

Đánh giá Chính sách

Bài 4: Thiết kế và thực hiện một thử nghiệm ngẫu nhiên có kiểm soát (RCT)

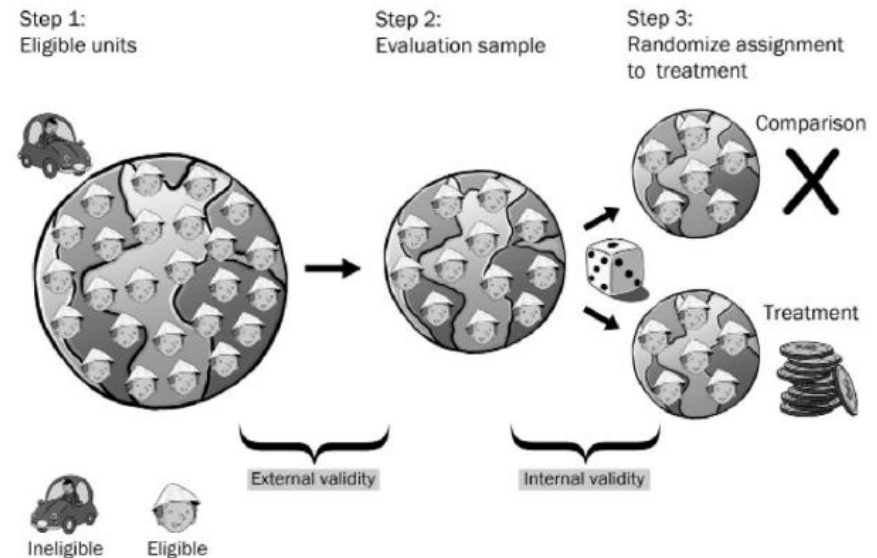
Edmund Malesky, Ph.D.

June 25, 2018

Duke University

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 xúc, 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



Bài giảng hôm nay

- Các ràng buộc thực tế
- Phương pháp ngẫu nhiên hóa
- Các dạng thực hiện đối với thử nghiệm tham gia-đối chứng đơn giản

Ràng buộc và nguồn lực

- Hầu hết các chương trình đều có hạn chế về nguồn lực
 - Số phiếu khuyến mãi, không gian thực hiện chương trình đào tạo, ngân quỹ cho những người hỗ trợ
- Dẫn đến có nhiều người muốn tham gia hơn là nguồn lực cho phép
- Việc bị hạn chế bởi nguồn lực cũng là cơ hội để đánh giá

Ràng buộc và tính công bằng

- Bốc thăm thực hiện khá đơn giản, phổ biến, và minh bạch
- Hữu ích khi không có các lý do nhân tiên để phải phân biệt đối xử
- Người tham gia biết ai được ai thua
- Bốc thăm đơn giản thường được nhìn nhận là công bằng

Ràng buộc: Tác động lan tỏa (contamination/spillover) và băng làn (cross-over)

- Nhóm đối chứng được sử dụng để ước lượng phản thực
- Nếu nhóm đối chứng khác với phản thực thì kết quả ước lượng có thể bị chệch
- Xảy ra khi có:
 - Tác động lan tỏa
 - Tác động băng làn

Tác động lan tỏa

- Khi có tác động lan tỏa, sự khác biệt đơn thuần giữa nhóm tham gia và nhóm kiểm soát không còn là tác động của chương trình nữa.
 - Có thể mang tính tích cực hoặc tiêu cực.
- Tác động lan tỏa có thể gây rắc rối cho thiết kế nghiên cứu khi mức độ tham gia bị hạn chế, tuy nhiên có thể thiết kế phương pháp xử lý khá dễ dàng để đo lường tác động lan tỏa trực tiếp.



Ước lượng tác động lan tỏa bằng thử nghiệm

Miguel & Kremer, *'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities'* (Giun: Nhận diện tác động lên giáo dục và sức khỏe khi xảy ra ngoại tác tham gia chương trình)

- Chương trình tẩy giun được áp dụng thử nghiệm ở cấp độ trường học
- Kiểm soát số học sinh trong một khoảng cách nhất định đến những học sinh không tham gia chương trình, và nghiên cứu tác động của chương trình lên kết quả đầu ra (giáo dục và sức khỏe) phụ thuộc như thế nào vào số học sinh được tham gia chương trình.
- Bởi vì chương trình được thực hiện ngẫu nhiên, cường độ tác động can thiệp cũng là ngẫu nhiên.

Baird, McIntosh, & Özler, *'Schooling, Income, & HIV Risk in Malawi'*. (Đi học, thu nhập, và rủi ro nhiễm HIV ở Malawi)

- Chương trình trợ cấp tiền mặt có điều kiện, được thử nghiệm ngẫu nhiên ở cấp độ làng xã.
- Mức độ bão hòa tình trạng tham gia chương trình được thử nghiệm ngẫu nhiên hóa, do đó cho phép so sánh những nữ sinh không được tham gia chương trình ở làng xã được chọn tham gia với nhóm kiểm soát được thiết lập là một hàm số của tỷ phần nữ sinh được chọn tham gia trong những làng được chọn tham gia.

