

Đánh giá Chính sách

Bài 3: Nhập môn Thử nghiệm Ngẫu nhiên có Kiểm soát (RCTs)

Edmund Malesky, Ph.D.

June 22, 2018

Duke University

Nội dung

- Tại sao lại thực hiện thử nghiệm ngẫu nhiên có kiểm soát?
- Các hình thức thử nghiệm ngẫu nhiên
 - ATE vs. ITE vs. TET
 - Ví dụ JTPA
 - Thiết kế thăng tiến/khuyến khích

Quá nóng bỏng

- Trong vài năm gần đây, việc sử dụng thử nghiệm ngẫu nhiên đã bùng nổ trong lĩnh vực khoa học xã hội, đặc biệt trong kinh tế học vi mô ứng dụng và kinh tế học phát triển.
- Gần đây, thử nghiệm ngẫu nhiên đã trở thành xu hướng trong số các nhà kinh tế học phát triển và nhà nghiên cứu chính trị.



Tiêu chuẩn vàng

- Thử nghiệm ngẫu nhiên có kiểm soát (RCT), mặc dù không được chấp nhận một cách phổ thông, đã bắt đầu được coi là “tiêu chuẩn vàng của đánh giá tác động chính sách”.
- Phòng thí nghiệm nghiên cứu về đói nghèo Jameel Poverty Action Lab (J-PAL) tại MIT, dưới sự lãnh đạo của Esther Duflo và Abhijit Banerjee, là trung tâm nghiên cứu quan trọng.
- J-PAL gây ra tranh cãi lớn khi cho định rằng chỉ khoảng 2% các dự án của Ngân hàng Thế giới được đánh giá đúng đắn sử dụng RCT.
- Các tranh cãi này vẫn đang tiếp tục...



Kết hợp nhiều kỹ năng

- Để thực hiện một thử nghiệm ngẫu nhiên, chúng ta cần phải sử dụng rất nhiều công cụ kỹ thuật sẽ được học trong môn học này.
 - Xác định đối tượng nghiên cứu
 - Đặt câu hỏi đúng
 - Thiết kế khảo sát
 - Kỹ năng thiết kế mẫu
 - Lựa chọn đối tượng khảo sát
 - Phân tích kinh tế lượng
 - Nếu làm đúng thì chúng ta không cần thiết phải sử dụng các kỹ thuật phân tích phức tạp.

Tóm tắt lịch sử



- Không hoàn toàn mới trong khoa học xã hội
- Các nhà tâm lý học đã tiến hành từ đầu thế kỷ 18
- Harold Gosnell bắt đầu sử dụng thử nghiệm trong nghiên cứu chính trị ở New York.
 - Việc sử dụng đội ngũ “tiếp thị chính trị tại nhà” cho các ứng cử viên có làm tăng tỷ lệ bỏ phiếu không?
 - Phân bổ ngẫu nhiên các khu phố được nhận thông báo về bầu cử
 - Tỷ lệ bỏ phiếu trong đợt bầu cử tổng thống tăng 1% năm 1924, tăng 9% trong bầu cử địa phương năm 1925.
- Nói chung, khoa học chính trị và kinh tế học thường đi sau trong ứng dụng.
- Mặc dù khoa học chính trị có cuộc cách mạng về nghiên cứu hành vi trong thập kỷ 50 và 60, các nghiên cứu chủ yếu dựa vào khảo sát thay vì thử nghiệm. Study of specific “real world” behavioral domain (unlike psychology)
 - Hoàn toàn có thể hiểu được về “tính ảo” của thử nghiệm.
 - Phương pháp thử nghiệm bắt buộc phải tạo ra tính chất ảo, do đó dễ bị nghi ngờ.
 - Không phải lúc nào cũng có thể kiểm soát và điều khiển hành vi

Đối tác trong công cuộc tìm kiếm sự thật

1. **Chính phủ**: Cần thiết đối với các chính sách áp dụng cho toàn bộ dân cư, tuy nhiên chúng ta cần phải thử trước khi mở rộng chương trình.
 - Để làm việc với chính phủ yêu cầu phải đạt được đồng thuận cao và có thể gặp khó khăn đối với những chính trị gia có cử tri bị ảnh hưởng bởi chương trình.
2. **NGOs**: Ít gặp phải vấn đề phân biệt đối xử bởi các chương trình của NGO đều khá độc lập và được tùy biến phụ thuộc vào bối cảnh.
 - Nhưng liệu kết quả thu được phụ thuộc vào các nhân tố đặc biệt không thể ứng dụng trong các môi trường khác như văn hóa tổ chức?
3. **Các tổ chức đa biên** như World Bank, Asian Development Bank...
4. **Các doanh nghiệp**: Đặc biệt trong vấn đề cung cấp tín dụng vi mô.

Tại sao phải sử dụng thử nghiệm ngẫu nhiên?

Ôn lại: Nhân quả và Phản thực

- Quan trọng nhất đối với đánh giá tác động chính sách là ước lượng phản thực bằng các nhóm so sánh hợp lệ
- Nhóm so sánh không hợp lệ dẫn đến ước lượng tác động của chương trình bị nhiễu bởi các nhân tố khác như khác biệt nhóm
- 2 phương pháp rất dễ gây ra vấn đề phản thực không hợp lệ:
 1. So sánh kết quả của nhóm tham gia trước và sau khi có chương trình.
 2. So sánh kết quả của nhóm có và không tham gia chương trình.
- Ngược lại, ngẫu nhiên hóa là tiêu chuẩn vàng để đánh giá tác động chính sách.
- Phân bổ ngẫu nhiên cho phép ước lượng phản thực hợp lệ.



Lợi ích của ngẫu nhiên hóa

- Phân bố ngẫu nhiên trong nhóm đối tượng có thể tham gia đảm bảo nguyên tắc công bằng và minh bạch khi thực hiện chương trình
- Do các hạn chế về ngân sách hay năng lực dẫn đến chương trình khó tiếp cận được nhóm đối tượng dự kiến.
- Phân bố ngẫu nhiên thay vì phân bố dựa trên các đặc tính quan sát được hoặc quy tắc đến trước phục vụ trước.
- Phân bố ngẫu nhiên đảm bảo hai nhóm tham gia và đối chứng tương đồng về mặt thống kê, nếu cỡ mẫu N đủ lớn.
- Nếu N lớn, phân bố ngẫu nhiên đảm bảo các đặc tính quan sát được và không quan sát được của 2 nhóm tương đồng.
- Do đó, phân bố ngẫu nhiên đảm bảo tìm được nhóm so sánh hợp lệ (phản thực).

Ôn lại: Khung lý thuyết kết quả tiềm năng

- ▶ Outcome

Y_i = Observed outcome for unit i

- ▶ Treatment

D_i : Indicator of whether unit i received treatment

$$D_i = \begin{cases} 1 & \text{unit } i \text{ received treatment} \\ 0 & \text{unit } i \text{ did not receive treatment} \end{cases}$$

- ▶ Potential Outcomes

Y_{1i} : Potential outcome for i with treatment

Y_{0i} : Potential outcome for i without treatment

Phân bổ ngẫu nhiên và vấn đề chệch mang tính lựa chọn

- ▶ Comparing outcomes for the treated and untreated often yields incorrect estimates

$$E[Y|D = 1] - E[Y|D = 0]$$

$$E[Y_1|D = 1] - E[Y_0|D = 0]$$

$$E[Y_1|D = 1] - E[Y_0|D = 0] + (E[Y_0|D = 1] - E[Y_0|D = 1])$$

which rearranges to:

$$E[Y_1|D = 1] - E[Y_0|D = 1] + E[Y_0|D = 1] - E[Y_0|D = 0]$$

which consists of the Average Treatment Effect on the Treated (ATET) and Selection Bias:

$$\underbrace{E[Y_1 - Y_0|D = 1]}_{\text{ATET}}$$

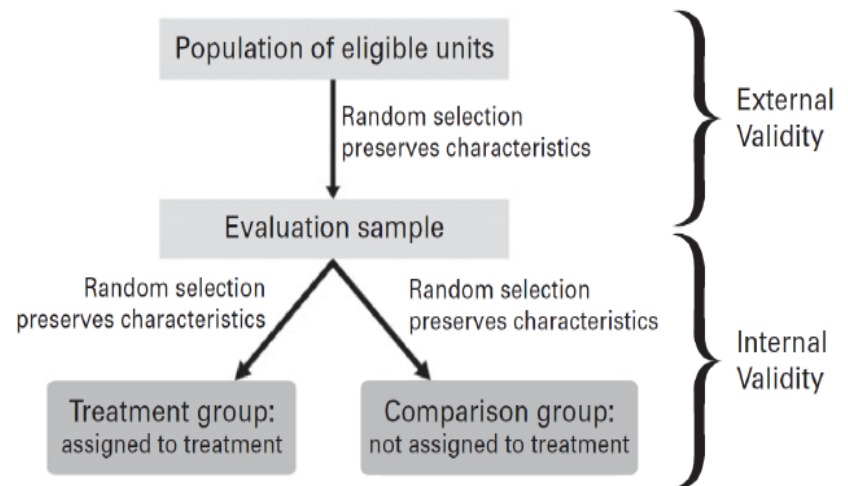
$$+ \underbrace{E[Y_0|D = 1] - E[Y_0|D = 0]}_{\text{Selection Bias}}$$

