
ABDUL LATIF JAMEEL

Poverty Action Lab



J-PAL LAC TRADUCIENDO LA INVESTIGACIÓN EN ACCIÓN

¿Por qué y cómo aleatorizar?

Enrique Seira
ITAM & J-PAL
enrique.seira@itam.mx

Contexto del Curso

1. Introducción a las Evaluaciones de Impacto
2. Teoría de Cambio e Indicadores
- 3. Por qué y Cómo Aleatorizar**
4. Tamaño de Muestra y Cálculos de Poder
5. Inferencia Causal
6. Evaluaciones Experimentales y Políticas Públicas
7. Evidencia de Evaluaciones Experimentales
8. Evaluación Experimental de Principio a Fin

Esquema de la presentación

1. Antecedentes
2. Qué es una evaluación experimental
3. Por qué aleatorizar
4. Cómo aleatorizar

1. Antecedentes

Relaciones Causales

- El objetivo de muchos estudios en medicina, física, economía, ciencias sociales y biología es encontrar relaciones de **CAUSA-EFECTO** entre variables o eventos.
- El interés en **RELACIONES CAUSALES** está motivado por cuestiones de políticas públicas, consideraciones teóricas.

Preguntas Causales

- Algunos ejemplos de preguntas causales pueden ser:
 1. ¿Cuál es el efecto del *crédito* en la *productividad* de una empresa?
 2. ¿Cuál es el efecto de que las *mujeres sean elegidas* a cargos de elección popular en la *inversión pública y percepción de las mujeres como líderes*?
 3. ¿Cuál es el efecto de *proveer información* a los consumidores sobre sus *decisiones de compra*?
 4. ¿Cuál es el efecto del *encarcelamiento* en la *reincidencia criminal*?

Problema de Atribuir Causalidad

- Contestar estas preguntas no es trivial, aun cuando ya identificaron: (a) cuál es el “tratamiento” exactamente, (b) cuales son las variables que van a medir el resultado.
- Por ejemplo, para medir cuál es el efecto de que las mujeres sean elegidas a cargos de elección popular, si comparamos inversión pública en lugares con mujeres elegidas y no elegidas podemos estar comparando peras con manzanas. ¿Por qué?
- Diferente percepción de antemano para poder ser elegidas, diferentes grados de educación o ingreso de la población, mismas que a su vez afectan la inversión publica y la propensión a elegir mujeres.

Problema de Atribuir Causalidad

- **Problema de selección:** las personas que escogen una acción (e.g. ir a la escuela o solicitar un crédito) son en general *diferentes a las personas que no lo hacen*.
- Por ejemplo: ¿Qué variables llevan a unas personas a solicitar un crédito?
 - ¿En aprietos económicos? Por lo tanto, si no somos cuidadosos podríamos concluir que el crédito daña a las empresas
 - ¿Mejor expectativa del futuro? por lo tanto sobre-estimaríamos el efecto.
- Otro ejemplo: si comparamos personas que fueron encarceladas con otras que no, **¿que nos llevaría a concluir?**

¿Cómo Atribuir Causalidad?

Condiciones para Atribuir Causalidad (John Stuart Mill): A es la presunta causa del efecto B si:

1. A sucede antes que B.
2. Si los niveles de A varían, tiene que haber consecuentemente variaciones en los niveles de B.
3. **No hay otra explicación posible para el efecto B, que no sea la presunta causa A.**

¿Cómo Atribuir Causalidad?

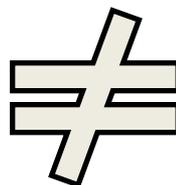
En esta clase, vamos a estudiar qué método nos asegura las 3 condiciones necesarias para poder atribuir causalidad.

- Si queremos **identificar el impacto verdadero** de un programa social necesitamos un buen **contrafactual**.

Impacto

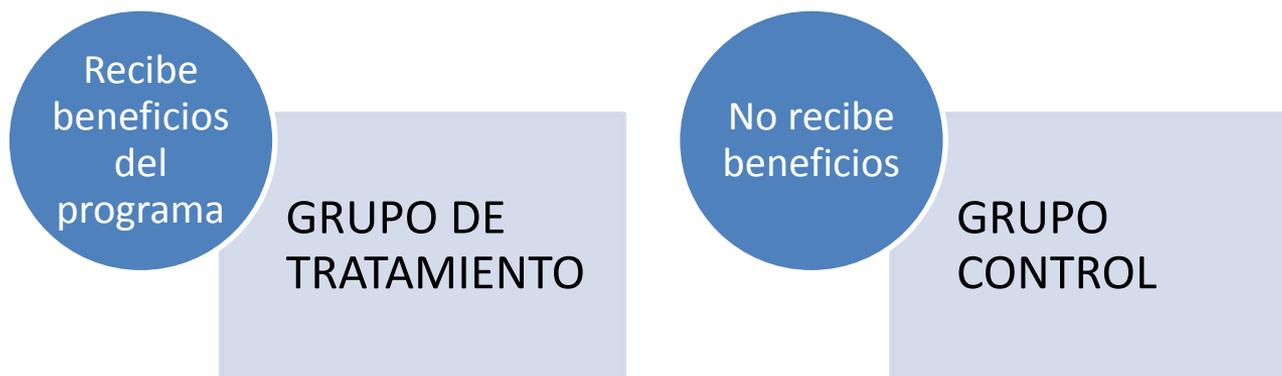
- El problema de la evaluación de impacto:

Evaluación típica: ¿cómo han cambiado las vidas de los clientes después del comienzo del programa?



Evaluación con grupo de control: ¿cómo han cambiado las vidas de los clientes por causa del programa?

11



Impacto

Para medir impacto necesitamos saber:

Qué ocurrió con el programa y
- Qué habría ocurrido sin el programa
(**contrafactual**)
=IMPACTO del programa

- Manteniendo todo lo demás constante (es decir, en lugares con la misma percepción de las mujeres, empresas idénticas con y sin el crédito, etc).
- Sin el contrafactual no tenemos contra que comparar los resultados!

2. Qué es una evaluación experimental

¿Qué es impacto?

