatment?

Are We Asking Non Causal Questions?

- But if the question is not causal, the method should be different. For example
  - What is the need for treatment?
  - How do stakeholders value treatments and outcomes?
  - How are resources distributed?
  - How is treatment implemented?
  - Costs of treatment?

Questionable Reasons to Reject Randomized Experiments

- Reasons that often apply to many nonrandomized experiments, as well:
  - The intervention is more than just the program, but includes staff, context etc.
  - We don’t know which component of treatment worked
- The constant treatment assumption: Randomized trials do not require constant treatments, especially not effectiveness trials. (Angrist et al. JASA)

Questionable Reasons, continued

- Randomized trials may or may not be more expensive, more time consuming, or less powerful than other studies, depending on what the other study design is. For example:
  - Can you do ITS with archival data or do you have to gather your own data anew? This will affect both time and expense.
- The cost of making an inferential error.
- But in general, more money buys you better data no matter what the design is.

Using Multiple Design Elements Single Component

- Quasi-experiments are not a set of fixed designs
- Rather, each quasi-experiment should be an aggregation of many design elements that each are thoughtfully chosen to compensate for the possible threats to causal inference
- Moreover, good quasi-experiments use multiple designs to address the same questions (e.g., Ask)

Summary of Agreements

- The question should drive the method
- Translational research involves more than just causal questions, and so requires many different kinds of design.
- Even when effects are at issues, randomized experiments are sometimes not possible or ethical, and nonrandomized experiments must be considered.
- In the hierarchy of nonrandomized designs, time series and regression discontinuity are generally preferred on theoretical grounds, when practical.

Summary of Agreements

- We have made a lot of progress on addressing problems in both randomized and nonrandomized experiments (things I've not heard mentioned here), for example:
  - Incomplete treatment implementation (Angrist)
  - Missing data
  - Selection bias and propensity scores
- Triangulating using different designs is a good thing (e.g., time series on archival data with surveys on other dv's; or RCT on proximal outcomes with observational study on long-term outcome).

Others are interested

- Nonexperimental versus experimental estimates of earning impacts: Steven Glazerman, Dan Levy, and David Myers
- Across studies comparisons

  73 metanalyses that Quasi Exp [NX] and RCT [X] methods mean effect size was .41 SD units in NX and .46 SD units in X , bulk differences between the NX and X for same type of intervention was close to zero [-.2 to +.4]

Reference

- The ANNALS of the American Academy of Political and Social Science 2003; 589: 63-93

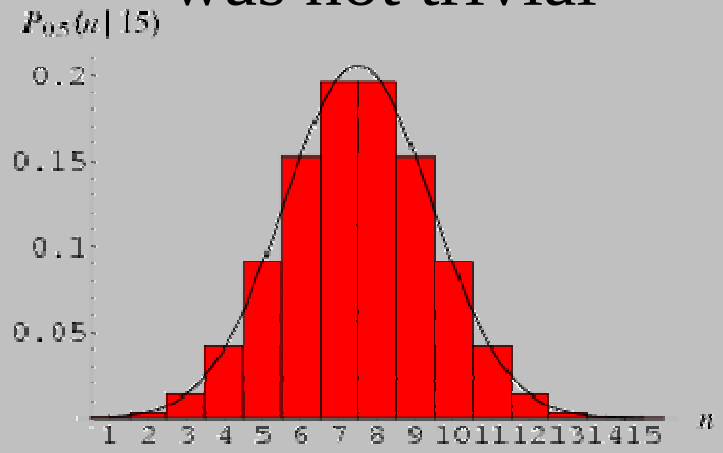
Conclusions For Labor Economics

Question: Can averaging multiple NX impact estimates approximate results from a RT?

