Multisite Causal Mediation Analysis

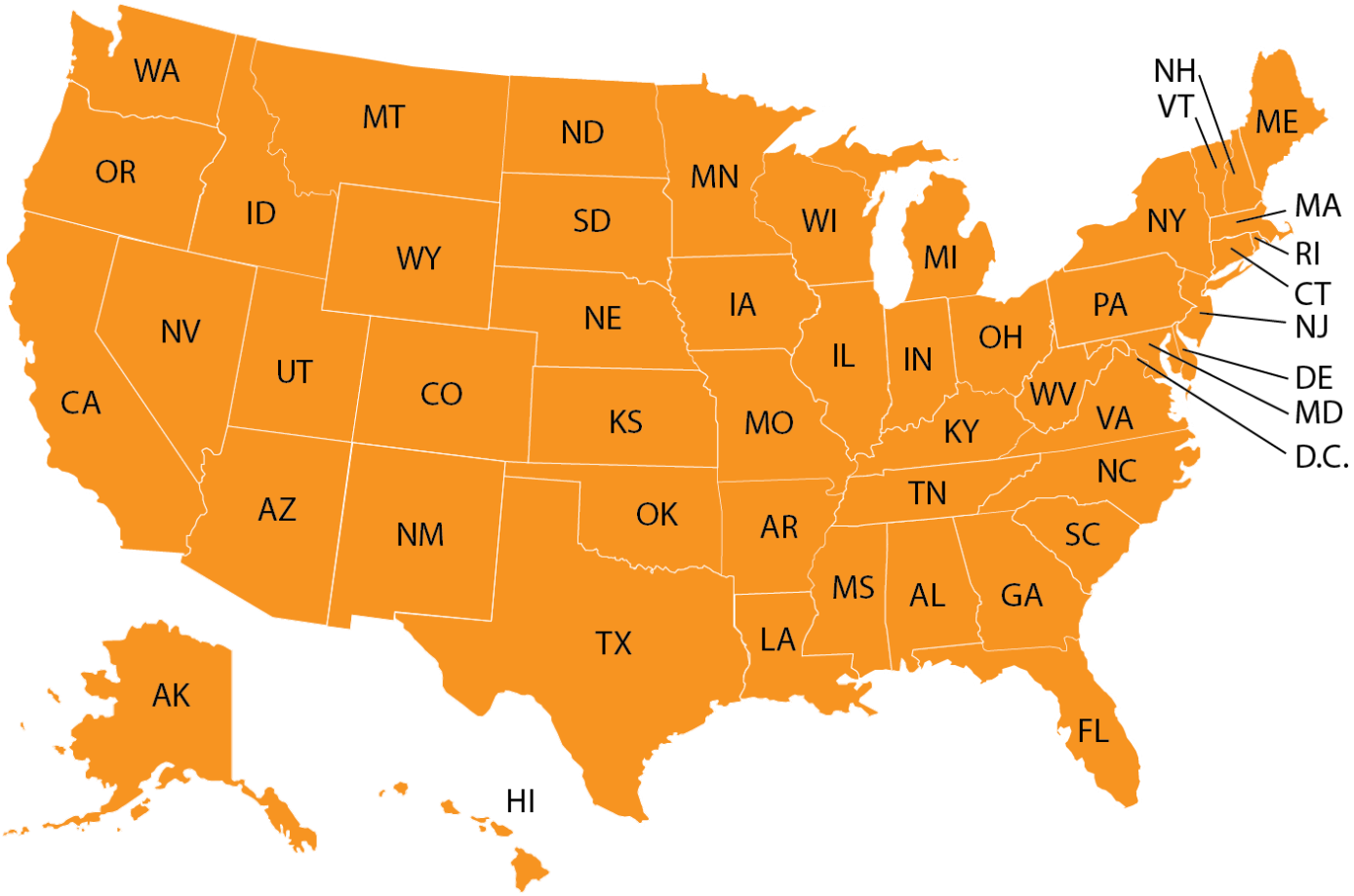
Xu Qin

Guanglei Hong

University of Chicago

Outline

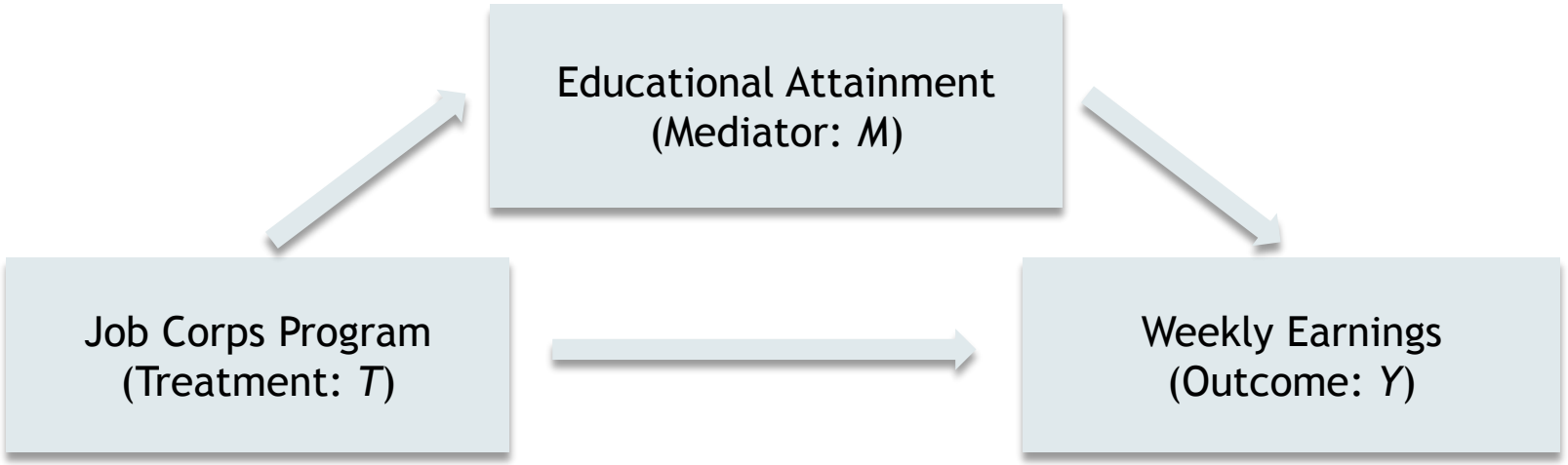
- National Job Corps Study
- Research Questions
- Methodological Challenges
- Identification
- Estimation
- Strengths and limitations


WA
OR
CA
ID
NV
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
NH
VT
ME
MA
RI
CT
NJ
DE
MD
D.C.
AK
HI

National Job Corps Study

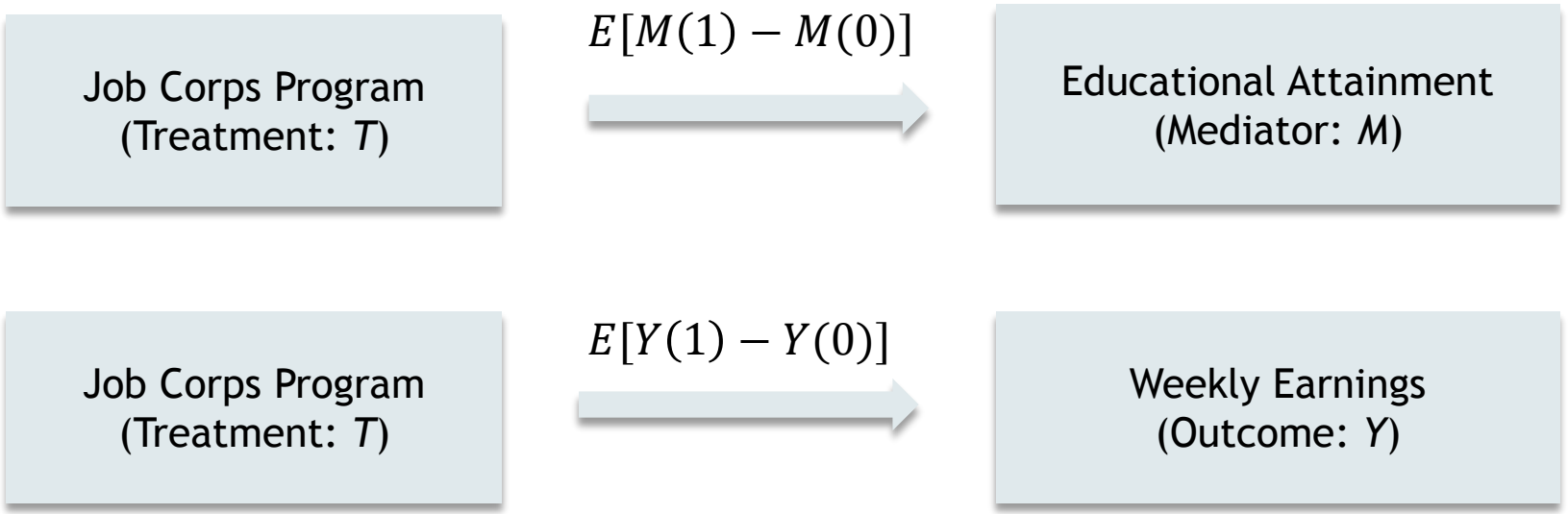
- **Job Corps**: the largest training to promote economic independence among disadvantaged youths ages 16-24 who are disconnected from school and work.
- **National Job Corps Study**: a multi-site randomized evaluation of Job Corps programs nationwide
- **Sample Size**: 9,409 applicants were randomly assigned to the experimental group and 5,977 to the control group. A total of 103 Job Corps centers served as the experimental sites. The number of applicants at each site ranged from 34 to 656.
- **Previous Findings**: on average, attaining an education or vocational training credential through the program mediated the Job Corps impacts on employment and earnings.