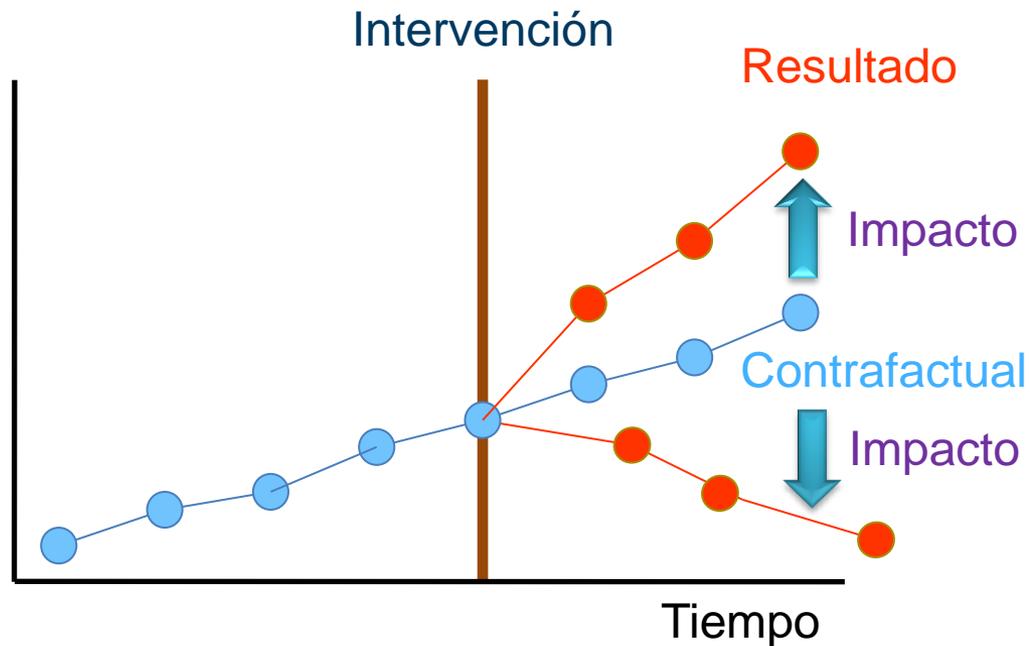
Resultados que los participantes del programa obtienen un tiempo después de participar en el programa

Resultados que esos mismos participantes hubieran obtenido en ese mismo momento si no habrían participado en el programa (hipotético)

IMPACTO

¿Cómo medir impacto?

Gráficamente:



Contrafactual

Es **hipotético**:

representa el estado del mundo que los participantes del programa habrían experimentado si no hubieran participado en el programa.

Problema: El contrafactual no se puede observar.

Solución: Necesitamos estimar o construir contrafactual.

Estimando el Contrafactual

- Se hace mediante la selección de un grupo de personas que **no** participaron en el programa.
- Este grupo se llama el **grupo de control** o el **grupo de comparación**.
- La forma de seleccionar este grupo es una **decisión clave** en el diseño de cualquier evaluación de impacto.

Seleccionando el grupo de control

- **Objetivo:** Seleccionar a un grupo que sea **exactamente igual** al grupo de participantes en todas las formas.
 - Sólo difieren en si reciben o no el programa.
 - Así podemos **atribuir** diferencias en los resultados entre el grupo de participantes y el grupo de comparación al programa (y no, a otros factores).

Métodos de evaluación de impacto

Evaluaciones experimentales o con grupo de control aleatorio (**muy confiables**)

Métodos no experimentales o cuasi experimentales

1. Antes y después
2. Diferencia simple en un mismo momento en el tiempo
3. Diferencias en diferencias (combinan (a) y (b))
4. Matching estadístico
5. Variables instrumentales
6. Regresión Discontinua

Métodos no experimentales

- Todos estos métodos tratan de hacer lo mismo:
 1. Estimar que hubiera pasado sin el programa por medio de **generar un grupo de control**.
 2. **Estimar la diferencia** entre el grupo de control y el grupo de tratamiento.
 3. La principal diferencia entre ellos es **cómo** estiman el contrafactual y **para cuales subpoblaciones**, i.e. quien es el grupo de control.

Métodos no experimentales

- Para definir el grupo de control los métodos no experimentales hacen supuestos, y lo que es peor: ¡supuestos que **no se pueden verificar!**
- Por ejemplo:
Antes y después, mismos individuos

Grupo de control: los tratados antes de recibir el tratamiento

Supuesto: que sin el tratamiento se hubieran quedado igual que antes

Amenaza: que haya tendencias, por ejemplo que la economía esté creciendo y se generen más ventas en las empresas.

Diferencia simple después del programa

Grupo de control: aquellos sin tratamiento, e.g. empresas sin crédito.

Supuesto: que las personas sin crédito son iguales a las empresas con crédito.

Métodos no experimentales

Matching:

Grupo de control: aquellos con iguales características observables

Supuesto: que observamos todas las características que podrían afectar el resultado

Variables instrumentales:

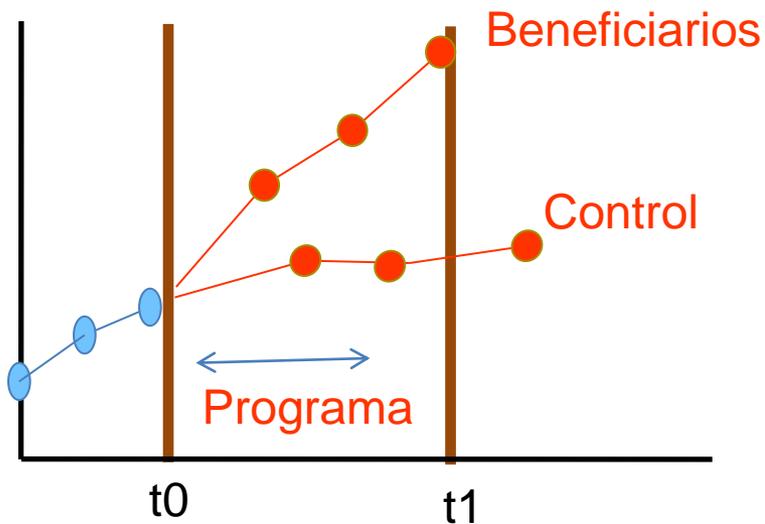
Idea: usar una variable que afecte la probabilidad de que reciba tratamiento pero no directamente el resultado. Por ejemplo, que a unos les llegó una invitación y a otros no.

Grupo de control: aquellos que no les llegó la invitación

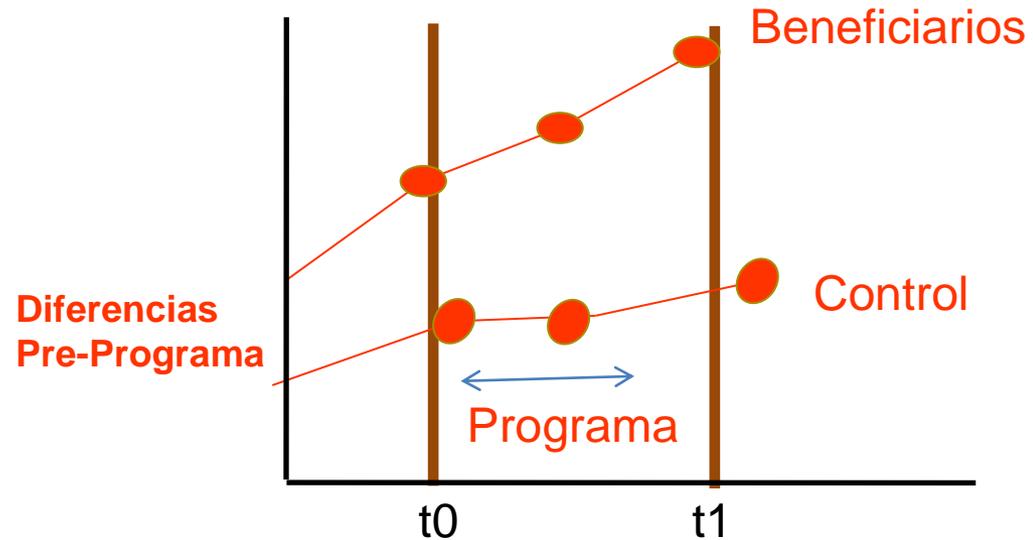
Supuesto: que la invitación no se mandó con base en el resultado esperado, e.g. a los mejores estudiantes.