Nếu việc tham gia chương trình được phân bổ ngẫu nhiên, kết quả của 2 nhóm T và C khác biệt nhau về mặt kỳ vọng là có được tham gia chương trình hay không. Nếu không có chương trình, kết quả của 2 nhóm sẽ giống nhau. **DO ĐÓ KHÔNG CÓ CHỆCH MANG TÍNH LỰA CHỌN!**

Hiệu lực nội tại và Hiệu lực ngoại vi

Lấy mẫu ngẫu nhiên và Phân bổ ngẫu nhiên rất quan trọng để đảm bảo hiệu lực nội tại và hiệu lực ngoại vi của ước lượng

- *Hiệu lực ngoại vi*: Mẫu đánh giá phản ánh đúng quần thể của các cá nhân có thể tham gia chương trình.
- Lấy mẫu ngẫu nhiên đảm bảo mẫu đánh giá là đại diện của toàn bộ quần thể
- *Hiệu lực nội tại*: Do nhóm so sánh hợp lệ nên không có các nhân tố phức hợp (confound) làm nhiễu kết quả ước lượng.
- Phân bổ ngẫu nhiên nên nhóm so sánh tương đồng với nhóm tham gia ở thời điểm trước khi có chương trình.



Source: Gertler, 2011.

Các vấn đề phức tạp nhất thường nằm ở giai đoạn chuẩn bị ban đầu

Thử nghiệm ngẫu nhiên đã nhanh chóng được chấp nhận trong nghiên cứu chính sách.

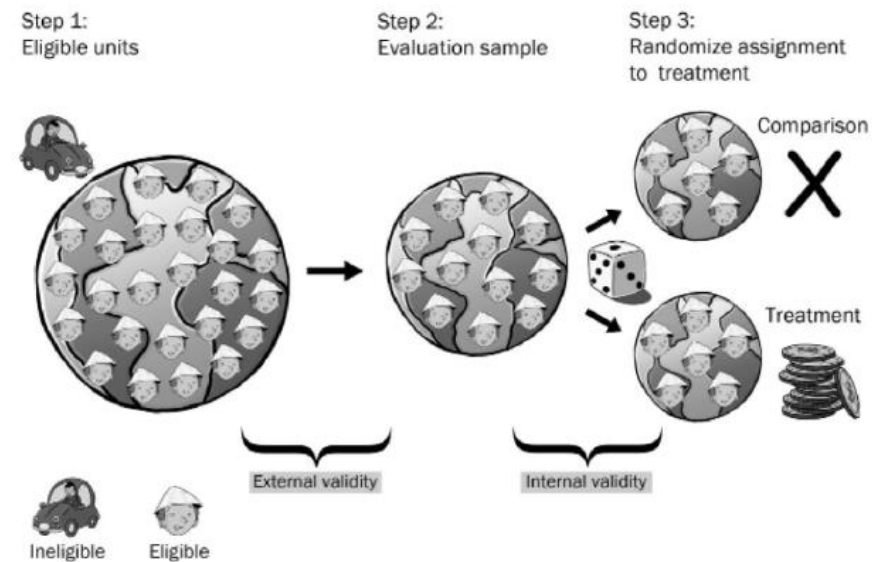
- Nếu mục tiêu của nghiên cứu chính sách là để tác động tới giới làm chính sách, bằng chứng của thử nghiệm ngẫu nhiên rất xác thực và minh bạch. Các chương trình thử nghiệm ngẫu nhiên như *Progres*a ở Mexico đã có tác động rất lớn lên chính sách được ban hành.
- Mặc dù công cụ phân tích định lượng sử dụng trong thử nghiệm ngẫu nhiên khá đơn giản, chúng ẩn chứa đằng sau những yêu cầu rất khắt khe về thiết kế nghiên cứu. Do đó, các bước thực hiện rất được coi trọng!
- Những thách thức lớn nhất không phải là cách thức thử nghiệm, mà là hiểu được giới hạn của công cụ thử nghiệm thành công đến đâu ở những môi trường khác nhau, và thiết kế nghiên cứu xung quanh những hạn chế của môi trường này.

Những quan ngại khác về hiệu lực ngoại vi

- **Hiệu lực ngoại vi:** Có thể áp dụng kết quả của chương trình đánh giá cho các mẫu nghiên cứu hay quần thể khác?
 - Hiệu lực nội tại là điều kiện cần nhưng không phải là điều kiện đủ để có hiệu lực ngoại vi
- Để khái quát hóa kết quả ra toàn bộ quần thể, việc thiết kế mẫu ngẫu nhiên rất quan trọng
- Đánh giá ngẫu nhiên hóa không thể tách biệt được tác động cân bằng tổng quát (general equilibrium effects) do:
 - So sánh sự khác biệt giữa nhóm tham gia và đối chứng trong bối cảnh cụ thể
 - Có thể áp dụng để đánh giá tác động cân bằng tổng quát ở cấp độ làng/xã, nhưng không thể áp dụng ở cấp độ quốc gia/quốc tế
- Khó có thể áp dụng đồng đều kết quả từ một quốc gia sang một quốc gia khác, hay từ một bối cảnh nghiên cứu/mục tiêu nghiên cứu cụ thể
- Thử nghiệm thường chỉ được tiến hành ở một khu vực nhỏ do vấn đề hậu cần
- Các thử nghiệm chỉ có giá trị cho chính các nghiên cứu đó, hay nó có thể được mô phỏng áp dụng ở những bối cảnh chung rộng lớn hơn?
 - Khó có thể xác định tác động của một chương trình được thiết kế gần giống nhưng không hoàn toàn tương đồng với một nghiên cứu gốc
 - Cần thiết phải vận dụng lý thuyết và mô phỏng để khái quát hóa một kết quả nghiên cứu ở quy mô lớn hơn.

Các bước để thực hiện phân bổ ngẫu nhiên

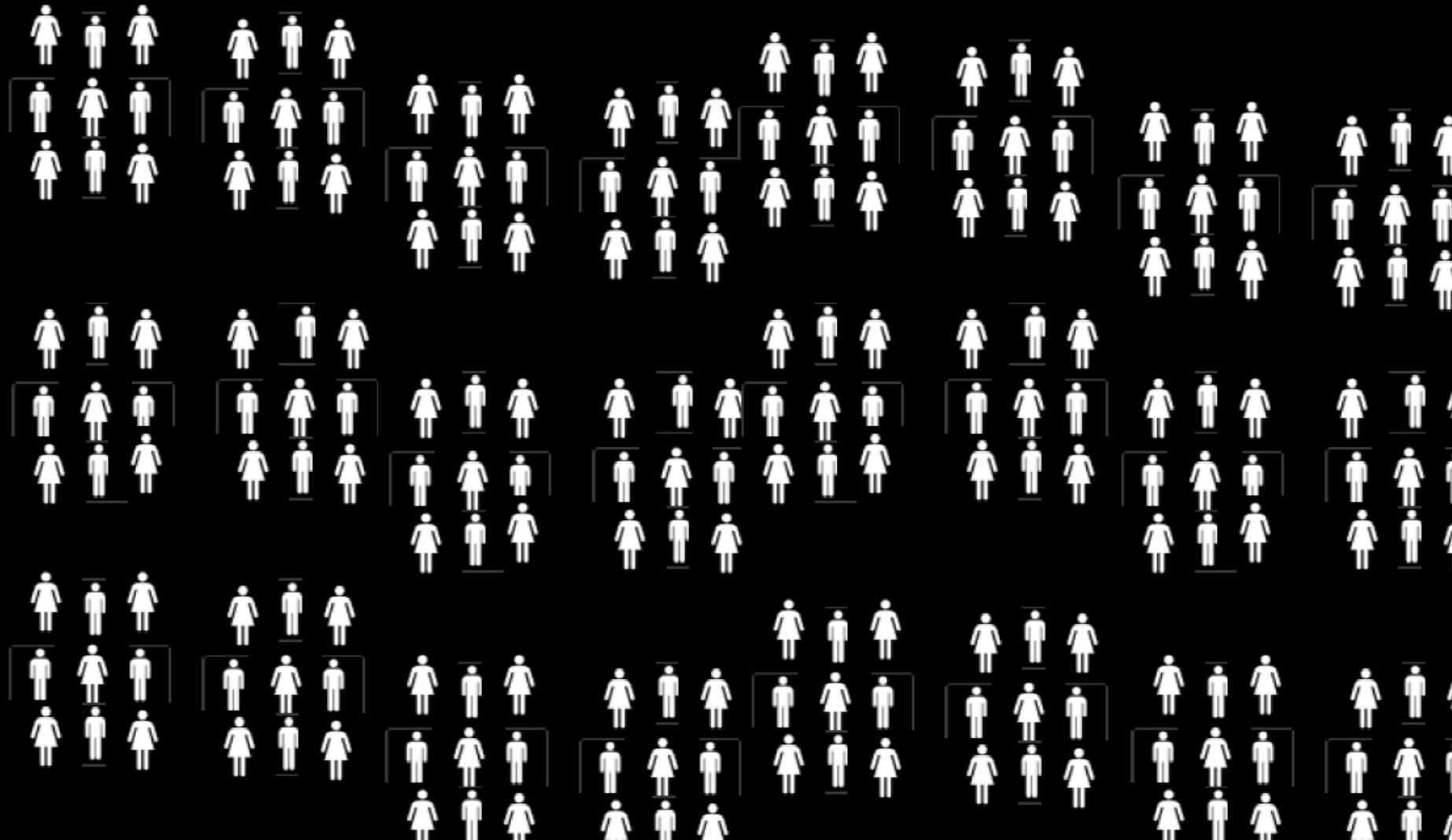
1. **Xác định các cá nhân/quan sát phù hợp có thể tham gia chương trình**
2. **Xác định kích cỡ mẫu sử dụng công thức tính độ vững và sai số (power calculation)**
 - Cần cỡ mẫu lớn nếu muốn phát hiện tác động nhỏ, tần suất thành công của chương trình thấp, hoặc có độ dao động lớn của kết quả, hoặc nếu muốn so sánh sự khác biệt giữa các nhóm trong mẫu
3. **Lựa chọn mẫu, tốt nhất là ngẫu nhiên**
 - Sử dụng các kỹ thuật được dạy ở lớp học
4. **Phân bổ nhóm tham gia và đối chứng sử dụng nguyên tắc minh bạch được xác lập trước khi bắt đầu thử nghiệm:**
 - Tung đồng xu, xúc xắc, bốc thăm, hay lấy số ngẫu nhiên
 - Ghi lại, hoặc mô phỏng lại được với mã số tham chiếu (seed) khi mô phỏng chuỗi ngẫu nhiên