Ràng buộc – tổ chức thực hiện

- Cần thiết phải nhận diện được các ràng buộc về tổ chức thực hiện trong khâu thiết kế chương trình.
 - Ví dụ từng nhân viên y tế thực hiện các hoạt động tẩy giun
 - Có rất nhiều nhiệm vụ khác, không chỉ là tẩy giun.
- Các nhân viên thực hiện nhiệm vụ với cả ba nhóm tham gia và đối chứng
- Có những nguyên tắc thủ tục khác nhau với các nhóm khác nhau?

Ràng buộc – tổ chức thực hiện

- Tính dễ nhận diện của việc được tham gia chương trình
- Ngẫu nhiên hóa ở cấp độ trẻ em trong mỗi lớp học
- Ngẫu nhiên hóa lớp học trong mỗi trường học
- Ngẫu nhiên hóa cấp độ thôn bản

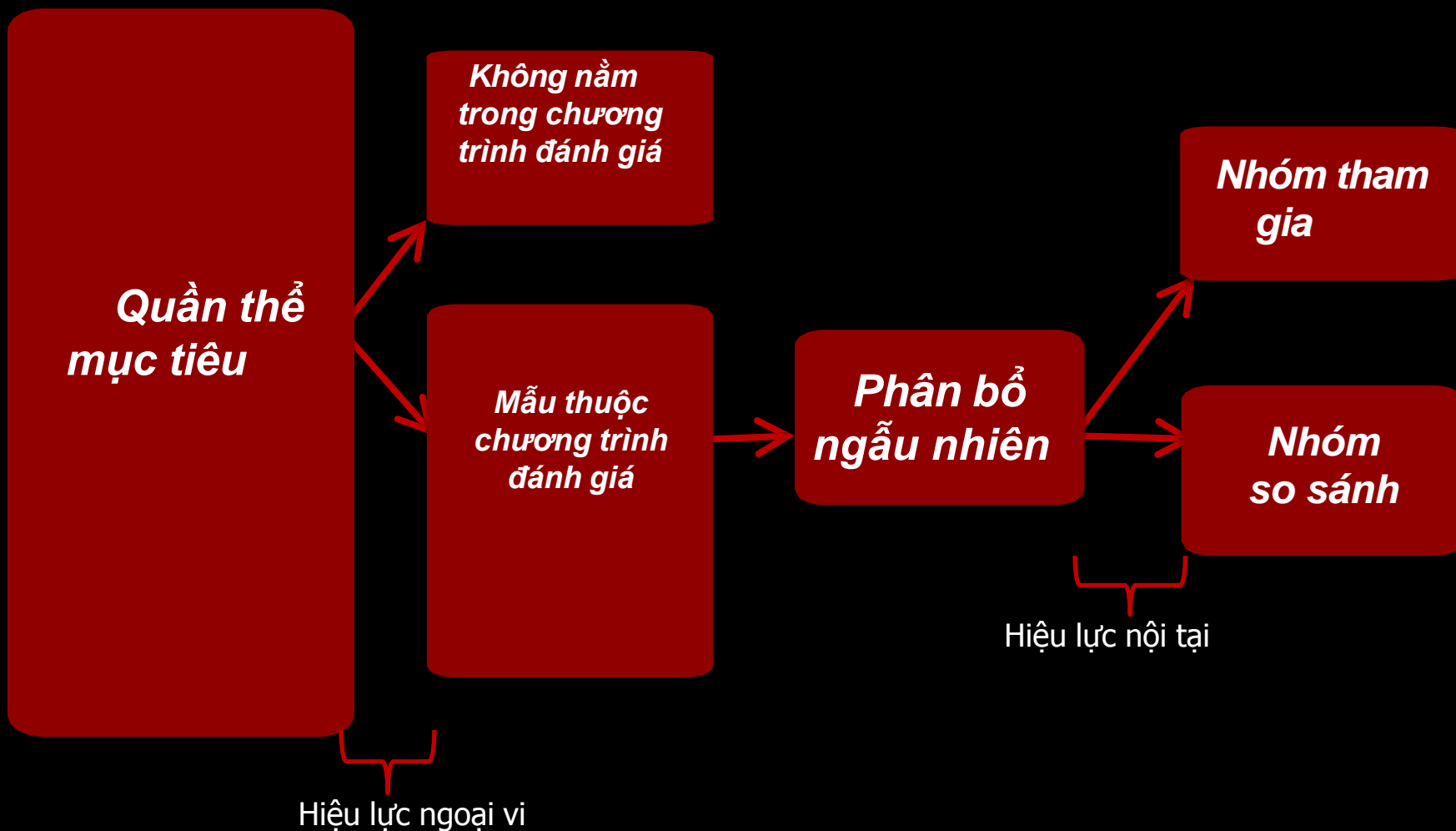
Ràng buộc – Số quan sát

- Chương trình chỉ có thể có quy mô phù hợp với một vài thôn bản
- Nhân tố rất quan trọng là khả năng ước lượng (statistical power): Quá ít quan sát thì khó có thể đo lường được tác động với độ chính xác cao
- Kích cỡ mẫu mong muốn được tính qua công thức tính khả năng phát hiện và sai số (power calculation), tuy nhiên không đề cập trong môn học này

Bài giảng hôm nay

- Các ràng buộc thực tế
- Phương pháp ngẫu nhiên hóa
- Các dạng thực hiện đối với thử nghiệm tham gia-đối chứng đơn giản