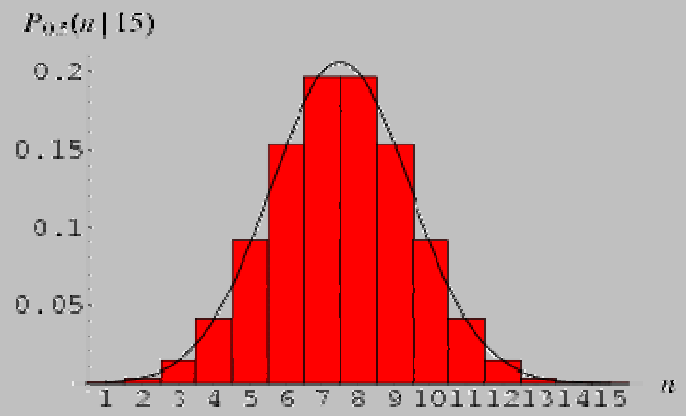
Answer: Maybe but we have not identified an aggregation strategy that consistently removes the bias or chance – estimated biases were both positive and negative ..centered roughly on zero

Overall mean of estimated intervention effects was identical but the intra-study difference in estimate effect was not trivial

Non-random controls



Random controls



Means are the same in gain in yearly income from training

Others are interested

- “Thus the between study evidence does not resolve whether differences in impact estimates are due to design or some other factor
- Within-study comparisons should give a ‘clean’ estimate of selection bias but cannot rule out random sampling error in both the experimental and NX estimators – design replication studies

Conclusions For Labor Economics

- “Thus the between study evidence does not resolve whether differences in impact estimates are due to design or some other factor

Other Social Scientists

Authors of the 12 design replication studies concluded:

4 that NX methods were well – no biases

4 that some NX methods were ok and some NX methods were biased

4 that NX methods did not perform well or evidence was insufficient

Stay tuned.

Extra Slides

Citations to Med/PH Examples

Cullen, K.W., Koehly, L.M., Anderson, C., Baranowski, T., Prokhorov, A., Basen-Engquist, K., Wetter, D., & Hergenroeder, A. (1999). Gender differences in chronic disease risk behaviors through the transition out of high school. American Journal of Preventive Medicine, 17, 1-7.

Finkelstein, M.O., Levin, B., & Robbins, H. (1996a). Clinical and prophylactic trials with assured new treatment for those at greater risk: I. A design proposal. American Journal of Public Health, 86, 691-695.

Finkelstein, M.O., Levin, B., & Robbins, H. (1996b). Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. American Journal of Public Health, 86, 696-705.

Summary

- Of the designs being considered for this intervention, RD is the only one that yields an unbiased estimate.
- RD can be used with both archival data and original data.
- But there is question about whether it can be implemented with sufficient power in this case.

Steve Gortmaker

- “I think we need larger effect sizes.”

Suppose

- You focused on a series of three churches and work with only one until you got truly big changes in exercise and weight/
- You worked in a series of three middle schools and work to produce big changes in nutrition and exercise practices
  - Refining the intervention in each successive school.

Suppose further,

- That we assume that we are trying to permanently alter the infrastructure of practices that affect public health.
  - Science as engineering the public healh infrastructure.

The Use of Interrupted Time-Series Designs in Behavior Analysis

- Behavior Analysis has focused on identifying independent variables that have big effects on dependent variables.
- An effect of 5 standard deviations is at the 25th percentile.
- The median effect size is 10 standard deviations.

Critique of COMMIT

- Some of the components included in COMMIT had not been shown to affect the processes they targeted.
- The example of Physician’s Advice.

Scenario 3: Promotion of Physical Activity

Ross C. Brownson, PhD

David Buchner, MD, MPH

Deborah A. Cohen, MD, MPH

Henry A. Feldman, PhD

Denise Simons-Morton, MD, PhD

Designs of Quasi-Experiments Studies for Assessing the Transport Enhancements and Physical Activity

What designs might be used to study the relationship between transportation enhancements and physical activity?

Why a Q-E Design?

- Random assignment is not possible
  - Lack of control over the “intervention”
- Not enough units of analysis

Research Questions

- *Primary aim:*

  1. What is the relationship between a transportation enhancement, specifically construction of light rail lines, and rates of moderate physical activity?