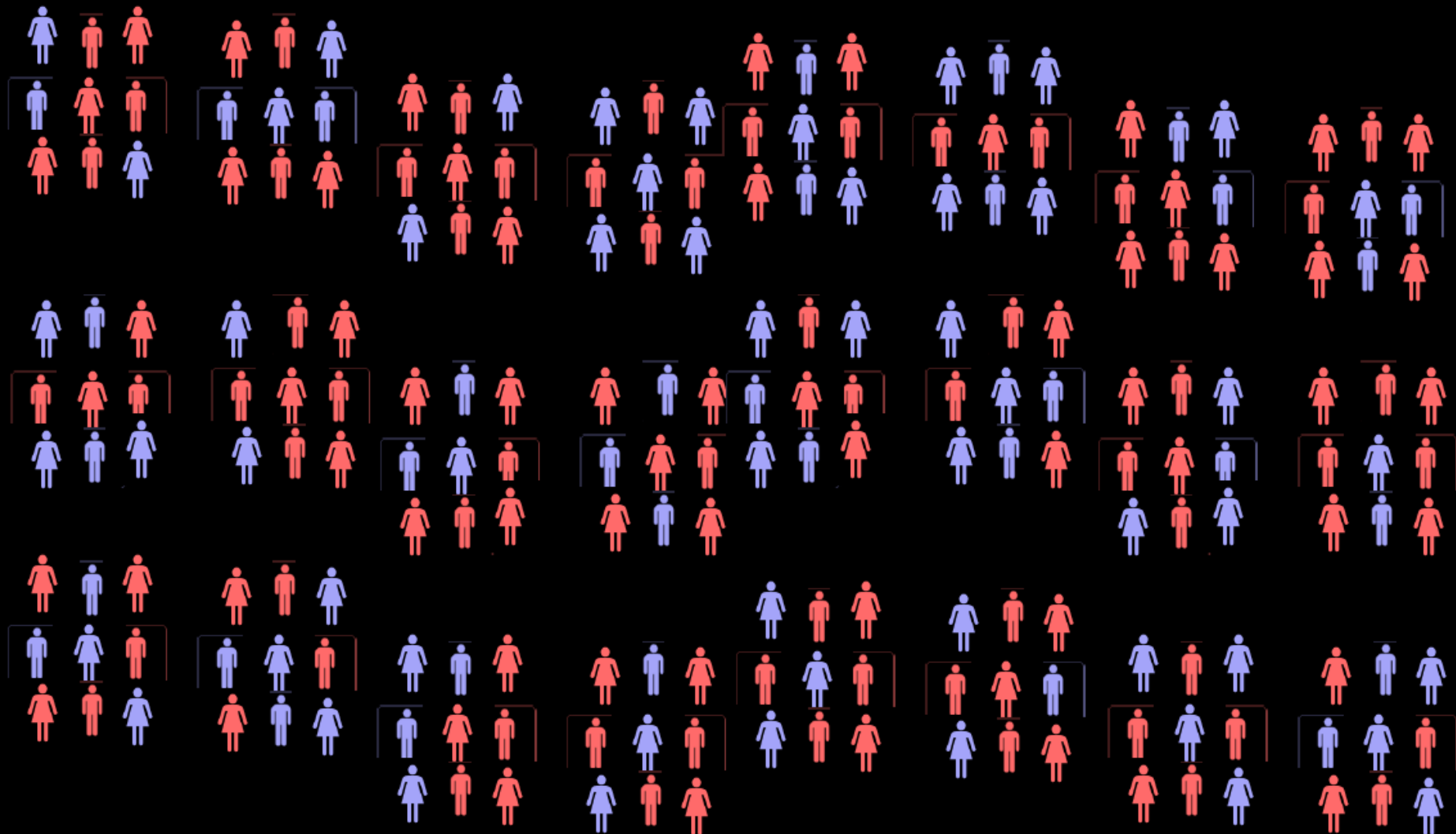
Loại hình đối tượng phân bổ ngẫu nhiên (tùy chọn)

- Có hai lựa chọn
 - Cấp độ cá nhân
 - Cấp độ nhóm: Thử nghiệm ngẫu nhiên ở cấp độ nhóm

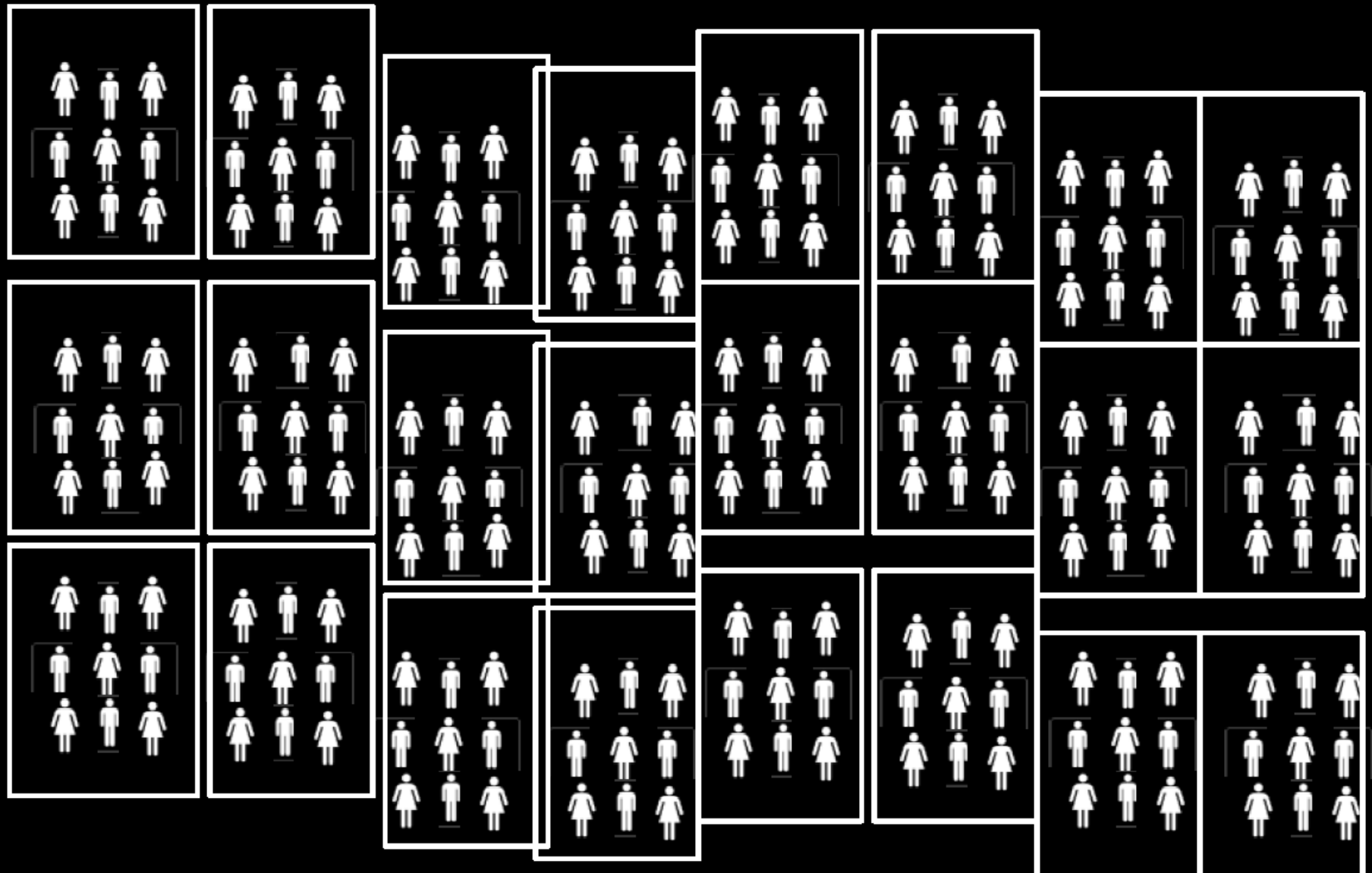
Unit of Randomization: Individual?