Hypothesized Mediation Mechanism

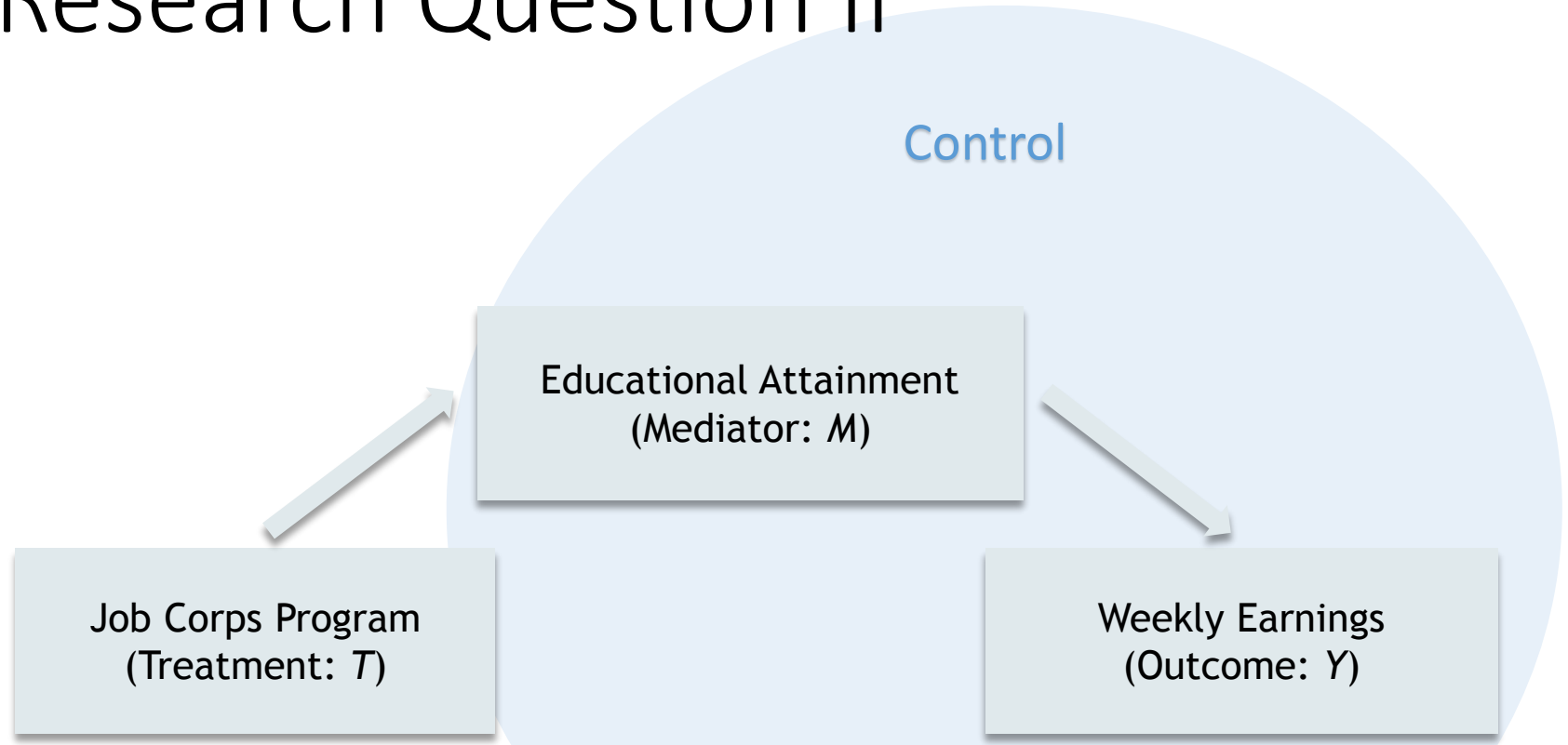


- **Mediator**: whether obtained an education/training credential during 30 months
- **Outcome**: weekly earnings in the 4th year after randomization
- Supplementary services: residential living, social skill training, individual counseling, drug and alcohol treatment, and health care