- *Secondary aims:*

  1. Do such physical changes differentially affect populations by economic status, race/ethnicity, age group, and gender?

  2. What is the cost-effectiveness of these physical changes?

Design 1

- Pre- Post- Q-E design
  - 3 intervention sites
  - 3 comparison sites
- Settings
  - Urban
  - Suburban

Complex Intervention

- As described by Deborah Cohen

Study Sites

- 6 sites
- Matched intervention, comparison sites
  - 3 intervention
  - 3 comparison
  - Comparison sites geographically separated from intervention sites

Comparison Sites

- No planned light rail
- Potential matching factors
  - size of the community
  - population density
  - racial/ethnic mix
  - proportion of the population under the poverty level
  - geographic location/weather

Sampling Study Participants

- Telephone survey
  - Cross-sectional repeated measures
    - Trade-offs with cohort sample (next slide)
  - Perceived environment & self-reported physical activity
- Study subjects in neighborhoods served by light rail
- Phone survey participants recruited to wear motion sensors
- Baseline measures and 2 follow-up samples
  - *e.g.*, 0, 12 months, 24 months

Survey Design

- Cross-sectional vs. cohort design
  - Avoid attrition in cross-sectional sample
  - Can assess reasons for moving
    - Important in relation to internal validity (self-selection)
  - Statistical power generally lower with cross-sectional sample

Objective Assessment of the Environment

- Community audits
  - Physical factors
    - Land use destinations
    - Recreational facilities
  - Social factors
    - Social disorder
    - Children at play

Key Variables

- Dependent variable
  - Physical activity behavior
    - Self-report, multiple domains (*e.g.*, IPAQ)
    - Motion sensors
- Independent variable
  - Physical environment (light rail)
- Moderators/mediators
  - Distance from transit
  - SES
  - Race/ethnicity

Design 2

- Time-series design
- Most aspects of design are similar to Design 1
  - 5-year study as in Design 1
- Phone interviews during first week of each month
- Six locations where light rail being built
  - Ea. site is its own comparison group prior to light rail