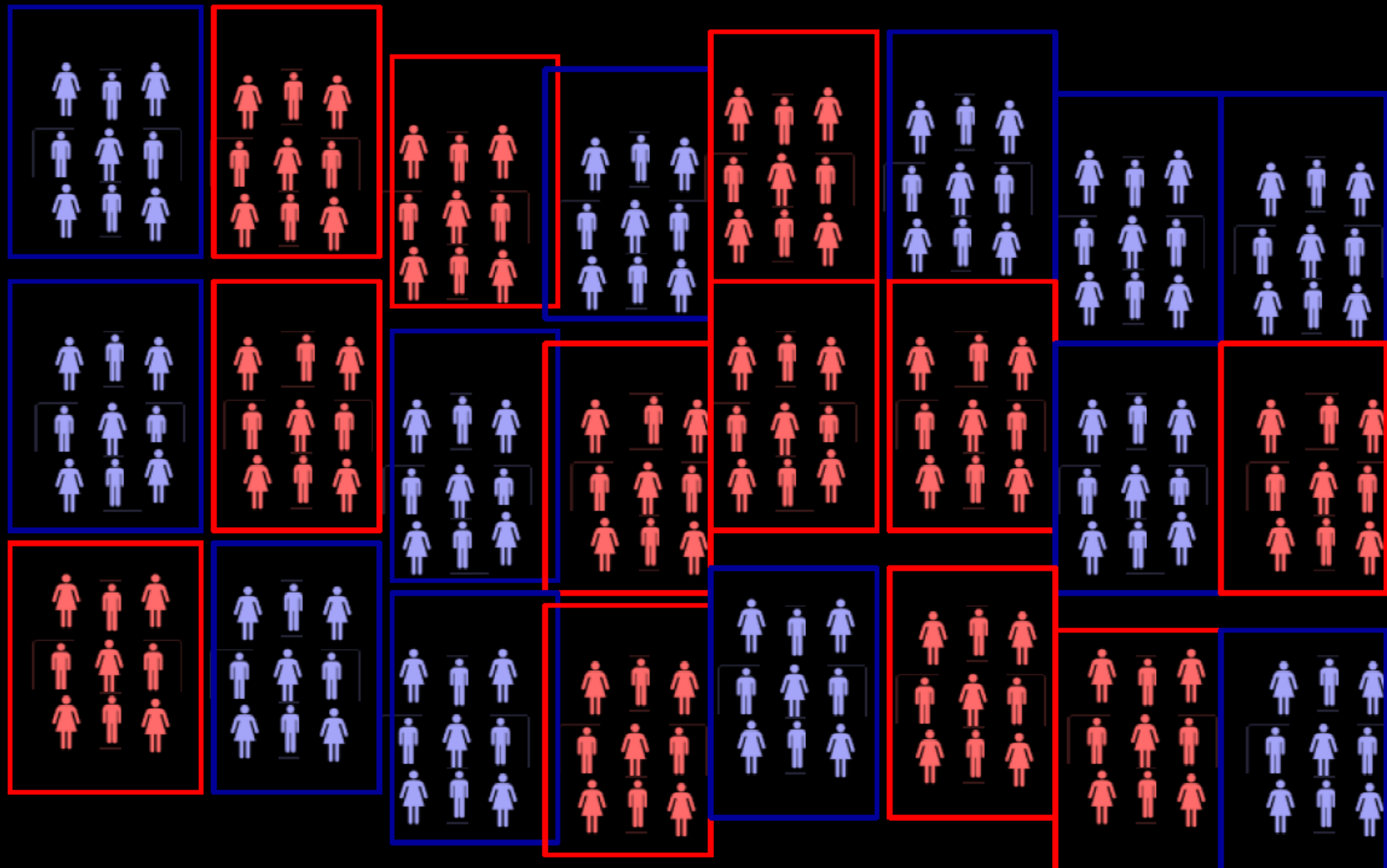
Unit of Randomization: Individual?



Unit of Randomization: Class?



Unit of Randomization: Class?



Lựa chọn cấp độ thử nghiệm nào?

- Bản chất của việc thử nghiệm
 - Chương trình can thiệp được thực hiện như thế nào?
 - Tác động tiềm năng có thể mở rộng đến cấp độ nào?
- Mức độ tích hợp các dữ liệu có sẵn
- Yêu cầu về kích cỡ mẫu và sai số
- Thông thường thì nên thiết kế thử nghiệm ngẫu nhiên ở cùng cấp độ với chính sách can thiệp

Phương pháp thử nghiệm ngẫu nhiên

1. Thiết kế thử nghiệm lâm sàng truyền thống

- Phân bổ ngẫu nhiên vào nhóm tham gia và nhóm đối chứng

2. Phương pháp đăng ký vượt (Oversubscription)

- Nguồn lực không cho phép tất cả mọi người đều tham gia chương trình. Do đó sử dụng biện pháp lựa chọn tham gia dựa trên bốc thăm. Động cơ của các cá nhân tham gia là giống nhau, do đó cho phép so sánh giữa các nhóm tham gia và không tham gia.

3. Trình tự ngẫu nhiên hóa theo giai đoạn

- Chương trình được thiết kế theo giai đoạn, và cho phép so sánh các nhóm ở các giai đoạn khác nhau (Miguel and Kremer)
 - Tuy nhiên cần cảnh giác là các nhóm chưa được lựa chọn hiện nay có thể thay đổi hành vi dựa trên kỳ vọng tương lai sẽ được lựa chọn tham gia. Do đó có thể có ảnh hưởng đến kết quả, có thể cả tích cực và tiêu cực.
 - Không thể ước lượng được tác động về dài hạn.

4. Ngẫu nhiên hóa trong nội nhóm

- Cho phép các nhóm nhỏ trong mẫu lớn được tham gia chương trình.
- Tương tự như trong một trường có thể có nhiều thang điểm
- Rủi ro lớn về tác động lan tỏa.

5. Thiết kế khuyến khích

- Đánh giá tác động của một can thiệp được cấp phổ quát cho tất cả mọi đối tượng, nhưng chỉ có một số cá nhân tham gia tự nguyện
 - Cung cấp thông tin

6. Thiết kế lan tỏa (spillover)

- Tận dụng khả năng lan tỏa, do đó thiết kế thử nghiệm ngẫu nhiên cấp độ áp dụng chương trình ở trong cùng một nhóm.
 - Thiết kế Mali về đại gia cặp chân dài (sugar daddy)