El problema de las evaluaciones no experimentales o cuasi-experimentales:

Los grupos a comparar pueden no ser comparables



Experimental



No experimental o cuasi-

¿Y el método experimental?

Un modelo básico

Tome una muestra de postulantes a un programa

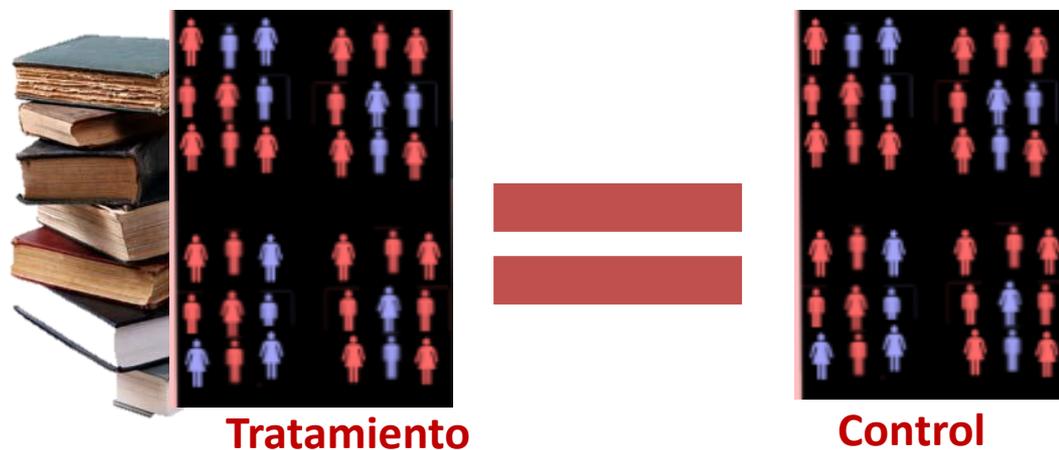
Asígnelos de forma aleatoria a:

- Un **grupo de tratamiento**: recibe el tratamiento
- Un **grupo de control**: no recibe el tratamiento durante el período de evaluación

3. ¿Por qué aleatorizar?

¿Por qué aleatorizar?

Al tener una asignación aleatoria, los grupos no difieren de ninguna manera sistemática.



Cualquier diferencia al finalizar al programa puede ser atribuida al **programa**.

Ventaja de las evaluaciones experimentales

- Se **elimina cualquier sesgo** que puede haber en una evaluación no experimental.
- La aleatorización hace que los grupos sean comparables en variables **observables** y **no observables**.
- No es necesario hacer demasiados supuestos, como en las evaluaciones cuasi-experimentales.

Evaluación de Programa de Tutorías en Chile: Comparación inicial entre Tratamiento y Control

Variables	Grupo de Tratamiento	Grupo de Control	Diferencia
Años Educación de la Madre	10.65	10.35	0.30 (0.31)
Ingresos de la Familia	261,615.80	248,089.60	13526.20 (16476.20)
Puntaje: Lenguaje	257.31	251.15	6.16 (4.82)
Puntaje: Matemáticas	249.98	243.33	6.65 (5.10)
Tamaño Escuela	71.70	74.30	-2.60 (7.70)

Pasos claves para diseñar un experimento

- 1. Diseñar** el estudio cuidadosamente (definir la subpoblación de interés)
- 2. Asignar aleatoriamente** las unidades a tratamiento o control.
- 3. Recolectar datos de línea de base** (no es estrictamente necesaria pero muy recomendable, ayuda a verificar balance, tamaño muestra, interacciones.)

Es muy importante el diseño de la Encuesta (especificación y medición de las variables de interés).

- 4. Verificar** que la asignación generó grupos similares

Pasos claves para diseñar un experimento

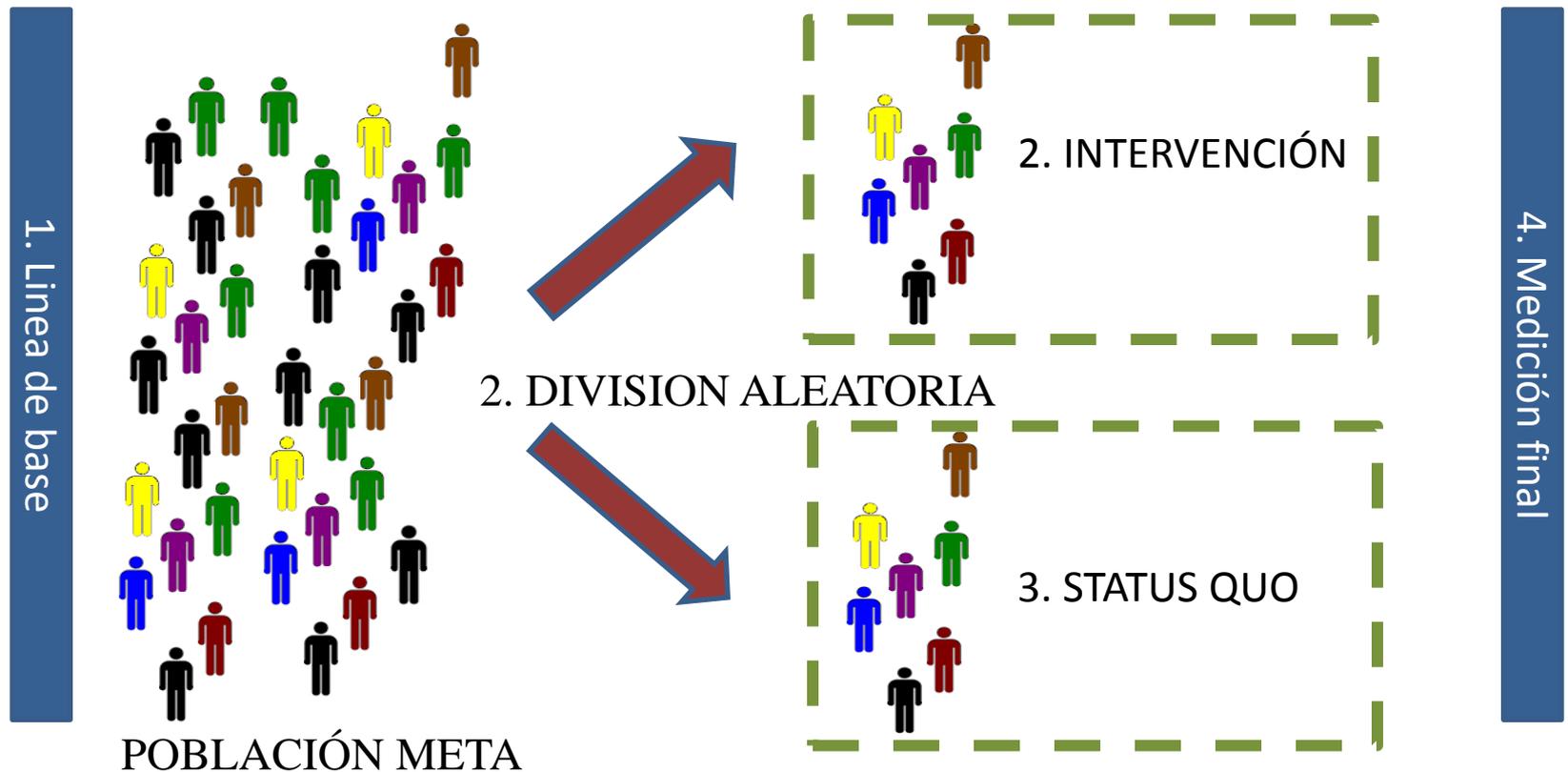
5. **Monitorear** el proceso para asegurar la integridad del experimento (que no se les dé el programa a los de control y sí se les dé a los de tratamiento)