Research Designs – RCT's

Improving External Validity and Responding to Concerns

Scenario 1: Prevention of Type 2 Diabetes

Marshall Chin, MD

Sanford A. Garfield, PhD

Carol Mangione, MD, MSPH

K.M. Venkat Narayan, MD, MPH, MBA

Myrlene A. Staten, MD

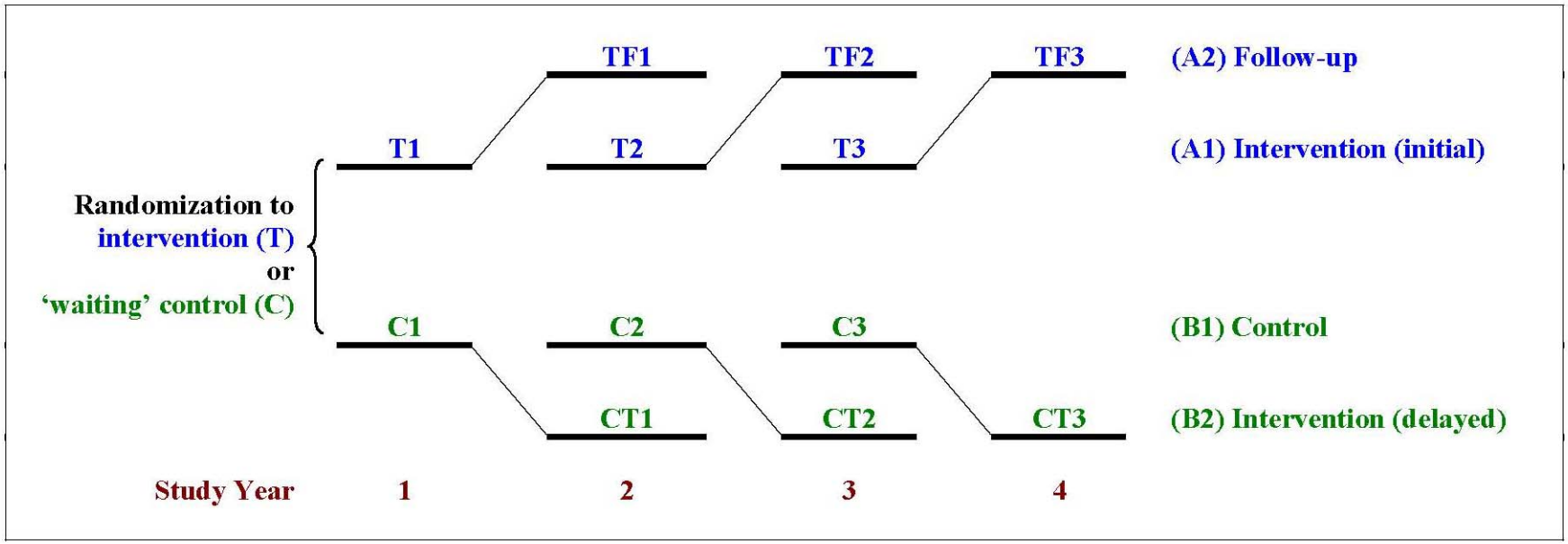
Staggered Enrollment

Marshall H. Chin, MD, MPH
Associate Professor of Medicine
University of Chicago

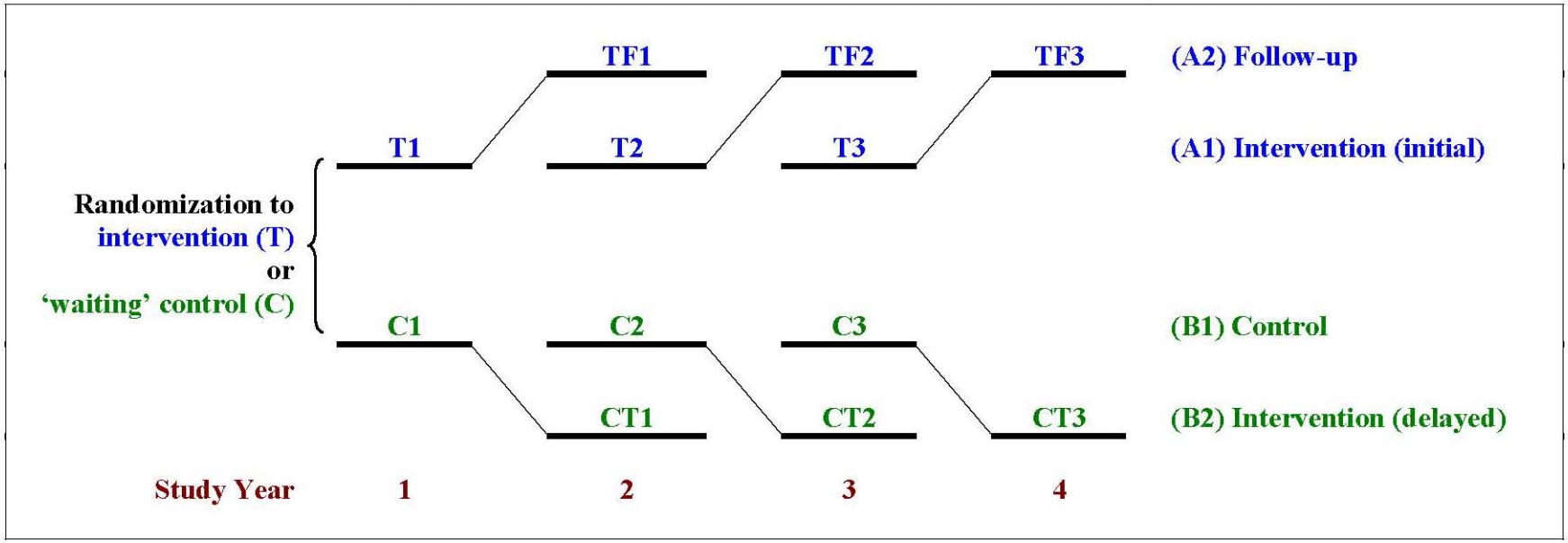
Design 2: Quasi-Experimental with Staggered Enrollment

- Randomization of initial assignment into intervention or control arm
- After 1 year, control participants transfer into intervention arm

Design: Staggered Enrollment with `Waiting' Control



Design: Staggered Enrollment with `Waiting' Control



Potential Analyses

Analysis	Contrasts	Cohort x Year	Comments
1. Intervention vs Control	(A1) vs (B1)	T1, T2, T3 vs C1, C2, C3	Standard RCT (1 yr outcome)
2. Follow-up vs Control	(A2) vs (B2)	TF1, TF2, TF3 vs C1, C2, C3	Addresses sustainability
3. Pre-post Comparison	(B1) vs (B2)	CT1, CT2, CT3 vs C1, C2, C3	Within subject comparison (uncontrolled)
4. Treatment (initial, delayed) vs Control	(A1), (B2) vs (B1)	T1, T2, T3 CT1,CT2, CT3 vs C1, C2, C3	Paritially pairwise comparisons Analytic issues: ▪ correlation ▪ appropriate contrasts
5. Analysis of Variance (4-group)	(A1), (A2), (B2) vs (B1)	(T1, T2, T3) (TF1, TF2, TF3) (CT1,CT2, CT3) vs C1, C2, C3	