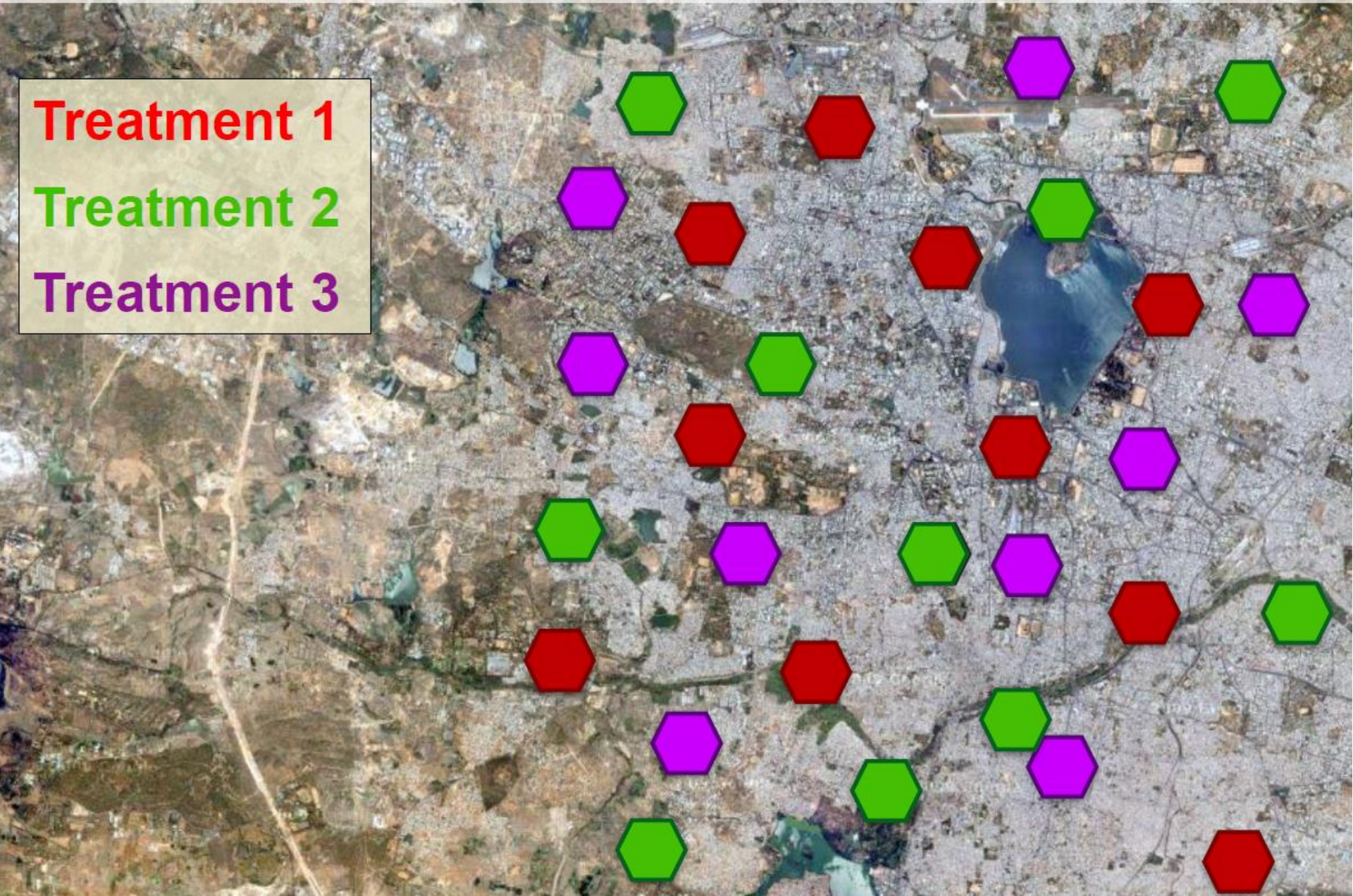


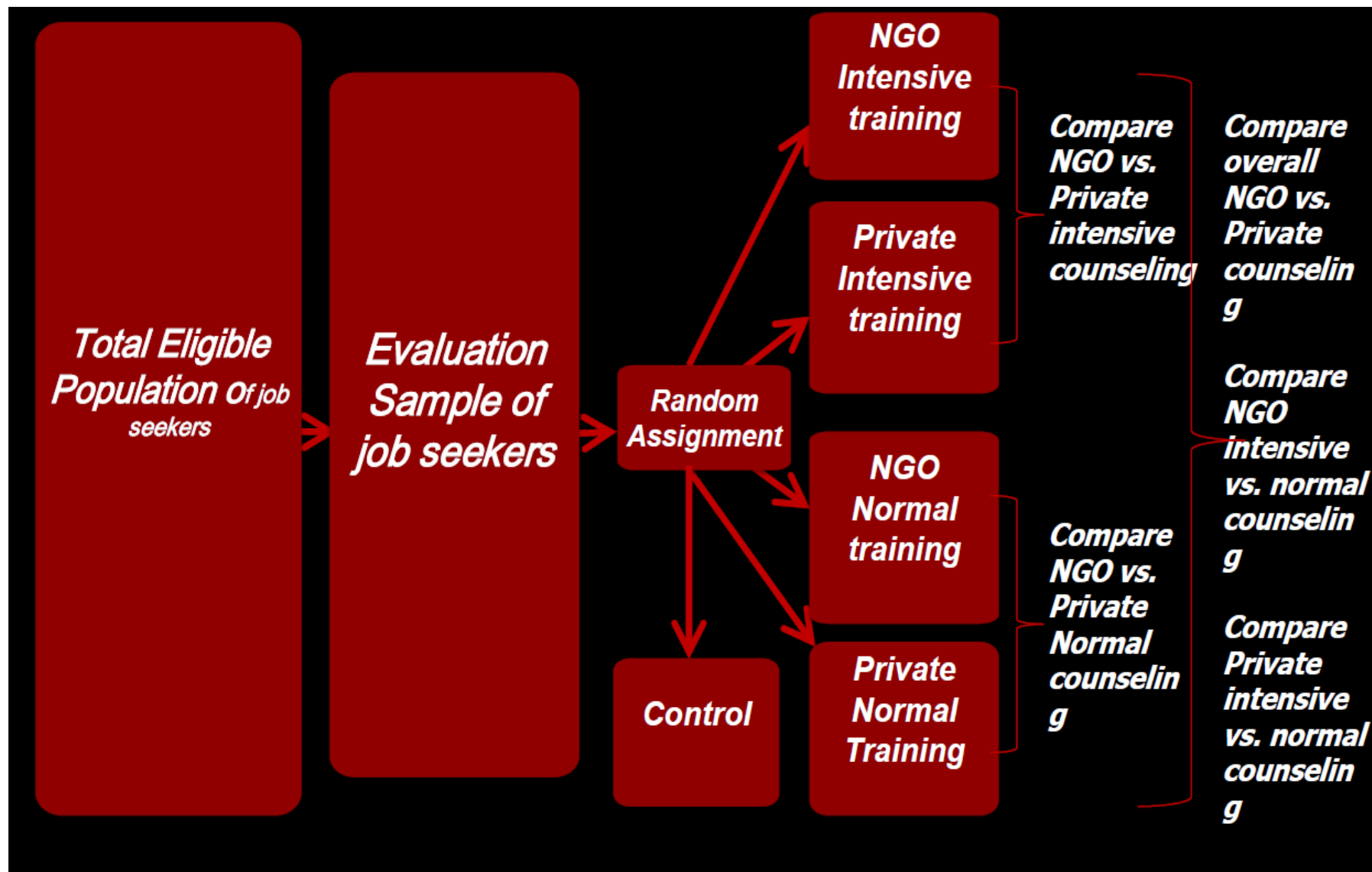
Multiple treatments

Treatment 1

Treatment 2

Treatment 3





Các nhân tố khác phải cân bằng

Household characteristics	Treatment villages (N = 2964)	Comparison villages (N = 2664)	Difference	t-stat
Health expenditures (\$ yearly per capita)	14.48	14.57	-0.09	-0.39
Head of household's age (years)	41.6	42.3	-0.7	-1.2
Spouse's age (years)	36.8	36.8	0.0	0.38
Head of household's education (years)	2.9	2.8	0.1	2.16*
Spouse's education (years)	2.7	2.6	0.1	0.006
Head of household is female = 1	0.07	0.07	-0.0	-0.66
Indigenous = 1	0.42	0.42	0.0	0.21
Number of household members	5.7	5.7	0.0	1.21
Has bathroom = 1	0.57	0.56	0.01	1.04
Hectares of land	1.67	1.71	-0.04	-1.35
Distance to hospital (km)	109	106	3	1.02

Source: Gertler, 2011.

Tác động: Phân bổ ngẫu nhiên với tuân thủ hoàn hảo

	Treatment	Comparison	Impact
	Average (Y) for the treatment group = 100	Average (Y) for the comparison group = 80	Impact = $\Delta Y = 20$
Enroll if, and only if, assigned to the treatment group			

(Comparison of Means)

	Treatment	Comparison	Difference	t-stat
Household health expenditures baseline	14.48	14.57	-0.09	-0.39
Household health expenditures follow-up	7.8	17.9	-10.1**	-25.6

(Regression Analysis)

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	-10.1** (0.39)	-10.0** (0.34)

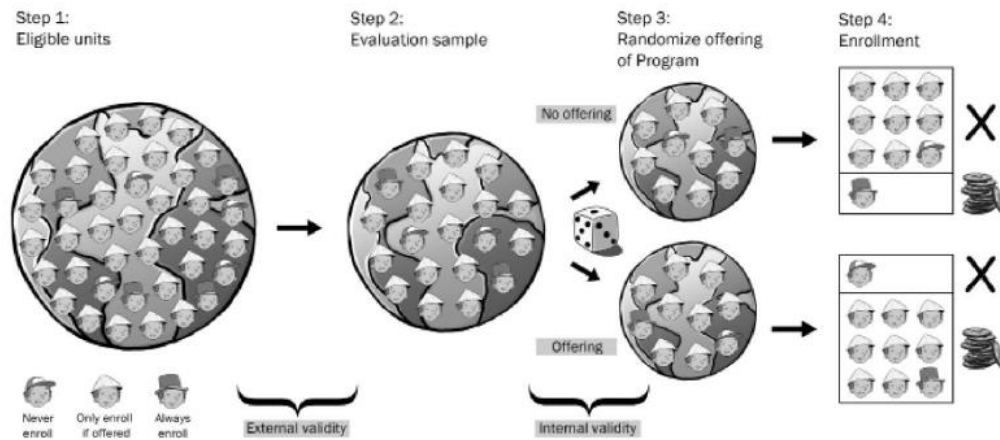
Source: Gertler, 2011.