Research Question I

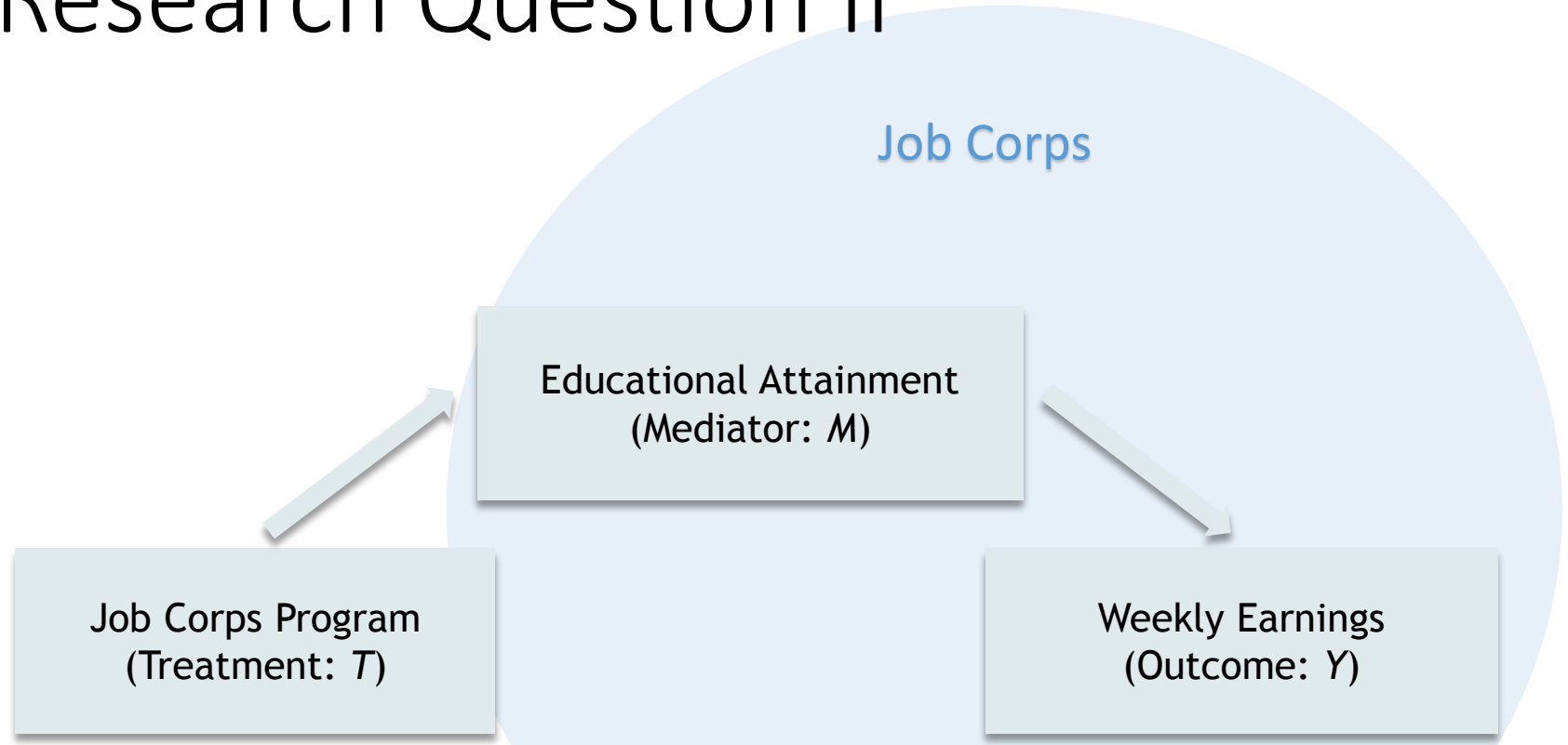


Research Question II



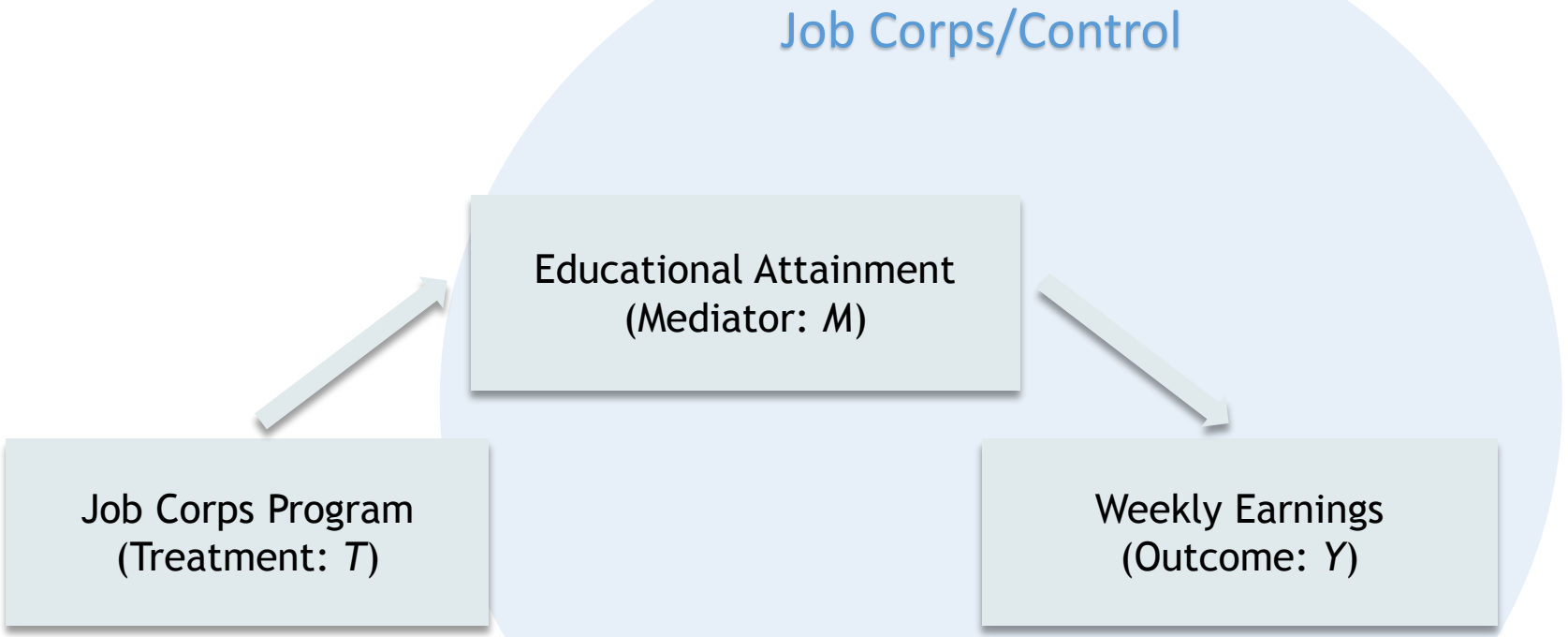
- How would the program-induced change in educational attainment exert an impact on earnings under the control condition (pure indirect effect) on average?
- How does it vary across sites and how is it correlated with site-specific program impact on educational attainment?

Research Question II



- How would the program-induced change in educational attainment exert an impact on earnings under JC condition (natural indirect effect) on average?
- How does it vary across sites and how is it correlated with site-specific program impact on educational attainment?

Research Question II



- The effect of program-induced change in education and training could be enhanced by the other supports provided by Job Corps that were generally unavailable under the control condition.
- If the above two are different, the extent to which this discrepancy (natural treatment-by-mediator interaction effect) was consistent across the sites.