6. **Recolectar datos post-intervención** tanto para el grupo de tratamiento como para el grupo control

7. Estimar **los impactos** del programa comparando los resultados promedio del grupo de tratamiento vs. los resultados promedio del grupo de control

8. Determinar si los impactos son **estadísticamente** significativos.

Pasos claves para realizar un experimento



¿Hay diferencia en la práctica entre aleatorizar o no?

Programa de Lectura en India.

- Implementado por **Pratham**, una ONG en India.
- El programa proporcionó **tutores** para ayudar a niños con dificultades en la escuela.
- El programa de tutores se implementó en escuelas primarias públicas en 2002-2003.
- Los maestros de estas escuelas decidieron cuáles niños (los de peor rendimiento) se educarían con los tutores.



Programa de Lectura: Indicadores de Resultados

A los niños se les aplican pruebas de lenguaje y matemáticas al comienzo del año escolar (pre-test) y al final del año (post-test)

PREGUNTA: ¿Cómo determinamos el impacto del programa?

Métodos para estimar impacto

Exploremos diferentes formas de calcular los impactos utilizando los datos de escuelas que obtuvieron un tutor

1. Pre – Post (Antes vs. Después)
2. Diferencia simple
3. Diferencias en Diferencias
4. Otros métodos no experimentales
5. Experimento Aleatorio

1. Pre-post (Antes vs. Después)

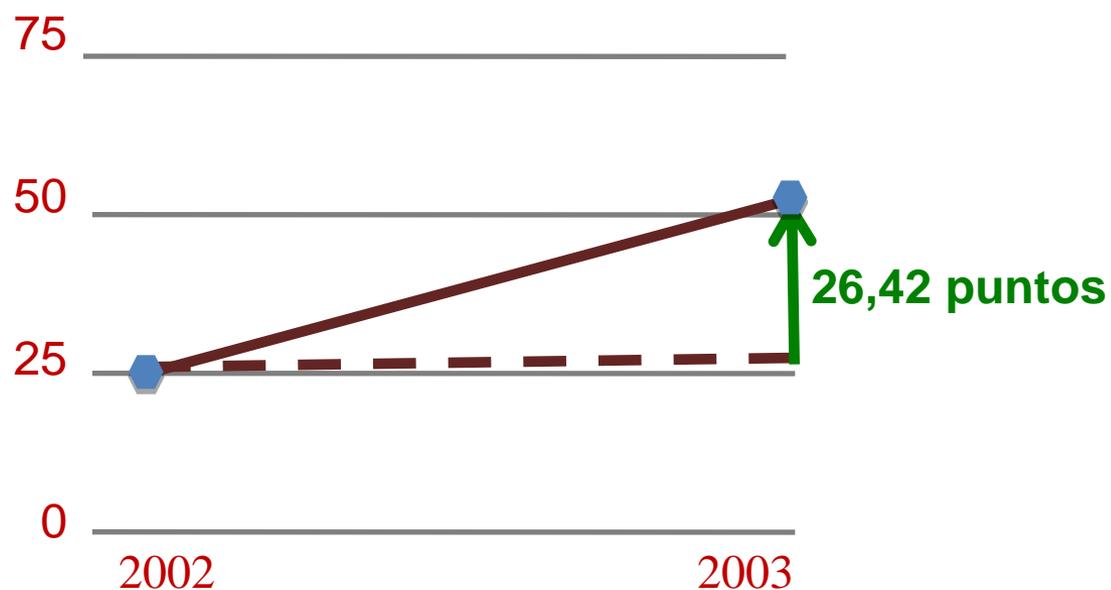
Puntaje promedio post-test para los niños con un tutores	51,22
Puntaje promedio pre-test para los niños con un tutores	24,80
Diferencia	26,42

¿Bajo qué condiciones se puede considerar esta diferencia (26,42) como el impacto del programa de tutores?

¿Que habría ocurrido sin el balsakhi?

Método 1: Antes vs. después

Impacto = 26.42 puntos?



2. Diferencia simple

Compare el **puntaje promedio** de...



con el
puntaje
promedio
de...