Thử nghiệm không tuân thủ

- **Nhiều chương trình mang tính tự nguyện, do đó tuân thủ là không hoàn hảo**
 - Một số đối tượng phân bổ vào nhóm hưởng lợi nhưng không tham gia
 - Một số đối tượng phân bổ vào nhóm đối chứng nhưng lại tham gia vào chương trình bằng cách nào đó
- **3 loại hình cá thể tồn tại trong một chương trình đánh giá (giả định không có những kẻ phá bĩnh - “defiers”):**
 - 1) Những người tuân thủ - Compliers (được phân bổ vào một nhóm, và thực hiện theo đúng yêu cầu của chương trình)
 - 2) Những người luôn tham gia - Always-takers (Luôn tham gia vào chương trình, bất chấp việc được phân bổ vào nhóm nào)
 - 3) Những người không bao giờ tham gia - Never-takers (Không bao giờ tham gia vào chương trình, bất chấp việc được phân bổ vào nhóm nào)
- **Không thể nhận diện, nhưng có cùng đặc tính với các nhóm hưởng lợi và đối chứng**
 - Phân bổ vào nhóm hưởng lợi làm tăng xác suất tham gia chương trình, nhưng không chắc chắn (xác suất tham gia < 1).
 - Với một số giả định (độc lập và đơn điệu), ước lượng Wald của tác động chính sách đối với nhóm tuân thủ được tính như sau:

$$\beta_W = \frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[T_i | Z_i = 1] - E[T_i | Z_i = 0]}$$

Ngẫu nhiên hóa đề xuất tham gia



	Group offered treatment	Group not offered treatment	Impact
	% enrolled = 90% Average Y for those offered treatment = 110	% enrolled = 10% Average Y for those not offered treatment = 70	$\Delta\% \text{ enrolled} = 80\%$ $\Delta Y = \text{ITT} = 40$ $\text{ToT} = 40/80\% = 50$
Never enroll			—
Only enroll if offered the program			
Always enroll			—

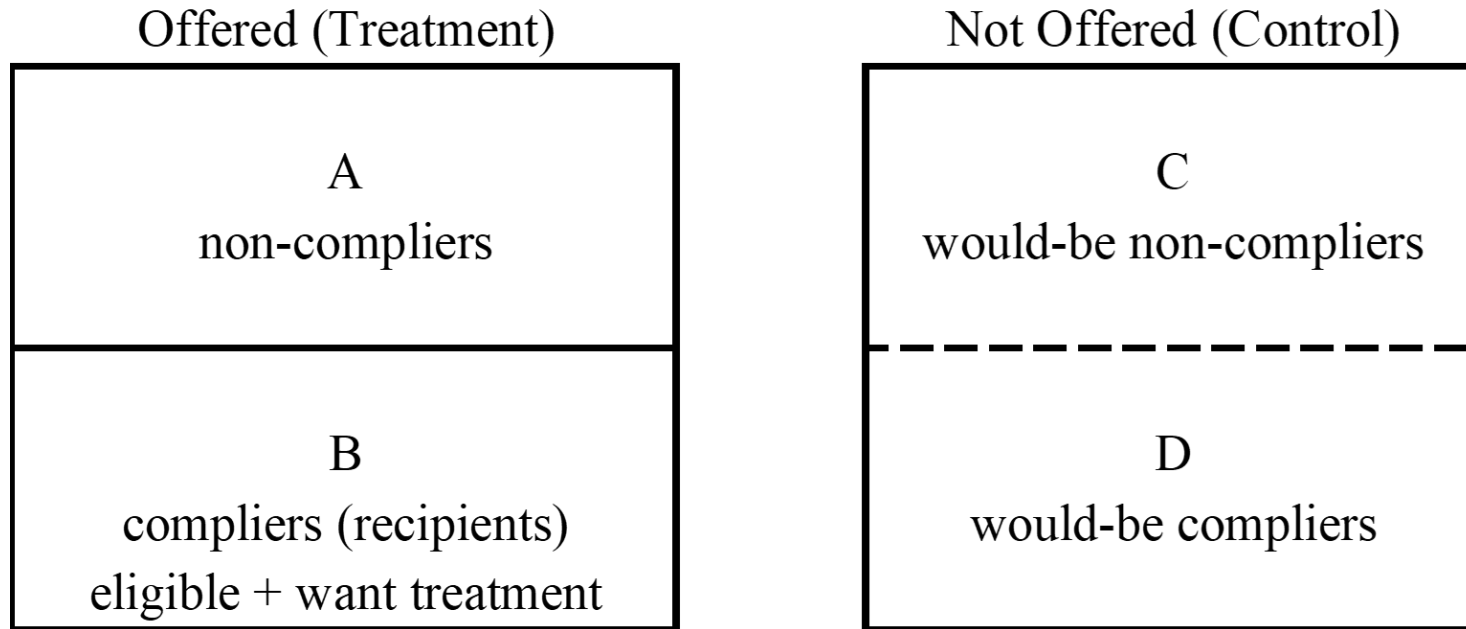
ITT: Intention to Treat.
Impact of program on those who are offered the treatment, *regardless of whether they actually enroll*

TOT: Treatment on Treated.
Impact of program on those who are offered treatment & *who actually enroll*

Source: Gertler 2011

- ITT: Tác động của chương trình đối với nhóm được đề xuất tham gia, bất chấp việc trên thực tế họ có tham gia hay không
- TOT: Tác động của chương trình đối với nhóm được đề xuất tham gia và trên thực tế có tham gia.

So sánh ITE và TET:



- ITE so sánh A & B với C & D.
- So sánh B với C & D có hàm chứa tác động chệch mang tính lựa chọn ngay (selection bias) cả khi việc lựa chọn tham gia chương trình được thực hiện ngẫu nhiên
- TET so sánh B với D, tuy nhiên làm thế nào để đảm bảo tuân thủ trong nhóm đối chứng?

Phân tích ý tưởng phân bổ ngẫu nhiên

Chúng ta quan tâm đến ước lượng nào: ATE, ITT hay TOT?

- Tác động can thiệp trung bình (Average Treatment Effect - ATE): Nếu chương trình được áp dụng phổ quát hoặc bắt buộc:
 - Tác động kỳ vọng đối với một cá nhân điển hình trong xã hội
- Nhà hoạch định chính sách quan tâm đến tác động đối với nhóm được đề xuất tham gia (intention to treat - ITT) khi mở rộng chương trình
 - Tác động có tính đến trường hợp đối tượng tham gia nhưng không tuân thủ
 - Phù hợp về mặt thực tiễn vì khó bắt buộc nhóm được lựa chọn tham gia
 - Ví dụ trẻ em không có mặt khi thực hiện chương trình tẩy giun, và khó có thể tìm địa chỉ nhà để cố gắng yêu cầu tham gia
 - Giúp trả lời câu hỏi: Tác động kỳ vọng lên cá nhân được đề xuất tham gia, bất kể trên thực tế cá nhân đó có thực hiện hay không
- Tuy nhiên tác động can thiệp lên đối tượng tham gia (treatment on treated - TOT) cũng rất quan trọng
 - Thể hiện tác động của việc tham gia chương trình nếu có thể thiết kế áp dụng với các công cụ khác
 - Ví dụ rất ít người muốn uống thêm vitamin sắt vi lượng, nhưng chúng ta muốn biết tác động của sắt lên cơ thể
 - Điều chỉnh cho thực tế là một số người được phân bổ vào nhóm hưởng lợi nhưng lại không tham gia vào chương trình, trong khi một số người phân bổ vào nhóm đối chứng nhưng lại tìm cách tham gia
 - Ước lượng phù hợp với nhóm tuân thủ, nhưng không phù hợp cho cả quần thể
 - Giúp trả lời câu hỏi: tác động kỳ vọng đối với cá nhân được đề xuất tham gia chương trình là gì nếu trên thực tế họ thực hiện theo đúng yêu cầu

Tác động muốn đánh giá sẽ quyết định thiết kế nghiên cứu

- Tác động can thiệp trung bình (Average Treatment Effect): có thể trực tiếp sử dụng thử nghiệm ngẫu nhiên.
- Tác động lên nhóm được đề xuất tham gia (Intention to Treat Effect):
 - Lấy mẫu ngẫu nhiên cho cả quần thể.
 - Tiến hành lựa chọn mẫu trong quần thể các nhóm đã được lựa một cách ngẫu nhiên
 - ITE được tính bằng sự khác biệt giữa kết quả của nhóm được đề xuất tham gia so với nhóm không.
- Tác động can thiệp lên đối tượng tham gia (Treatment Effect on Treated):
 - Chỉ sử dụng một nhóm các cá nhân được đề xuất tham gia và sẽ làm theo đúng yêu cầu. Đó là nhóm những người tuân thủ.
 - Phân bổ ngẫu nhiên việc tham gia chương trình trong nhóm tuân thủ này.
 - TET được tính bằng sự khác biệt về kết quả giữa nhóm có tham gia và không tham gia của những người tuân thủ.
 - Không dễ ước lượng TET trên thực tế vì yêu cầu phải xác định được nhóm những người tuân thủ và yêu cầu họ đăng ký vào chương trình nghiên cứu thử nghiệm. Việc thực hiện tương tự như bốc thăm.