RCTs | Cấu trúc căn bản



Nhân tố cơ bản của RCT – Đo lường tác động

- Yêu cầu dữ liệu
 - Dữ liệu kết quả của nhóm tham gia và nhóm đối chứng
 - Dữ liệu tham chiếu (Baseline data) nếu có
- Tác động
 - Tác động can thiệp trung bình (Average Treatment Effect)

$$\blacksquare E[Y \downarrow 1 - Y \downarrow 0 \mid D=1] @ = y \downarrow T - y \downarrow C$$

- Thử nghiệm – Phản thực
- Trung bình tham gia – Trung bình đối chứng

THỨ TỰ NGẪU NHIÊN CỦA THIẾT KẾ THỬ NGHIỆM THEO GIAI ĐOẠN (PHASE-IN DESIGN)

Thử nghiệm theo giai đoạn: Tận dụng lợi thế của việc mở rộng chương trình

- Vấn đề đạo đức: Cuối cùng thì ai cũng được tham gia
- Thực tế: Phương pháp tiếp cận tự nhiên khi mở rộng quy mô của chương trình trong khi phải đối mặt với các hạn chế về nguồn lực
- Ngẫu nhiên hóa: Nhân tố nào quyết định trường học nào, chi nhánh nào... sẽ được tham gia chương trình vào năm nào?

Đặc tính của thiết kế ngẫu nhiên hóa theo giai đoạn

- **Phản thực:**

- Sau năm 1, những người/địa điểm bắt đầu tham gia chương trình ở năm 2, 3... sẽ được sử dụng làm nhóm đối chứng. Sau năm 2, những người bắt đầu tham gia chương trình ở năm 3, 4... sẽ được sử dụng làm nhóm đối chứng...

- **Yêu cầu dữ liệu:**

- Dữ liệu tham chiếu (tùy thuộc) và kết quả

- **Cân nhắc:**

- Theo thời gian, các nhóm đối chứng sẽ chuyển thành nhóm tham gia, do đó sẽ bị mất đi
- Tác động nhiều do nhóm những người chưa tham gia hiện nay dự kiến sẽ được tham gia trong tương lai