Potential Analyses: Staggered Enrollment/Waiting Control Design

Analysis	Contrasts	Cohort x Year	Comments
		Intervention vs Control	
1. Standard RCT (1 yr)	(A1) vs (B1)	T1, T2, T3 vs C1, C2, C3	Intervention vs Control
2. Follow-up vs Control	(A2) vs (B2)	TF1, TF2, TF3 vs C1, C2, C3	Intervention Follow-up vs Control Addresses sustainability
3. Pre-post Comparison: Waiting Controls	(B1) vs (B2)	CT1, CT2, CT3 vs C1, C2, C3	Within subject comparison (uncontrolled)
4. Treatment (initial & delayed) vs Control	(A1), (B2) vs (B1)	T1, T2, T3 CT1,CT2, CT3 vs C1, C2, C3	Paritially pairwise comparisons Analytic issues: ▪ correlation ▪ appropriate contrasts
5. Analysis of Variance (4-group)	(A1), (A2), (B2) vs (B1)	(T1, T2, T3) (TF1, TF2, TF3) vs C1, C2, C3 (CT1,CT2, CT3)	

Analyses

- Intervention vs. control
- Among control subjects that crossover into intervention, each subject can serve as own control

Weaknesses

- Secular trends
- Possible contamination of control groups
- Learning effects – control group has more time in study
- Shorter F/U time in initial control group
- Analytic challenges in adjusting for correlation

Back-up slides

RET Sample Size Considerations

RET:

- P_1 Adoption rate in the intervention group
- P_0 Adoption rate in the control group
- RET the incremental adoption rate is: $P_d = P_1 - P_0$

RCT

- Q_1 Adoption rate in the intervention group
- Q_0 Adoption rate in the control group
- RCT with perfect adherence adoption rate is: $Q_d = Q_1$

Assume treatment effect is constant M, then RET intervention effects are P_dX M and RCT effects are Q_dX M and the inflation factor is: $(Q_d / P_d)^2$

Society of Prevention Research Committee Report

- Committee: Standards of Evidence
- Committee Chair: Brian R. Flay
- Report date March 8th, 2004

Research Designs – Group RCT's

The assignment condition should maximize confidence that the intervention rather than some other condition caused the outcome – [sufficient sample size without significant pre-intervention differences – possible in many if not most with very substantial resources – **SPR**

RET Strengths and Weaknesses

Internal validity for treatment adoption	RET>>OBS>>RCT
Internal validity for effectiveness	RCT>>RET>>OBS
Internal validity for public health benefit	RET>>OBS>>RCT
External Validity: Sample representativeness	OBS>RET>>RCT
External Validity: Content representativeness	OBS>RET>>RCT
Sample Size and Costs	OBS>>RCT>>RET

> Moderate dominance, >> strong dominance

Duan N., et al. Personal Communication

Strengths

- Increased enrollment compared with std RCT
- Increased subject retention compared with std RCT
- Intervention and control subjects drawn from same population
- Subjects initially randomized to control group can serve as own controls

RET Strengths and Weaknesses

Internal validity for treatment adoption	RET>>OBS>>RCT
Internal validity for effectiveness	RCT>>RET>>OBS
Internal validity for public health benefit	RET>>OBS>>RCT
External Validity: Sample representativeness	OBS>RET>>RCT
External Validity: Content representativeness	OBS>RET>>RCT
Sample Size and Costs	OBS>>RCT>>RET

> Moderate dominance, >> strong dominance

Duan N., et al. Personal Communication

RET Sample Size Considerations

- DPP, assume that the treatment effect is prevention of **6.2 cases** of DM per 100 person-years
- Then in the RET, if adoption P_d=0.5, then RET treatment effect is **3.1 cases** of DM per 100 person years
- **Then inflation factor is $(1/0.5)^2$ = 4 times more sample needed for the same power!**

Sample size needed based on the observed effect size in the DPP

Study	Placebo	Lifestyle	Alpha	Power	Sample Size
DPP	11 cases per 100 person-yr	4.8 cases per 100 person-yr	.05 one tailed	.90	**353 per arm**
RET	11 cases per 100 person-yr	7.9 cases per 100 person-yr	.05 one tailed	90	**1582 per arm**