Chương trình đào tạo nghề JTPA

- Đạo luật đào tạo nghề (JTPA): Nghiên cứu bởi Bộ Lao động Mỹ năm 1986 để đánh giá các chương trình đào tạo
- Ứng viên trong 16 chương trình JTPA được phân bổ ngẫu nhiên vào nhóm hưởng lợi và đối chứng
 - Nhóm hưởng lợi cho phép tham gia, nhóm đối chứng không cho phép tham gia ngay
- Thu thập dữ liệu ban đầu, 2 cuộc phỏng vấn điện thoại, và dữ liệu của từng bang
- 2/3 trong nhóm hưởng lợi trên thực tế đăng ký vào các chương trình đào tạo JTPA, trong khi dưới 2% thuộc nhóm đối chứng cũng đăng ký
- Không phải tất cả đều được tăng lương do chương trình JTPA, do có sự sụt giảm về thu nhập trước khi áp dụng chương trình
- Hỗ trợ tài chính cho chương trình đào tạo dành cho vị thành niên bị cắt giảm sau khi kết quả nghiên cứu được công bố.

Cân bằng giữa các nhóm thuộc chương trình JTPA (Với giới tính Nam chỉ cân bằng một phần)

MEANS AND STANDARD DEVIATIONS

	Entire Sample	Assignment		Difference (t-stat.)
		Treatment	Control	
A. Men				
Number of observations	5,102	3,399	1,703	
<i>Treatment</i>				
Training	.42 [.49]	.62 [.48]	.01 [.11]	.61 (70.34)
<i>Outcome variable</i>				
30 month earnings	19,147 [19,540]	19,520 [19,912]	18,404 [18,760]	1,116 (1.96)
<i>Baseline Characteristics</i>				
Age	32.91 [9.46]	32.85 [9.46]	33.04 [9.45]	-.19 (-.67)
High school or GED	.69 [.45]	.69 [.45]	.69 [.45]	-.00 (-.12)
Married	.35 [.47]	.36 [.47]	.34 [.46]	.02 (1.64)
Black	.25 [.44]	.25 [.44]	.25 [.44]	.00 (.04)
Hispanic	.10 [.30]	.10 [.30]	.09 [.29]	.01 (.70)
Worked less than 13 weeks in past year	.40 [.47]	.40 [.47]	.40 [.47]	.00 (.56)

Kết quả của đào tạo nghề (người lớn)

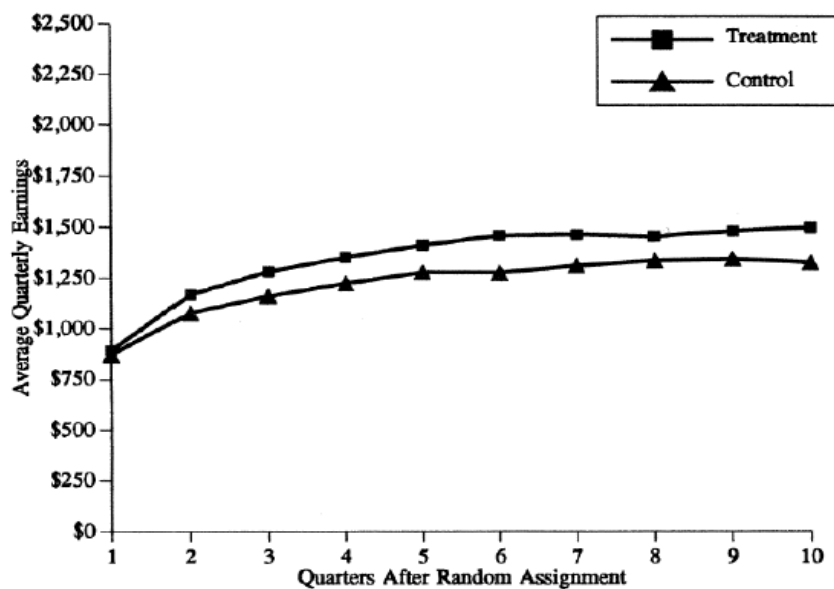


Figure 1A
Mean Earnings, by Quarter: Adult Women

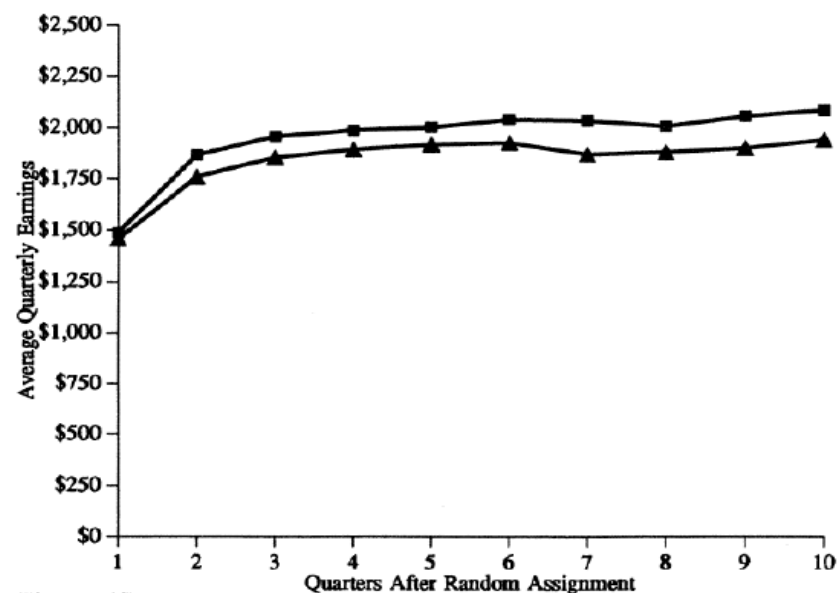


Figure 1B
Mean Earnings, by Quarter: Adult Men

Kết quả của đào tạo nghề (vị thành niên)

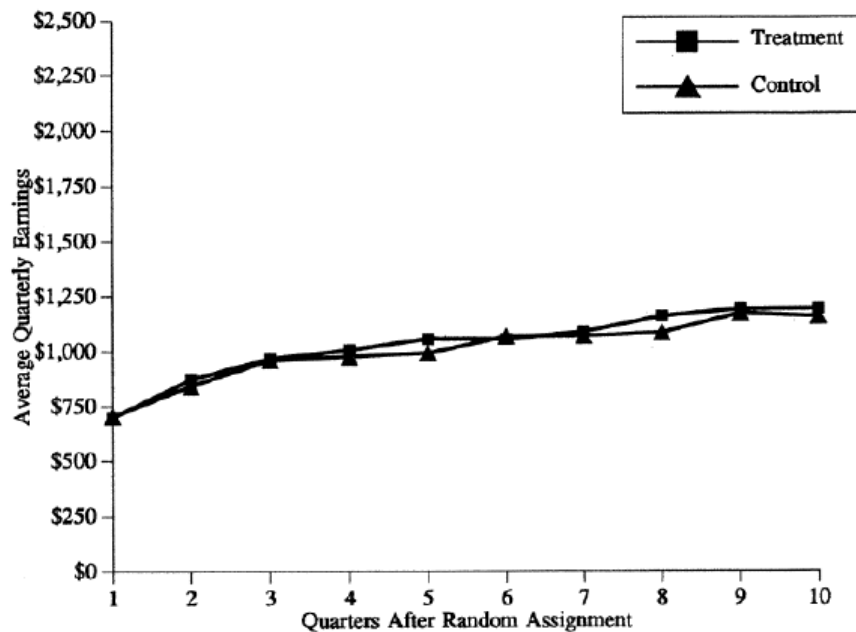


Figure 2A
Mean Earnings, by Quarter: Female Youth

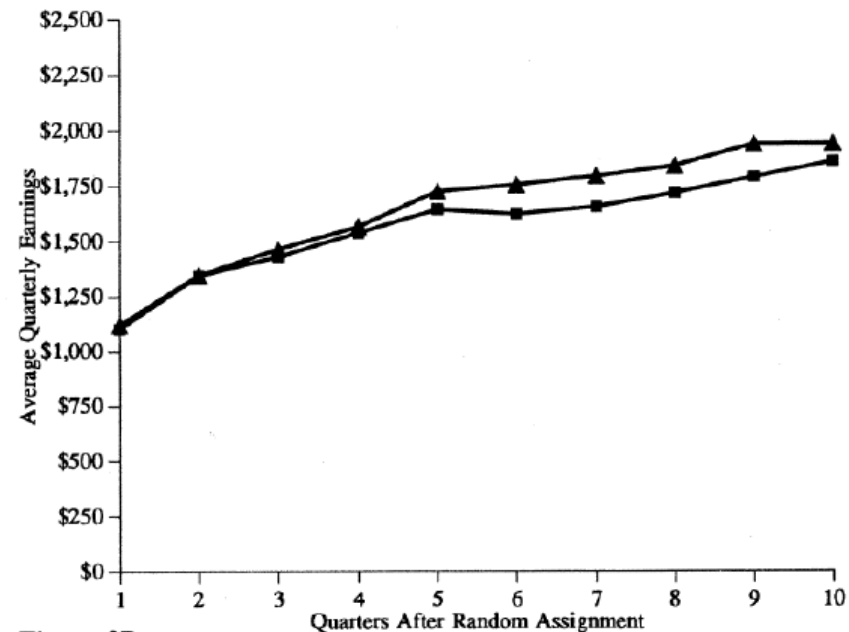


Figure 2B
Mean Earnings, by Quarter: Male Youth Non-arrestees

TET so với ITE

	Enrolled in Training	Not Enrolled in Training	Total
Assigned to Training	4,804	2,683	7,487
Assigned to Control	54	3,663	3,717
Total	4,858	6,346	11,204