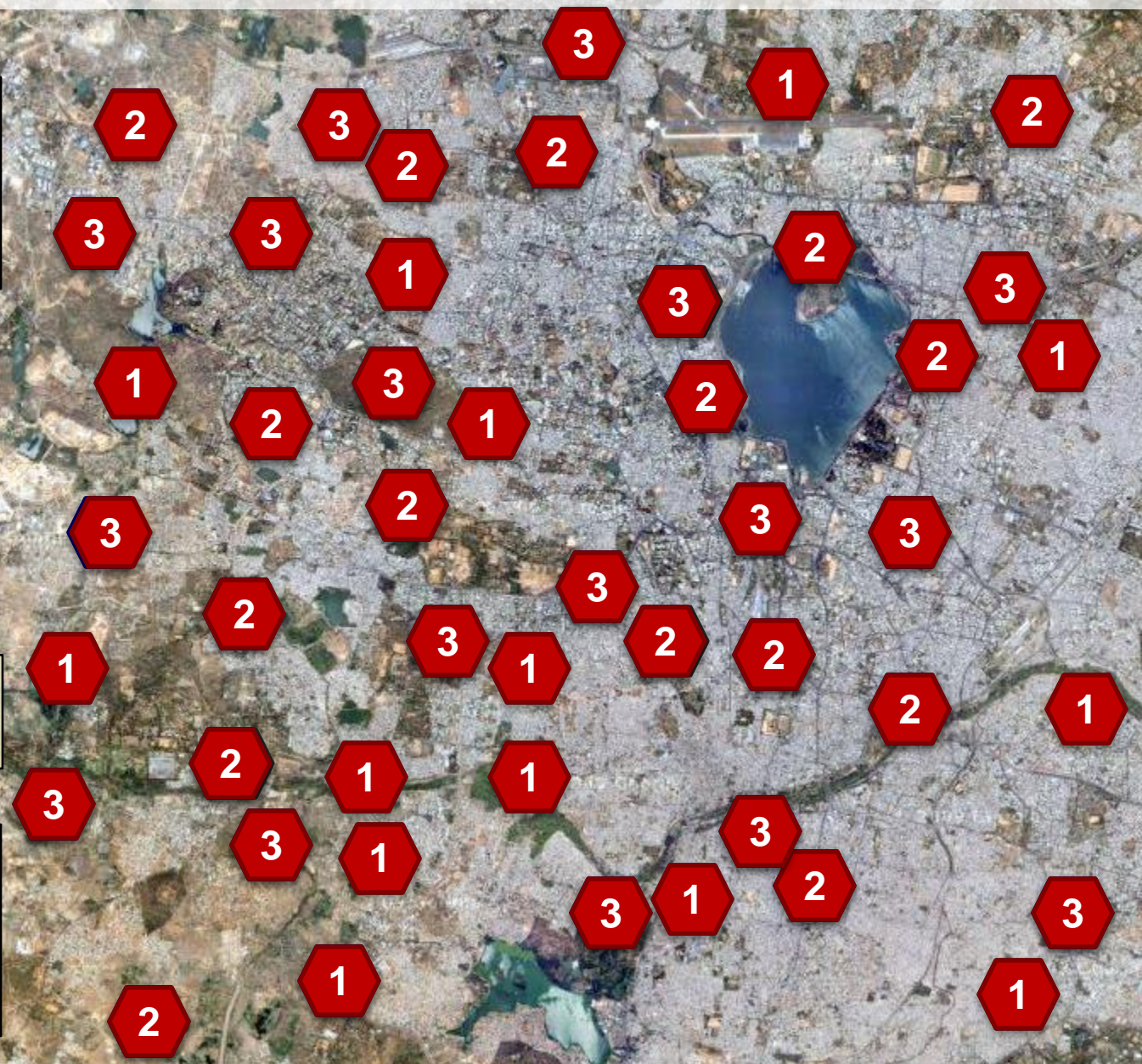
Thiết kế theo giai đoạn

Vòng 1
Tham gia: 1/3
Đôi chứng: 2/3

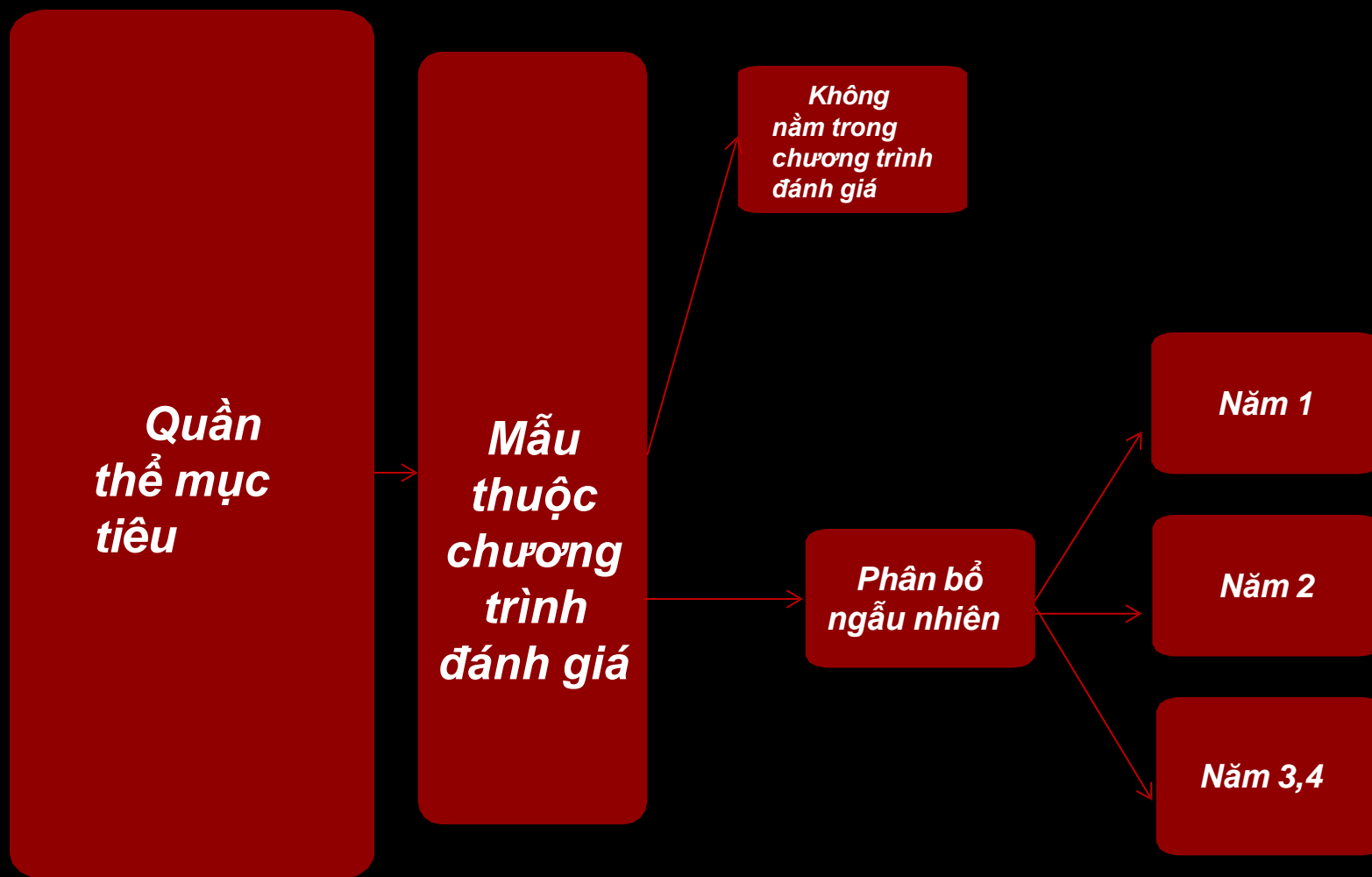
Vòng 2
Tham gia: 2/3
Đôi chứng: 1/3

Kết thúc đánh giá thử nghiệm

Vòng 3
Tham gia: 3/3
Đôi chứng: 0



RCTs | Thiết kế theo giai đoạn



Thiết kế theo giai đoạn – Đo lường tác động

- Tác động

- Sau năm 1: Trung bình của nhóm tham gia (những người được tham gia năm 1) trừ đi trung bình của nhóm sẽ tham gia vào năm 2 & 3.

$$\begin{aligned} E[Y_1 - Y_0 | D = 1] \\ = \bar{y}_{Y1} - \bar{y}_{Y2\&3} \end{aligned}$$

- Sau năm 2: Trung bình của nhóm tham gia năm 1 & 2 trừ đi trung bình của nhóm sẽ tham gia năm 3 & 4.

$$\begin{aligned} E[Y_1 - Y_0 | D = 1] \\ = \bar{y}_{Y1\&2} - \bar{y}_{Y3\&4} \end{aligned}$$

Thiết kế theo giai đoạn: Ưu và nhược điểm

- Ưu điểm
 - Mọi người cuối cùng sẽ được tham gia nên có động lực để giữ liên hệ
- Một số vấn đề
 - Có thể làm phức tạp vấn đề ước lượng tác động dài hạn
 - Theo thời gian, có thể mất nhóm đối chứng
 - Yêu cầu phải cân trọng với các khung thời gian thực hiện theo giai đoạn
 - Liệu kỳ vọng được tham gia có làm thay đổi hành vi hiện tại không?
 - Có thể bị nhiễu do tác động dự báo của những người sẽ được tham gia trong tương lai.