Niños que **recibieron** el
tutor

Niños que **no recibieron**
el tutor

2. Diferencia simple (post test)

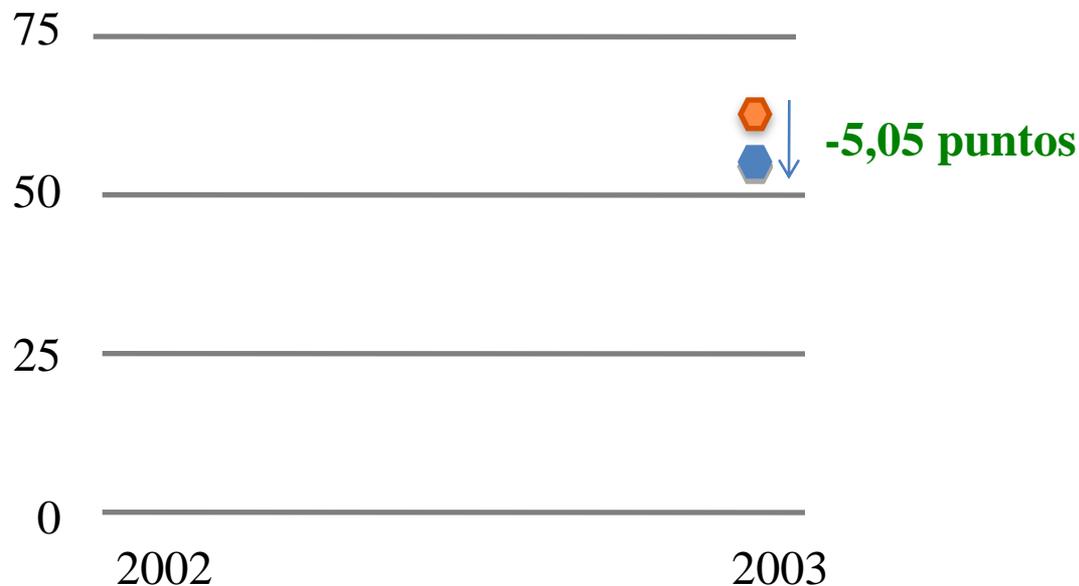
Puntaje promedio para los niños con un balsakhi	51,22
Puntaje promedio para los niños sin un balsakhi	56,27
Diferencia	-5,05

¿Bajo qué condiciones se puede considerar esta diferencia (-5,05) como el impacto del programa de tutores?

¿Que habría ocurrido sin el tutor?

Método 2: Comparación simple

Impacto = -5.05 puntajes?



3. Diferencias en Diferencias

Compare el cambio en puntaje promedio de...



con el cambio
en puntaje
promedio
de...



Niños que recibieron tutor

Niños que no recibieron
tutor

3. Diferencias en Diferencias

	Pre-test	Post-test	Diferencia
Puntaje promedio para los niños con un tutor	24,80	51,22	26,42
Puntaje promedio para los niños sin un tutor	36,67	56,27	19,60
Diferencia			6,82

¿Bajo qué condiciones se puede considerar esta diferencia (6,82) como el impacto del programa de tutores?

4. Experimento Aleatorio

Suponga que evaluamos el programa de tutores usando un experimento aleatorio

- PREGUNTA #1: ¿Qué implica eso?
¿Cómo lo hacemos?
- PREGUNTA #2: ¿Cuál sería la ventaja de utilizar este método para evaluar el impacto del programa de tutores?

Impacto del Programa de Tutores

Método	Impacto Estimado
(1) Pre-post	26.42*
(2) Diferencia simple	-5.05*
(3) Diferencias en diferencias	6.82*
(4) Regresión	1.92
(5) Experimento Aleatorio	5.87*

*: Significativo estadísticamente al nivel de 5%

Conclusión: ¡El método que utilizamos importa!

4. ¿Cómo Aleatorizar?

Cómo aleatorizar: estrategias

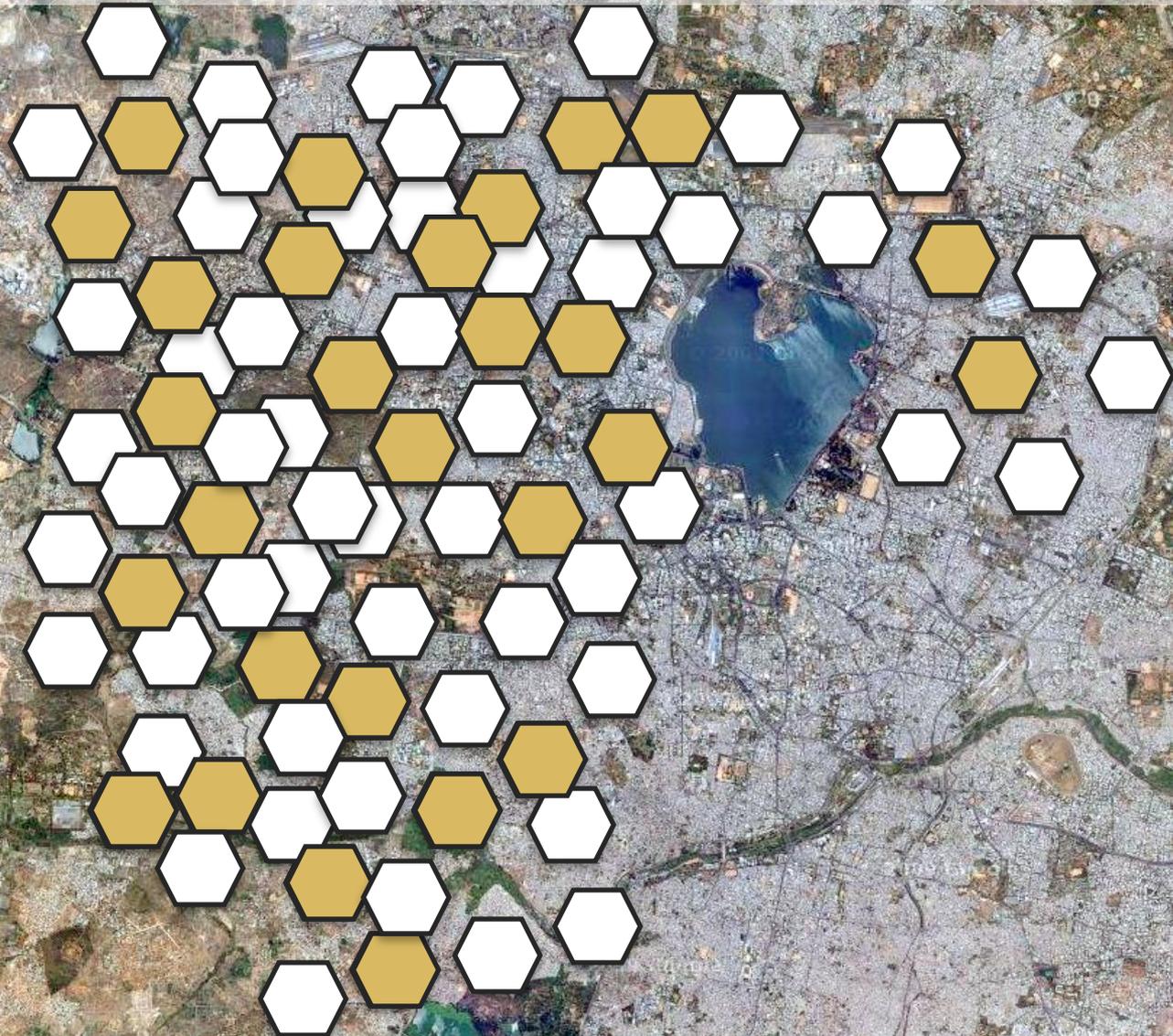
En la primera parte platicamos de porqué aleatorizar trae grandes ventajas.