Definition

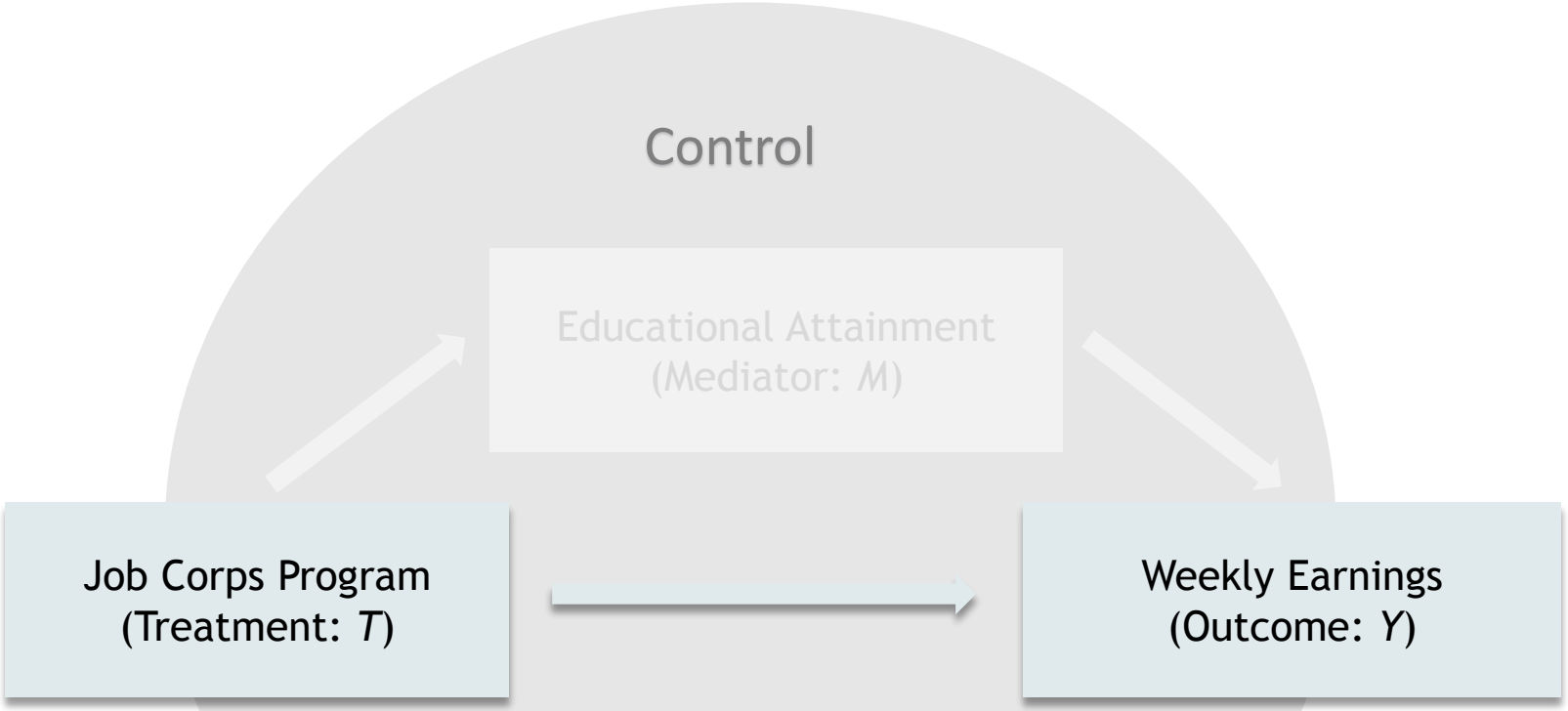
	Job Corps $T = 1$	Control $T = 0$
Educational attainment if assigned to Job Corps $M(1)$	$E[Y(1, M(1))]$	$E[Y(0, M(1))]$
Educational attainment if assigned to control $M(0)$	$E[Y(1, M(0))]$	$E[Y(0, M(0))]$

Natural Indirect Effect

Pure Indirect Effect

Natural Treatment-by-Mediator Effect

Research Question III



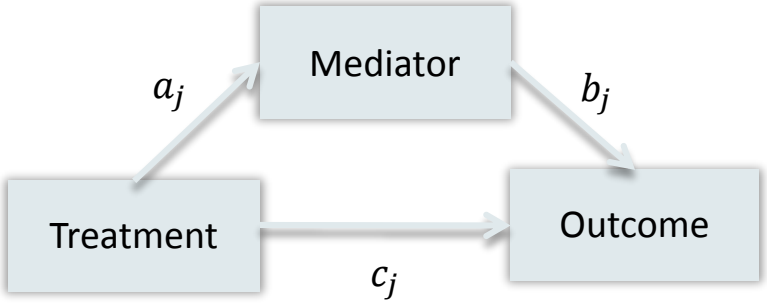
- How does the natural direct effect, which reflects the amount of variation across the sites in the relative contribution to the program impact of unspecified pathways vary across sites?
- How is it correlated with site-specific natural indirect effect?

Definition

	Job Corps $T = 1$	Control $T = 0$
Educational attainment if assigned to Job Corps $M(1)$	$E[Y(1, M(1))]$	$E[Y(0, M(1))]$
Educational attainment if assigned to control $M(0)$	$E[Y(1, M(0))]$	$E[Y(0, M(0))]$