THIẾT KẾ KHUYẾN KHÍCH

Khuyến khích

- *Phải làm gì khi bạn không thể sử dụng phân bố ngẫu nhiên?*
 - Đôi khi phân bố ngẫu nhiên là không thực tế hay phi đạo đức
 - Nhưng có rất nhiều chương trình có tỷ lệ chấp nhận tham gia dưới 100%
 - Có thể ngẫu nhiên hóa việc khuyến khích tham gia chương trình

Khuyến khích là gì?

- Một thứ gì đó làm cho đối tượng có xu hướng chấp nhận đăng ký chương trình đánh giá
- Khuyến khích không phải là việc được tham gia hay hưởng lợi từ chương trình
- Chúng ta ước lượng tác động đối với nhóm đối tượng nào?
- Nghĩ về “ai sẽ phản ứng đối với các khuyến khích?”

RCTs | Thiết kế khuyến khích

- Yêu cầu dữ liệu:
 - Tham chiếu (nên có) và kết quả đối với nhóm được nhận khuyến khích và nhóm không
- Cân nhắc:
 - Khuyến khích cần thiết phải được thiết kế sao cho tăng xác suất đăng ký chương trình
 - Tác động can thiệp trung bình có thể khác giữa nhóm đăng ký chương trình do được khuyến khích và quần thể hay dân số nói chung.