Ahora platicaremos de como hacerlo. Tanto desde el punto de vista operativo, como de la mecanica en sí.

Antes de hacer esto quisiera detenerme a hacer una distinción entre muestra aleatoria vs tratamiento aleatorio.

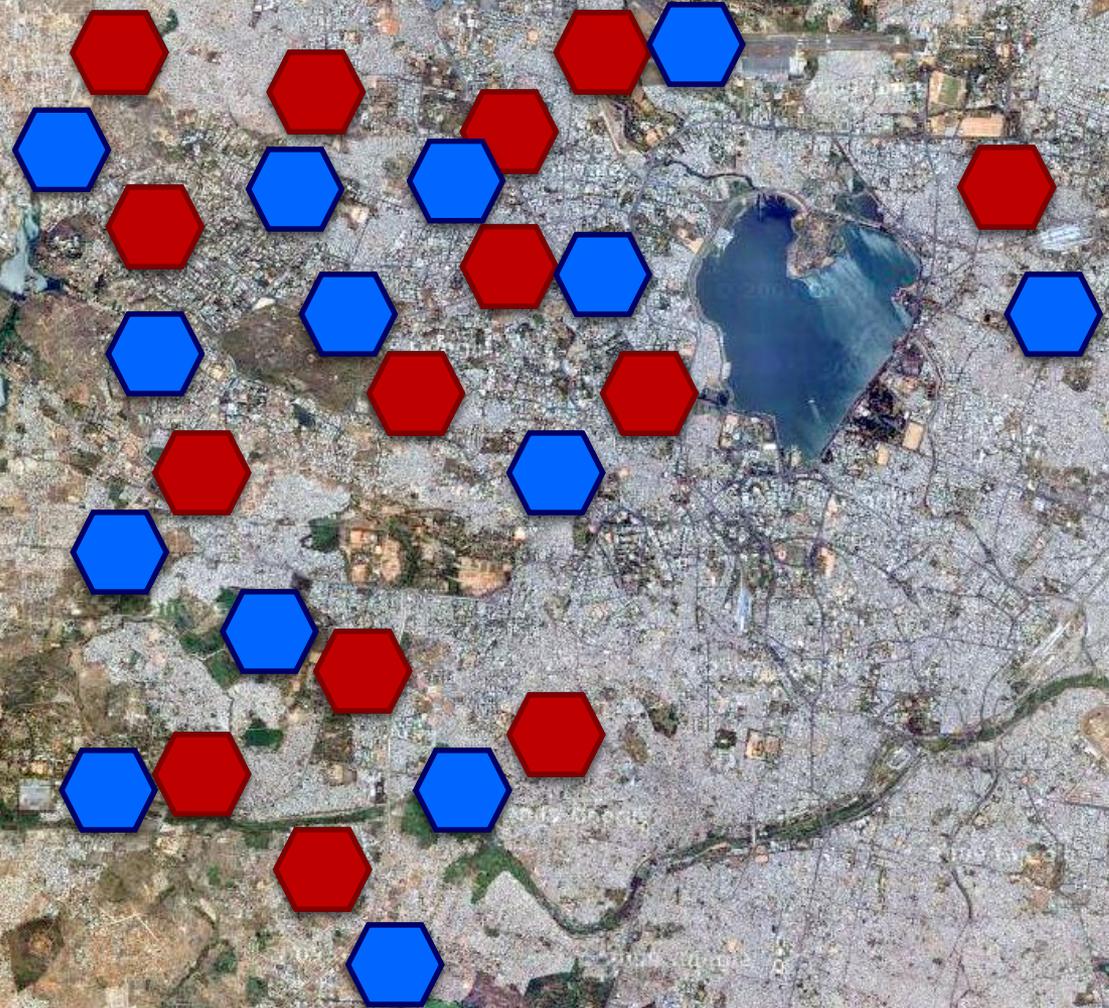
Muestreo aleatorio vs. Asignación aleatoria

***Muestra aleatoria**
es para
representar a la
población*



Muestreo aleatorio vs. Asignación aleatoria

*Asignar aleatoriamente a **tratamiento** y control es para evitar sesgos de selección*



Cómo aleatorizar: estrategias

Típicamente la primera reacción al mencionar que se propone aleatorizar es la resistencia: por motivos “éticos” o políticos.

En gran parte de los casos randomizar es más justo que otra metodología de selección y no es tan complicado como se cree.

En proyectos piloto: es común que antes de implementar un programa a gran escala se haga un piloto. Para poder evaluarlo bien es conveniente ***escoger aleatoriamente a las unidades en donde se implementará.***

En Oportunidades antes de expandirlo a decenas de miles de localidades se implementó en 200+ *localidades* piloto seleccionadas aleatoriamente dejando 200+ como control. Al probar que funcionaba se expandió. Este es un factor que contribuye a la continuación del programa.

Cómo aleatorizar: estrategias

Método de exceso de demanda: si hay mayor demanda de la que se puede atender, forzosamente tenemos que dejar fuera a algunos. Si **escogemos aleatoriamente quien entra** -además de ser transparente y justo- nos permite tener un grupo de control.

En Colombia se implementó un sistema de Vouchers para escuela. Como había exceso de demanda por las escuelas se escogió por lotería cual *alumno* entraba y cual no.

Randomización en punto de corte: Consiste en dejar entrar a personas que apenas se quedaron afuera. Esto tiene la ventaja de sólo cambiar poco el método de selección en el margen.

Karlan y Zinman aleatoriamente le dieron crédito a *personas* que habían sido rechazadas pero estaban cerca del punto de corte del banco. Esto permitió medir si el banco era demasiado estricto en su política.

Cómo aleatorizar: estrategias

Orden de implementación aleatorio: consiste en ir implementando el programa para diferentes unidades en diferentes momentos, pero donde el orden es aleatorio.

En un proyecto de desparasitación de niños en *escuelas*, se escogieron 75 escuelas para recibir la medicina, pero 25 entraron el 1er año, 25 el 2do y las restantes el 3ro.

Aleatorización por sub-grupos:

Un ejemplo lo explica mejor:

En el programa de Balsakhi en unas escuelas se les dió tutor a los niños de 3er año y en otras a los de 4to año. Cuál año escoger en cada escuela se determinó de forma aleatoria.

De esta forma todas las escuelas fueron atendidas al mismo tiempo.

Cómo aleatorizar: estrategias

Diseño de invitación aleatoria: consiste en aleatorizar quien recibe promoción/invitación especial para recibir el tratamiento (que no es lo mismo que aleatorizar el tratamiento).

Puede decirse que este método es el menos intervencionista, porque no le prohíbe a nadie entrar al programa y tampoco a nadie lo fuerza a tomarlo.