TET

OLS AND IV ESTIMATES

ITE

IMPACTS

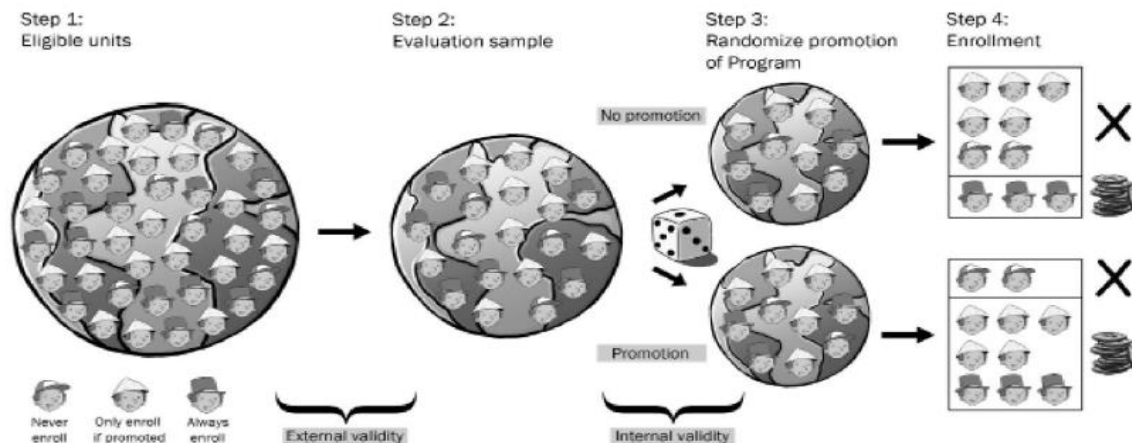
Corrected

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970 (555)	3,754 (536)	1,117 (569)	970 (546)	1,825 (928)	1,593 (895)
B. Women	2,133 (345)	2,215 (334)	1,243 (359)	1,139 (341)	1,942 (560)	1,780 (532)

Note: Columns (1) and (2) show the differences in earnings by training status; columns (3) and (4) show differences by assignment status. Columns (5) and (6) report the result of using assignment status as an instrument for training. The covariates used in columns (2), (5) and (6) are *High school or GED, Black, Hispanic, Married, Worked less than 13 weeks in past year, AFDC* (for women), plus indicators for the service strategy recommended, age group and second follow-up survey. Robust standard errors in parenthesis.

Thiết kế khuyến khích/Đề bạt ngẫu nhiên

- Thử nghiệm đề bạt ngẫu nhiên để ước lượng tác động của chương trình trong trường hợp không thể kiểm soát được vấn đề tham gia
 - Mặc dù có thể có ngân sách thực hiện, nhưng không khả thi về chính trị hay đạo đức để ngăn cản ai đó tham gia
 - Có thể lựa chọn ngẫu nhiên người được nhận đề bạt/khuyến khích (không phải là một can thiệp chính sách, bởi tất cả mọi đối tượng đều có thể tham gia chính sách)
- Các chương trình phổ biến thông tin hay khuyến khích có thể làm tăng mức độ chấp nhận trong nhóm ngẫu nhiên được lựa từ quần thể.
 - Dùng để tăng mức độ chấp nhận chương trình, nhưng không tác động đến kết quả trực tiếp (do quá tốn kém)
 - Các chương trình khuyến khích ảnh hưởng đến nhóm tuân thủ, bất chấp việc luôn có những đối tượng luôn tham gia và không bao giờ tham gia
 - Nhằm tạo ra nhóm so sánh (việc nhận được khuyến khích là đại diện cho nhóm hưởng lợi)



Ngẫu nhiên hóa khuyến khích

	Promoted group	Non-promoted group	Impact
	% enrolled = 80% Average Y for promoted group = 110	% enrolled = 30% Average Y for non-promoted group = 70	$\Delta\%$ enrolled = 50% $\Delta Y = 40$ Impact = $40/50\% = 80$
Never enroll			—
Only enroll if promoted			
Always enroll			—

Ngẫu nhiên hóa khuyến khích: Chương trình bảo hiểm y tế

Table 4.4 Case 4—HISP Impact Using Randomized Promotion (Comparison of Means)

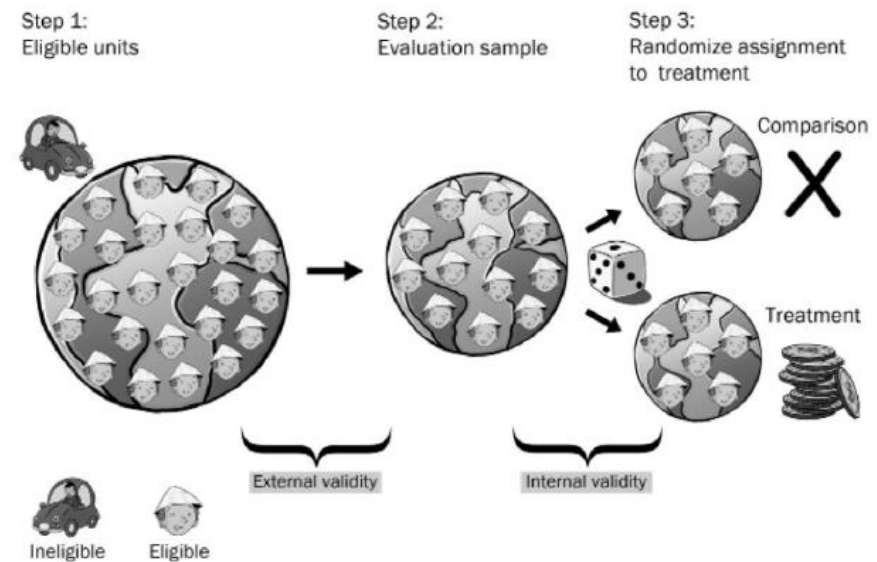
	Promoted villages	Nonpromoted villages	Difference	t-stat
Household health expenditures baseline	17.1	17.2	-0.1	-0.47
Household health expenditures follow-up	14.9	18.8	-3.9	-18.3
Enrollment in HISP	49.2%	8.4%	40.4%	

Table 4.5 Case 4—HISP Impact Using Randomized Promotion (Regression Analysis)

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	-9.4** (0.51)	-9.7** (0.45)

Các bước để thực hiện phân bổ ngẫu nhiên

1. **Xác định các cá nhân/quan sát phù hợp có thể tham gia chương trình**
2. **Xác định kích cỡ mẫu sử dụng công thức tính độ vững và sai số (power calculation)**
 - Cần cỡ mẫu lớn nếu muốn phát hiện tác động nhỏ, tần suất thành công của chương trình thấp, hoặc có độ dao động lớn của kết quả, hoặc nếu muốn so sánh sự khác biệt giữa các nhóm trong mẫu
3. **Lựa chọn mẫu, tốt nhất là ngẫu nhiên**
 - Sử dụng các kỹ thuật được dạy ở lớp học
4. **Phân bổ nhóm tham gia và đối chứng sử dụng nguyên tắc minh bạch được xác lập trước khi bắt đầu thử nghiệm:**
 - Tung đồng xu, xúc xắc, bốc thăm, hay lấy số ngẫu nhiên
 - Ghi lại, hoặc mô phỏng lại được với mã số tham chiếu (seed) khi mô phỏng chuỗi ngẫu nhiên



Kết luận

- Thử nghiệm là giải pháp lý tưởng đối với bài toán ước lượng quan hệ nhân quả.
- Các vấn đề phức tạp nhất thường nằm ở giai đoạn chuẩn bị ban đầu.
- Có rất nhiều các thử nghiệm để xử lý các vấn đề gặp phải trong khi thực hiện.
- Có nhiều hình thức ước lượng tác động của chương trình.
- Quan trọng nhất là sử dụng các công cụ phù hợp cho từng mục tiêu cụ thể.