Thiết kế khuyến khích

 Khuyến khích

 Không khuyến khích

So sánh nhóm được khuyến khích với nhóm không được khuyến khích

Các nhóm này tương quan nhau

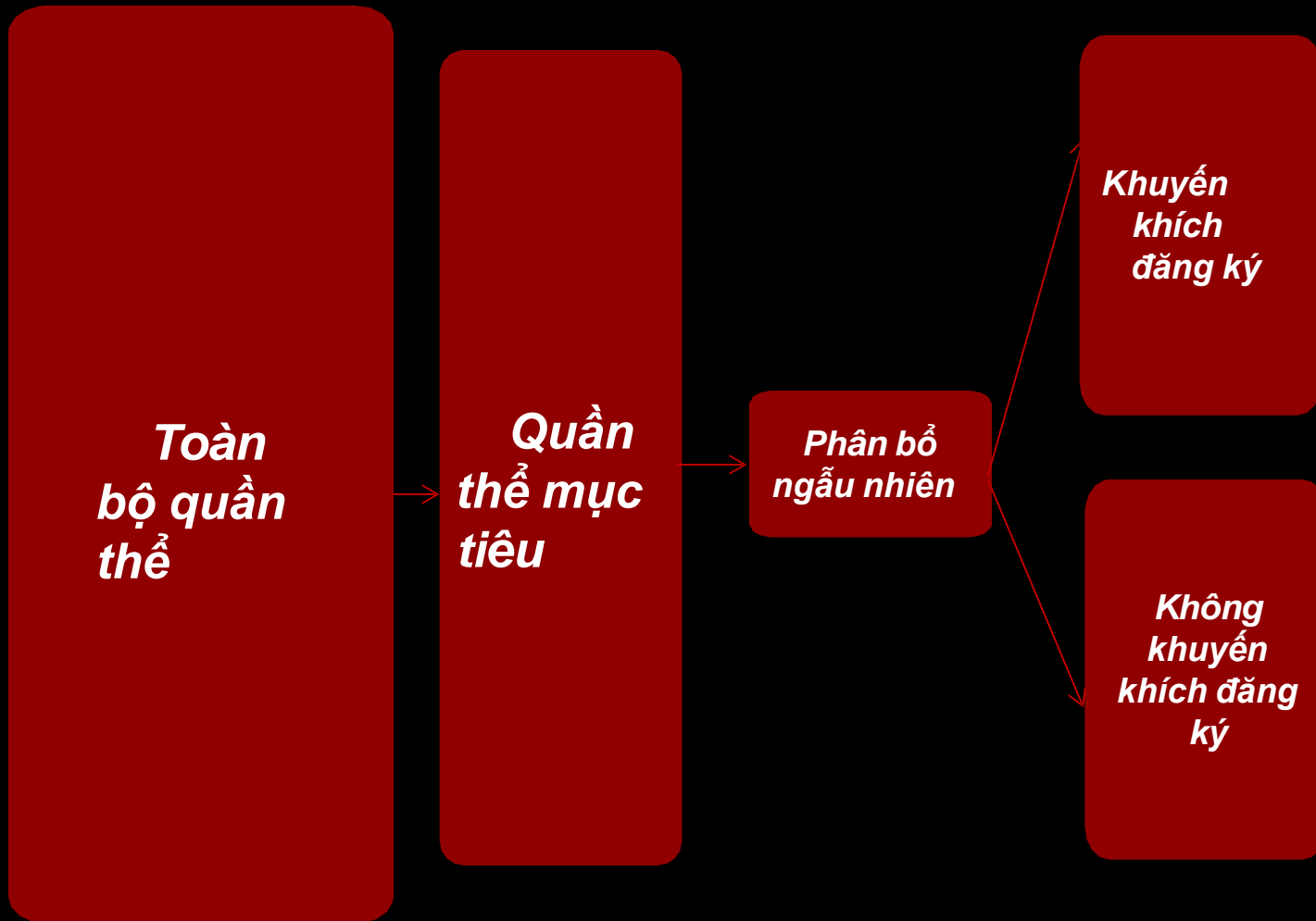
 Đăng ký tham gia

 Không tham gia

Không so sánh nhóm tham gia với nhóm không tham gia



RCTs | Thiết kế khuyến khích



Thiết kế khuyến khích – Đo lường tác động

- Tác động

- Trung bình nhóm tham gia (người nhận được khuyến khích) trừ đi trung bình nhóm không được khuyến khích.
 $E[Y_{-1} - Y_{-0} | D=1]$
- Chia cho phần trăm khác biệt về đăng ký chương trình.
- Quan trọng: Tác động này gọi là tác động can thiệp với nhóm muốn tham gia chương trình (intention to treat effect-ITE), không phải là tác động can thiệp trung bình (ATE).

$$\begin{aligned} & E[Y_{-1} - Y_{-0} | D = 1] \\ &= \frac{y_E - \bar{y}_C}{e_E - e_C} \end{aligned}$$

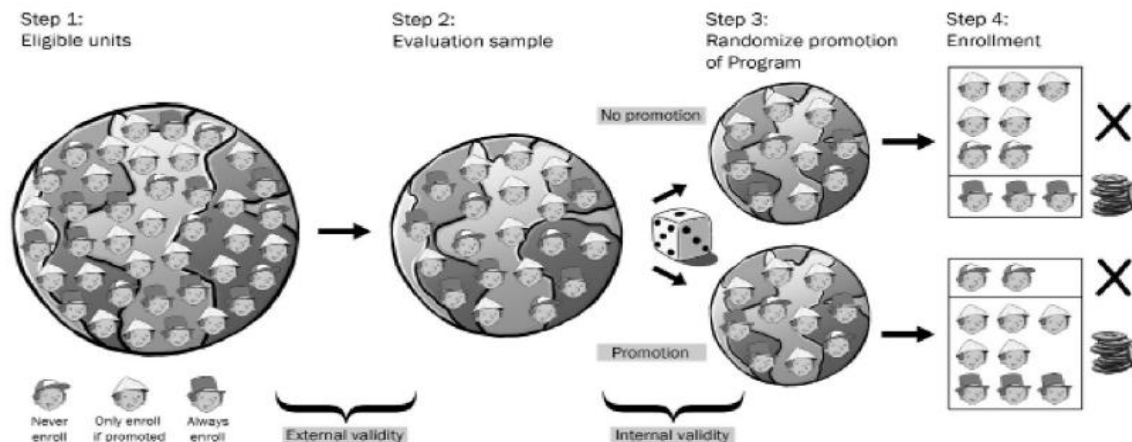
Thiết kế khuyến khích

Ví dụ tác động








- Bạn mở một chương trình đào tạo nghề cho tất cả các đối tượng và phân bổ ngẫu nhiên ở một số khu vực tại đó các cá nhân nhận được khuyến khích để tham gia.
- Bạn phát hiện ra là ở khu vực nhận được khuyến khích tỷ lệ dân số đăng ký tham gia là 25% cao hơn so với khu vực khác. Sau 1 năm, thu nhập trung bình ở khu vực nhận được khuyến khích là \$100, so với khu vực khác là \$80.
- Tác động can thiệp ước lượng được là:
$$ITE = (\$100 - \$80) / .25 = \$25$$

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 chọn 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)

Phương pháp ngẫu nhiên hóa - Ôn tập

Thiết kế	Hiệu quả nhất khi...	Lợi thế	Bất lợi
Bốc thăm giản đơn	<ul style="list-style-type: none">•Nhiều người muốn tham gia chương trình	<ul style="list-style-type: none">•Thông dụng•Dễ hiểu•Dễ thực hiện•Có thể thực hiện công khai	<ul style="list-style-type: none">•Nhóm đối chứng có thể không hợp tác•Rơi rớt mẫu giữa các nhóm

Phương pháp ngẫu nhiên hóa – Ôn tập

Thiết kế	Hiệu quả nhất khi...	Lợi thế	Bất lợi
Theo giai đoạn	<ul style="list-style-type: none">•Chương trình mở rộng theo thời gian•Sau cùng thì Sau mọi người đều được tham gia	<ul style="list-style-type: none">•Dễ hiểu•Dễ giải thích các hạn chế•Nhóm kiểm soát tuân thủ do kỳ vọng được tham gia trong tương lai	<ul style="list-style-type: none">•Kỳ vọng được tham gia trong tương lai có thể ảnh hưởng đến hành vi hiện tại•Khó đo lường tác động dài hạn