Natural Direct Effect

Path Analysis Models For Multi-site Data



$$M_{ij} = (d_M + u_{Mj}) + (a + u_{aj})T_{ij} + e_{Mij}$$

$$Y_{ij} = (d_Y + u_{Yj}) + (c + u_{cj})T_{ij} + (b + u_{bj})M_{ij} + e_{Yij}$$

$$\begin{pmatrix} e_{Mij} \\ e_{Yij} \\ u_{Mj} \\ u_{a_j} \\ u_{Y_j} \\ u_{c_j} \\ u_{b_j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{e_M} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{e_Y} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tau^2_M & \tau_{M,a} & \tau_{M,Y} & \tau_{M,c} & \tau_{M,b} \\ 0 & 0 & \tau_{M,a} & \tau^2_a & \tau_{a,Y} & \tau_{a,c} & \tau_{a,b} \\ 0 & 0 & \tau_{M,Y} & \tau_{a,Y} & \tau^2_Y & \tau_{Y,c} & \tau_{Y,b} \\ 0 & 0 & \tau_{M,c} & \tau_{a,c} & \tau_{Y,c} & \tau^2_c & \tau_{c,b} \\ 0 & 0 & \tau_{M,b} & \tau_{a,b} & \tau_{Y,b} & \tau_{c,b} & \tau^2_b \end{pmatrix} \right]$$

- Population Average Indirect Effect: $ab + \tau_{a,b}$
- Between-Site Variance of Indirect Effect: $b^2\tau_a^2 + a^2\tau_b^2 + \tau_a^2\tau_b^2 + 2ab\tau_{a,b} + \tau_{a,b}^2$

Challenges to Path Analysis

- Rely on correct specifications of both the mediator model and the outcome model including
  - the functional forms
  - the distributional forms
- Because the average indirect effect is represented as a combination of multiple parameters, it adds considerable complications to estimation and statistical inference especially when research interests lie in
  - the between-site variance of the indirect effect
  - the covariance between the site-specific direct effect and indirect effect
- Their applications to discrete mediators and outcomes are particularly limited in both single-site and multisite studies

Counterfactual Framework

	Job Corps $T_{ij} = 1$	Control $T_{ij} = 0$
Educational attainment if assigned to Job Corps $M_{ij}(1)$	$E[Y_{ij}(1, M_{ij}(1))]$	$E[Y_{ij}(0, M_{ij}(1))]$
Educational attainment if assigned to control $M_{ij}(0)$	$E[Y_{ij}(1, M_{ij}(0))]$	$E[Y_{ij}(0, M_{ij}(0))]$

Under SUTVA:

No interference between units within a site

No interference between sites

Identification Assumptions

- Sequential randomization within levels of pretreatment covariates at each site:
  - Within levels of the observed pretreatment covariates, the treatment assignment was randomized at each site.
  - Within levels of pretreatment covariates, the mediator value assignment was also randomized under each treatment at each site.

Identification Results

	Job Corps $T_{ij} = 1$	Control $T_{ij} = 0$
Educational attainment if assigned to Job Corps $M_{ij}(1)$	$E[Y_{ij}(1, M_{ij}(1))\|S_{ij} = j]$ $\|\|$ $E[Y_{ij}\|T_{ij} = 1, S_{ij} = j]$	$E[Y_{ij}(0, M_{ij}(1))\|S_{ij} = j]$ $\|\|$ $E[W'_{ij}Y_{ij}\|T_{ij} = 0, S_{ij} = j]$
Educational attainment if assigned to control $M_{ij}(0)$	$E[Y_{ij}(1, M_{ij}(0))\|S_{ij} = j]$ $\|\|$ $E[W_{ij}Y_{ij}\|T_{ij} = 1, S_{ij} = j]$	$E[Y_{ij}(0, M_{ij}(0))\|S_{ij} = j]$ $\|\|$ $E[Y_{ij}\|T_{ij} = 0, S_{ij} = j]$

$$W_{ij} = \frac{\Pr(M_{ij} = m|T_{ij} = 0, \boldsymbol{X}_{ij} = \boldsymbol{x}, S_{ij} = j)}{\Pr(M_{ij} = m|T_{ij} = 1, \boldsymbol{X}_{ij} = \boldsymbol{x}, S_{ij} = j)}$$

$$W'_{ij} = \frac{\Pr(M_{ij} = m|T_{ij} = 1, \boldsymbol{X}_{ij} = \boldsymbol{x}, S_{ij} = j)}{\Pr(M_{ij} = m|T_{ij} = 0, \boldsymbol{X}_{ij} = \boldsymbol{x}, S_{ij} = j)}$$

Analytic Procedure

- Step 1 Propensity score estimation
- Step 2 Site-specific effect estimation
- Step 3 Population average effect estimation
- Step 4 Between-site variance and covariance estimation
- Step 5 Sensitivity analysis