Staggered Enrollment Analyses

- Intervention vs. control
- Among control subjects that crossover into intervention, each subject can serve as own control

Staggered Enrollment Strengths

- Increased enrollment compared with std RCT
- Increased subject retention compared with std RCT
- Intervention and control subjects drawn from same population
- Subjects initially randomized to control group can serve as own controls

Time-series v. Pre-post Design: Statistical Issues

- **Analytic strategy (crude, refined)**
- **Correlation**
  - **Within community (ICC), pair (r)**
- **Detectable effects (crude)**
  - **Common elements, key differences**
  - **Analytic options**
- **Trade-offs**
  - **Validity, generalizability, power**

Detectable Effect Calculation: Common Elements

- **80% power**
- **Crude analysis**
- **H_0 tested with critical p-value 0.05**
- **Standard deviation 101 min/wk MVPA**
- **Community clustering of MVPA:**

  ICC (intraclass correlation)

  $$= \frac{\text{Community variance}}{\text{Community + person variance}} = 0.009$$

- **Correlation between community pairs: r = ?**

Detectable Effects: Do the Math

Detectable [diff in] Mean Δ	Pre-post *paired*	Pre-post *unpaired*	Time series
= SD (min/wk)	101	101	101
× $(t_{\alpha/2,df} + t_{\beta,df})$	2 df	4 df	5 df
× [4 (1–ICC)	0.009	0.009	0.009
× groups compared	2	2	1
÷ communities/group	3	3	6
÷ subjects/community	n	n	n
× (1 – r)]^{½}	?	0	0

Detectable Effects: Bottom Line

Design	Analysis	Min/wk 6n=3000	Min/wk 6n=6000
Time series	One-sample t	12.8	9.1
Pre-post	Indep-sample t	27.3	19.3
Pre-post	Paired t, r = 0	39.4	27.8
“	“ , r = 0.2	35.2	24.9
“	“ , r = 0.5	27.8	19.7

Time-series v. Pre-post Design: Statistical Issues

- **Analytic strategy (crude, refined)**
- **Correlation**
  - **Within community (ICC), pair (r)**
- **Detectable effects (crude)**
  - **Common elements, key differences**
  - **Analytic options**
- **Trade-offs**
  - **Validity, generalizability, power**

Translational Questions may be focused at different levels

- **Public Policy**
- **Health systems and plans**
- **Clinical practice**
- **Patient**