Phương pháp ngẫu nhiên hóa – Ôn tập

Thiết kế	Hiệu quả nhất khi...	Lợi thế	Bất lợi
Khuyến khích	<ul style="list-style-type: none">•Chương trình phải dễ ngỏ đối với tất cả các đối tượng•Khi việc chấp nhận chương trình còn thấp, nhưng có thể cải thiện nếu được khuyến khích	<ul style="list-style-type: none">•Có thể ngẫu nhiên hóa ở cấp độ cá nhân ngay cả khi chương trình không thực hiện ở cùng cấp độ đó	<ul style="list-style-type: none">•Đo lường tác động của chương trình đối với nhóm phản ứng với khuyến khích•Cần khuyến khích đủ lớn để cải thiện mức độ chấp nhận chương trình•Khuyến khích tự nó có thể có tác động trực tiếp lên kết quả

Bài giảng hôm nay

- Các ràng buộc thực tế
- Phương pháp ngẫu nhiên hóa
- Các dạng thực hiện đối với thử nghiệm tham gia-đối chứng đơn giản

Nhiều can thiệp cùng lúc

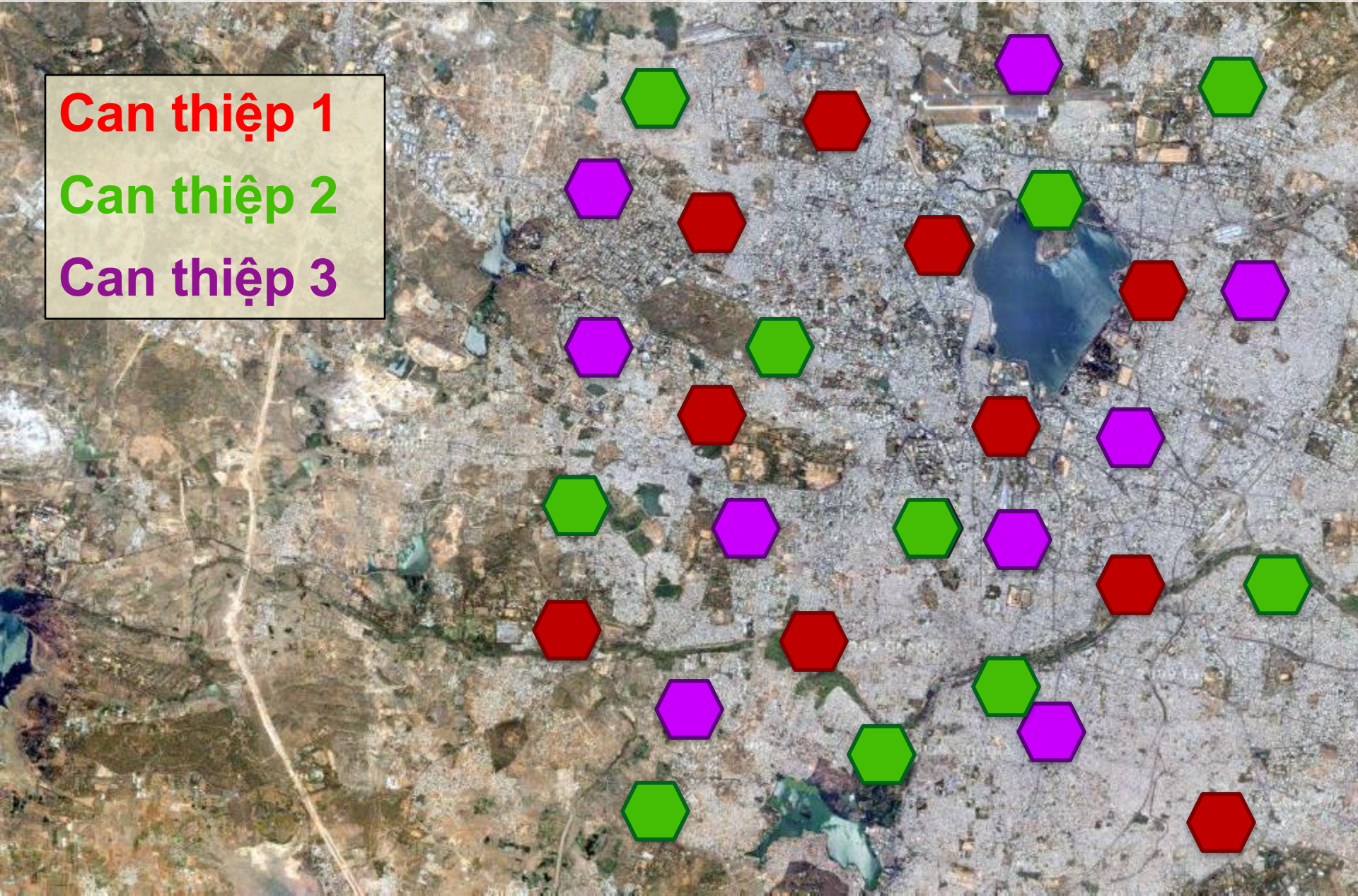
- Đôi khi câu hỏi trọng tâm là quyết định hình thức can thiệp tối ưu trong các hình thức can thiệp khác nhau cùng lúc
- Có thể ngẫu nhiên hóa các can thiệp đó
- Liệu điều này có giúp chúng ta biết được tác động của chỉ một can thiệp?
- Bạn có nhóm kiểm soát không?