Step 1 Propensity score estimation

Let $p_{tij}^{(t)} = Pr(M_{ij} = 1 | T_{ij} = t, \boldsymbol{X}_{ij} = \boldsymbol{x}, S_{ij} = j)$,

$$logit\left(\frac{p_{tij}^{(t)}}{1 - p_{tij}^{(t)}}\right) = \boldsymbol{X}'_{tij}\boldsymbol{\alpha}_t + \boldsymbol{S}'_{tij}\boldsymbol{r}_{tj}, \boldsymbol{r}_{tj} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_t), t = 0,1$$

- $\boldsymbol{X}$: age, gender, race, criminal involvement, drug use, employment, and earnings at the baseline. $\boldsymbol{S}$: vector of covariates with random effects $\boldsymbol{r}_j$.
- Balance checking: assess whether the distributions of $\boldsymbol{X}$ are similar across combinations of the treatment and mediator levels at all the sites after controlling for the pretreatment covariates.

Step 2 Site-specific effect estimation

	Job Corps $T_{ij} = 1$	Control $T_{ij} = 0$
Educational attainment if assigned to Job Corps $M_{ij}(1)$	$E[Y_{ij}\|T_{ij} = 1, S_{ij} = j]$ $\uparrow$ $\frac{\sum_{i=1}^{n_j} T_{ij}Y_{ij}}{\sum_{i=1}^{n_j} T_{ij}}$	$E[W'_{ij}Y_{ij}\|T_{ij} = 0, S_{ij} = j]$ $\uparrow$ $\frac{\sum_{i=1}^{n_j} \widehat{W}'_{ij}(1 - T_{ij})Y_{ij}}{\sum_{i=1}^{n_j} \widehat{W}'_{ij}(1 - T_{ij})}$
Educational attainment if assigned to control $M_{ij}(0)$	$E[W_{ij}Y_{ij}\|T_{ij} = 1, S_{ij} = j]$ $\uparrow$ $\frac{\sum_{i=1}^{n_j} \widehat{W}_{ij}T_{ij}Y_{ij}}{\sum_{i=1}^{n_j} \widehat{W}_{ij}T_{ij}}$	$E[Y_{ij}\|T_{ij} = 0, S_{ij} = j]$ $\uparrow$ $\frac{\sum_{i=1}^{n_j} (1 - T_{ij})Y_{ij}}{\sum_{i=1}^{n_j} (1 - T_{ij})}$

Challenge – Two-Step Estimation

- Sampling variability in weight cannot be ignored
    - The sampling variability in weight, estimated through multilevel logistic regressions in step 1, needs to be reflected in the sampling variability of site-specific effect estimates.
- Solution: treat the two-step estimators as a one-step generalized method of moments (GMM) estimator, which solves the moment functions stacked from both steps.

Step 3 & 4

- Step 3: Population average effect estimation
  - Take average over site-specific estimate estimates
- Step 4: Between-site variance estimation
  - Purge average within-site variance from between-site variance of site-specific effect estimates.

Step 5: Sensitivity analysis

- Omitted pretreatment covariates may cause bias, and there is a lack of adjustment strategies for post-treatment covariates.
- Sensitivity analysis may be conducted to assess the impact of omitted pretreatment and posttreatment confounders.
- Conceptualizing and investigating the relationships among multiple mediators may help with handling observed post-treatment covariates.

Analysis Results

	Average Effect		Between-Site Variance	
	Estimate	Effect Size	Estimate	95% Plausible Values
ITT on mediator (odds ratio)	2.39 (p<0.001)			
ITT on outcome	21.84 (SE=5.19, p<0.001)	0.12	635.44 (p=0.03)	(-27.57, 71.25)
Direct Effect	13.47 (SE=5.18, p=0.009)	0.07	597.97 (p=0.045)	(-34.46, 61.40)
Indirect Effect	8.38 (SE=1.40, p<0.001)	0.04	30.26 (p=0.05)	(-2.40, 19.16)
Correlation			0.03	

Strengths

Unlike multilevel SEM, our method

- can directly estimate not only the population average direct effect and indirect effect, but also the between-site variance of the effects. This enables researchers to investigate possible heterogeneity of mediation mechanisms across sites. Such empirical information is essential for improving future intervention designs.
- relaxes the assumption of no treatment-by-mediator interaction.
- greatly simplifies the outcome model specification and therefore avoids possible outcome model misspecifications.

Potential Concerns

- The proposed estimation procedure may not be optimal when sites sizes are extremely small.
- SUTVA needs to be relaxed if an individual's potential mediators and potential outcomes are each a function not only of one's own treatment assignment but also of other individuals' treatment assignments, or if an individual's potential outcomes could additionally be affected by other individuals' mediator values.