Sources: Berwick; Shortell

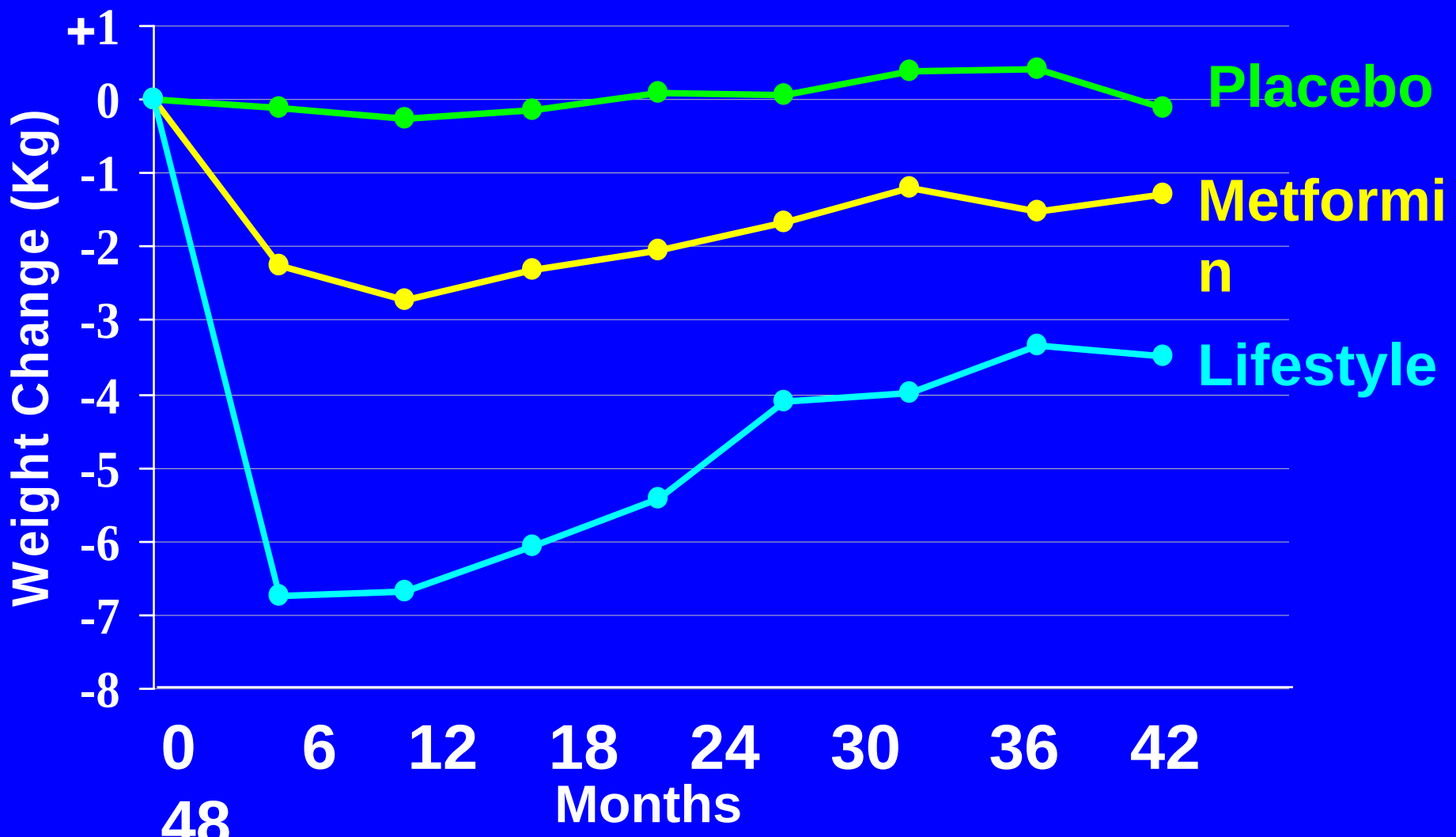
Workshop method

- Six expert teams created scenarios with two research designs
  - Prevention type 2 diabetes
  - Prevention of childhood obesity
  - Improving management of asthma
  - Promoting physical activity
  - Prevention of underage drinking
  - Prevention of youth cigarette use
- Discussion of trade offs of each design

Questions in the background

- Since sometimes randomization is not feasible – i.e. policy changes, unethical [everyone must be treated], lack of trust in a community, what can be done in this situation?
- Randomization is possible at the group level but expensive
- What are the appropriate review process/criteria ?