Nhiều can thiệp cùng lúc

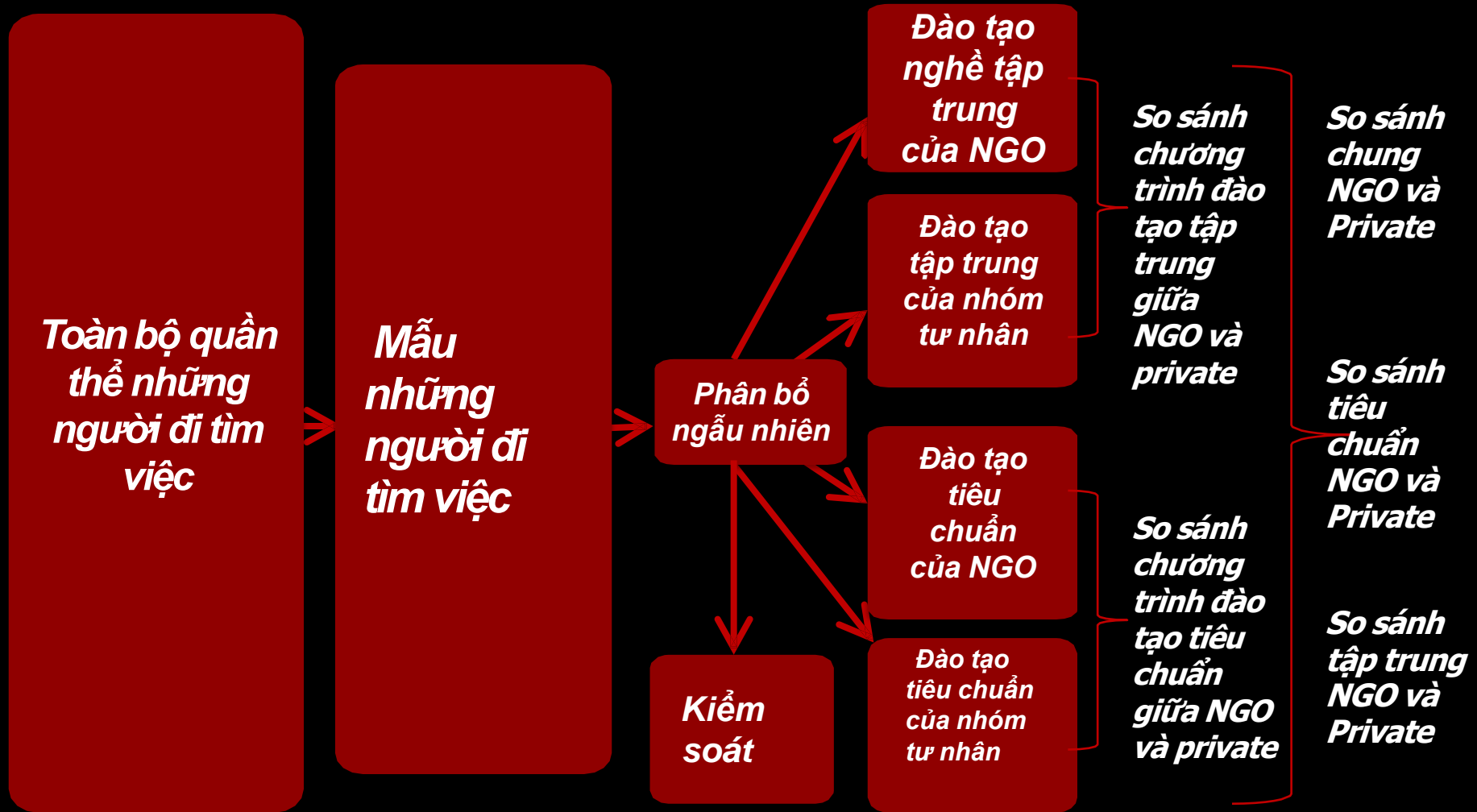
Can thiệp 1

Can thiệp 2

Can thiệp 3



RCTs | RCTs với nhiều nhánh



Can thiệp có nhiều cấu phần (cross-cutting treatment)

- Kiểm định các cấu phần khác nhau của cùng một can thiệp
- Kiểm định liệu các cấu phần bổ sung hay thay thế cho nhau
- Những cấu phần nào là hiệu quả nhất về mặt chi phí?
- Lợi thế: ưu điểm về mặt hoạt động, và giúp trả lời các câu hỏi liên quan, thay vì chỉ hỏi tác động.

Thiết kế nhân tố kết hợp (Factorial Design)

2 loại can thiệp

1. Chương trình đào tạo cho doanh nhân
2. Chương trình cho vay tín dụng vi mô

	Cho vay	Không cho vay
Đào tạo	Vay + đào tạo	Chỉ đào tạo
Không đào tạo	Chỉ vay	Không vay hay đào tạo

Can thiệp với nhiều mức độ

- Một số trường học được tham gia chương trình đầy đủ
- Tất cả trẻ em đều được uống thuốc
- Một số trường học chỉ được tham gia một phần
- 50% trẻ được uống thuốc
- Kiểm định tác động của trợ cấp và giá cả

Tổng hợp

- Vấn đề thực hiện rất quan trọng
- Có rất nhiều phương pháp tiếp cận để xử lý các vấn đề gặp phải khi thực hiện trên thực tế.
- Hình thức thiết kế và phân tích có sự khác nhau nhưng không nhiều, phụ thuộc vào loại hình thử nghiệm bạn thực hiện.
- Chú ý phân biệt các khái niệm ATE, ITE, TET, và LATE.