Mean Weight Change

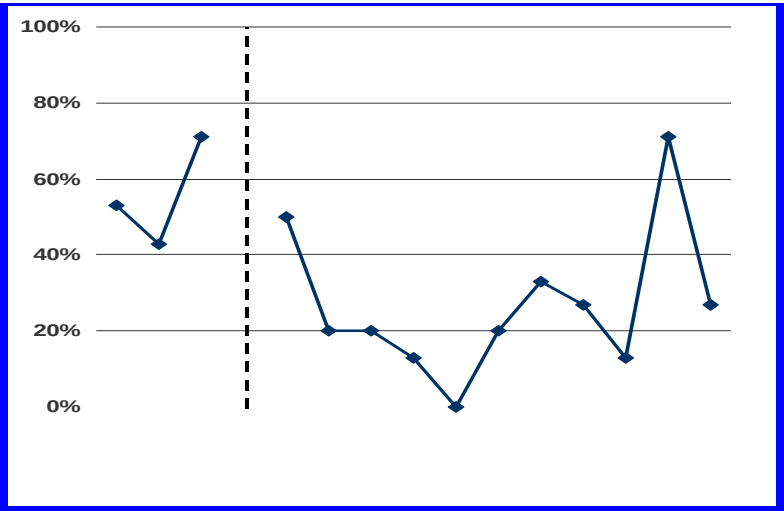
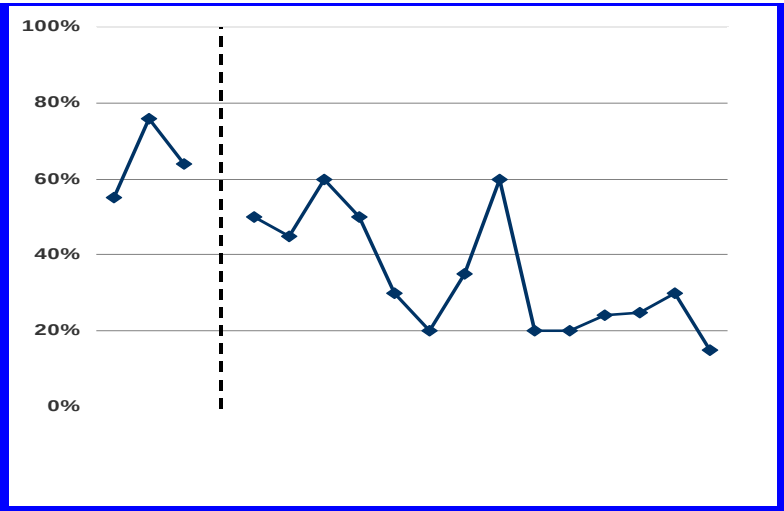
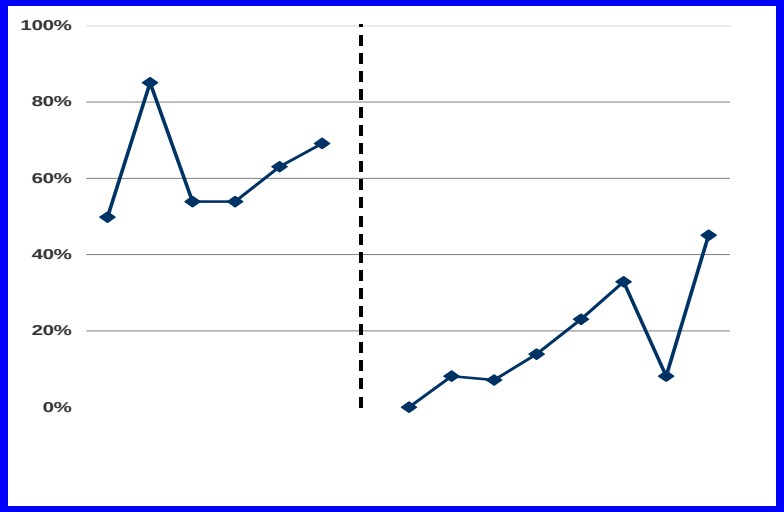
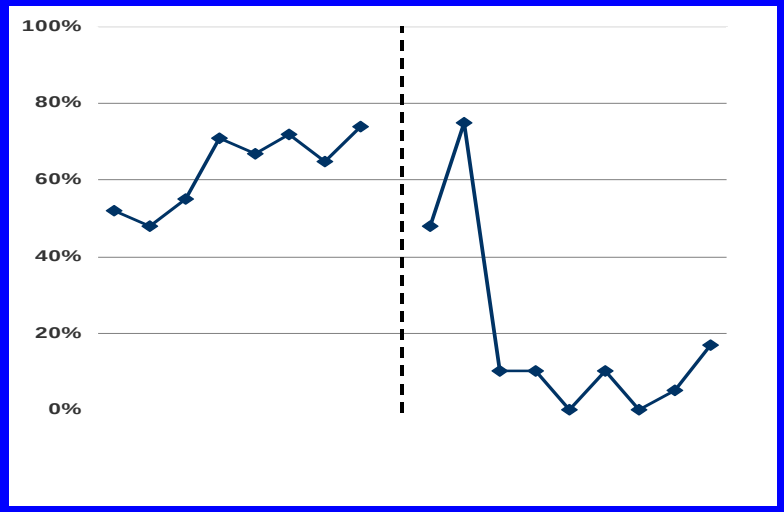


Translational research often asks non causal but important Q's

- RE-AIM www.re-aim.org
- Reach the target population
- Efficacy or effectiveness

1. Adoption by target settings or institutions
2. Implementation consistency of delivery of intervention
3. Maintenance of intervention effects
4. IMPACT = RxExAxIxM

Workshop Report

Insights on improving external validity of RCT's and internal validity of research studies without random assignment

: Access Reduction in Four Communities