Sin embargo tiene algunas desventajas: lo que se estudia no es en sí el efecto del programa sino el efecto del programa inducido por la invitación.

Se puede estimar el efecto del programa en los que fueron convencidos por la invitación, no en la población en general.

Si el efecto de la invitación es bajo se requieren muestras muy grandes.

La población a estudiar

Es importante notar que estrictamente las conclusiones tienen validez sólo para la población que se aleatorizó. Dependiendo del contexto se podrá o no generalizar el resultado fuera de la muestra.

Por eso, para obtener resultados para la población elegible, algunas evaluaciones hacen la asignación aleatoria para las personas que *ya pasaron los criterios de elegibilidad y los filtros del programa*.

Si la pregunta es ¿cómo afectaría el programa a los no elegibles –por ejemplo porque se quiere cambiar la elegibilidad-- entonces la evaluación se debe realizar sobre esta población

Cómo aleatorizar: mecánica

Una vez decidida la estrategia de aleatorización y la unidad de aleatorización (por verse), la mecánica es muy fácil.

Se tiene una lista de las unidades en Excel o en Stata u otro programa, donde cada renglón corresponde a una unidad: e.g. una escuela, una persona, un municipio, una sucursal, etc.

En excel tecleas “=RAND()” y te da un número entre 0 y 1 uniformes. En Stata tecleas =uniform().

Si quieres seleccionar $\frac{1}{2}$ de la muestra, creas una variable que sea=1 si la variable random es $< .5$.

Cómo aleatorizar: mecánica

Si no es posible tener una lista podemos usar otros métodos menos rigurosos como:

Sacar papeles de un bote

Usar la primera letra del nombre

First-come-first served

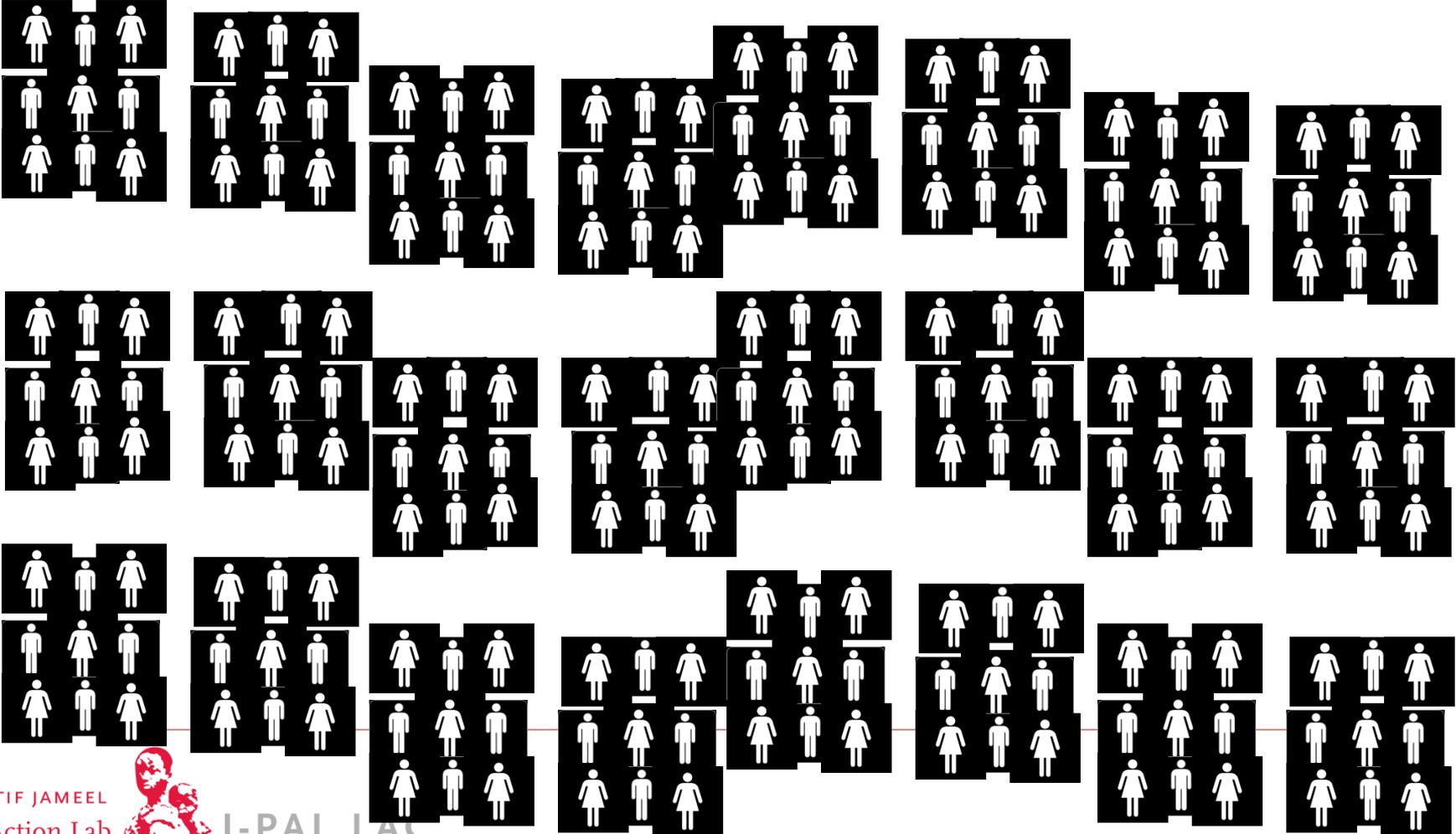


Nótese que esto puede tener problemas:

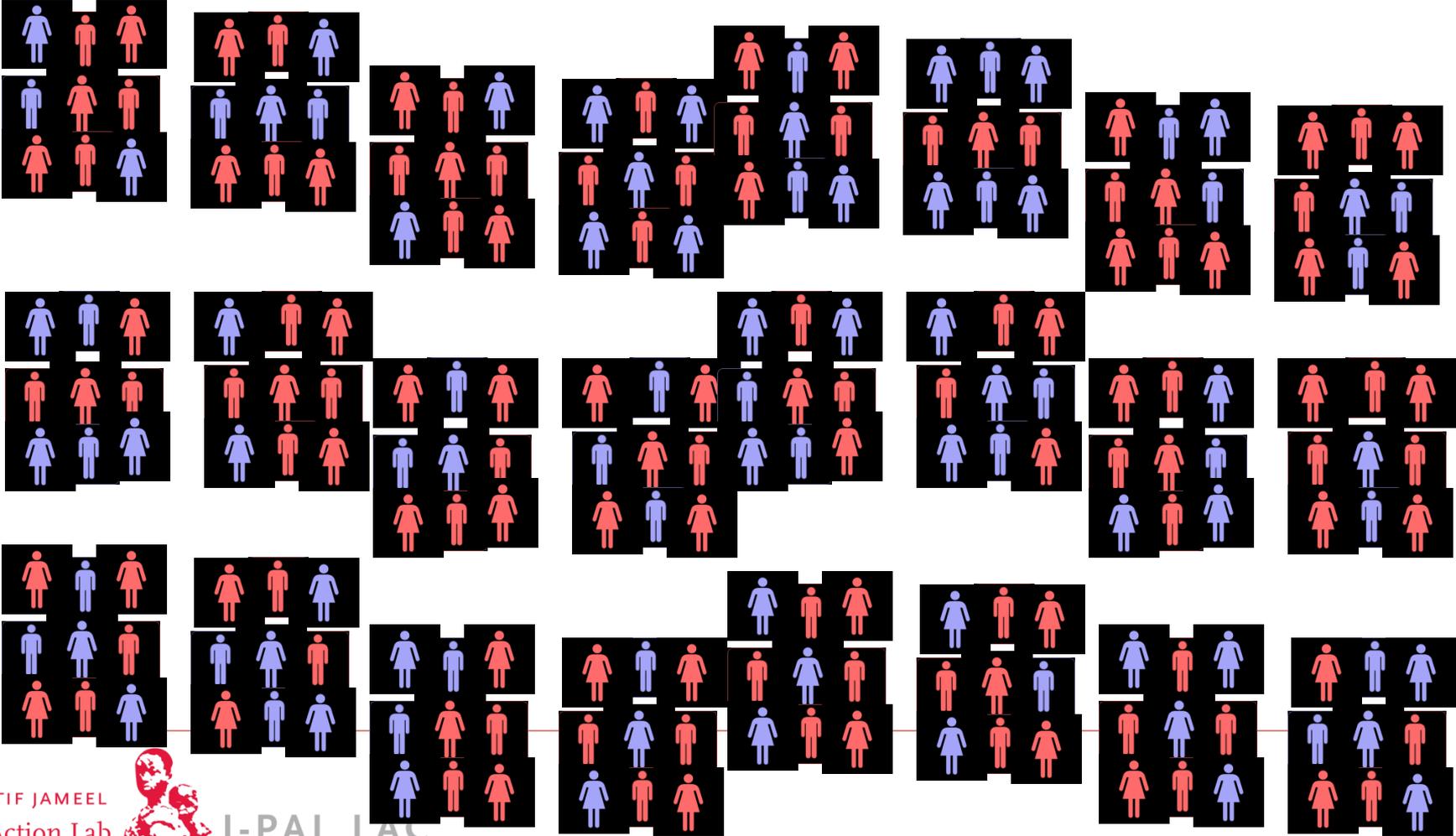
¿Ejemplos?

Siempre verifique que el resultado final sí parezca aleatorio (e.g. pruebas de medias o de distribuciones entre los grupos). Si no lo son entonces vuelva a asignar aleatoriamente.

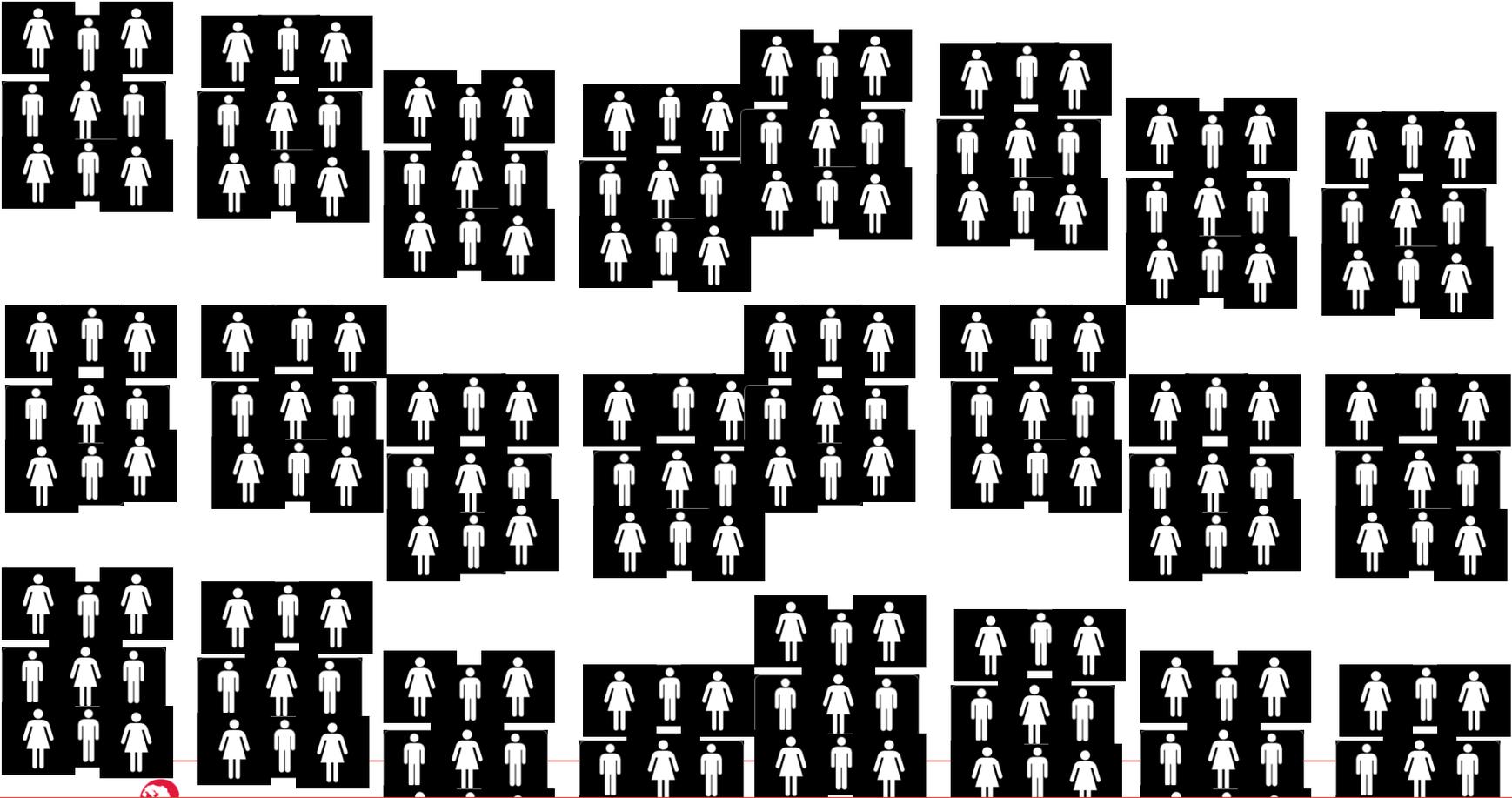
Unidad de aleatorización: ¿individual?



Unidad de aleatorización: ¿individual?

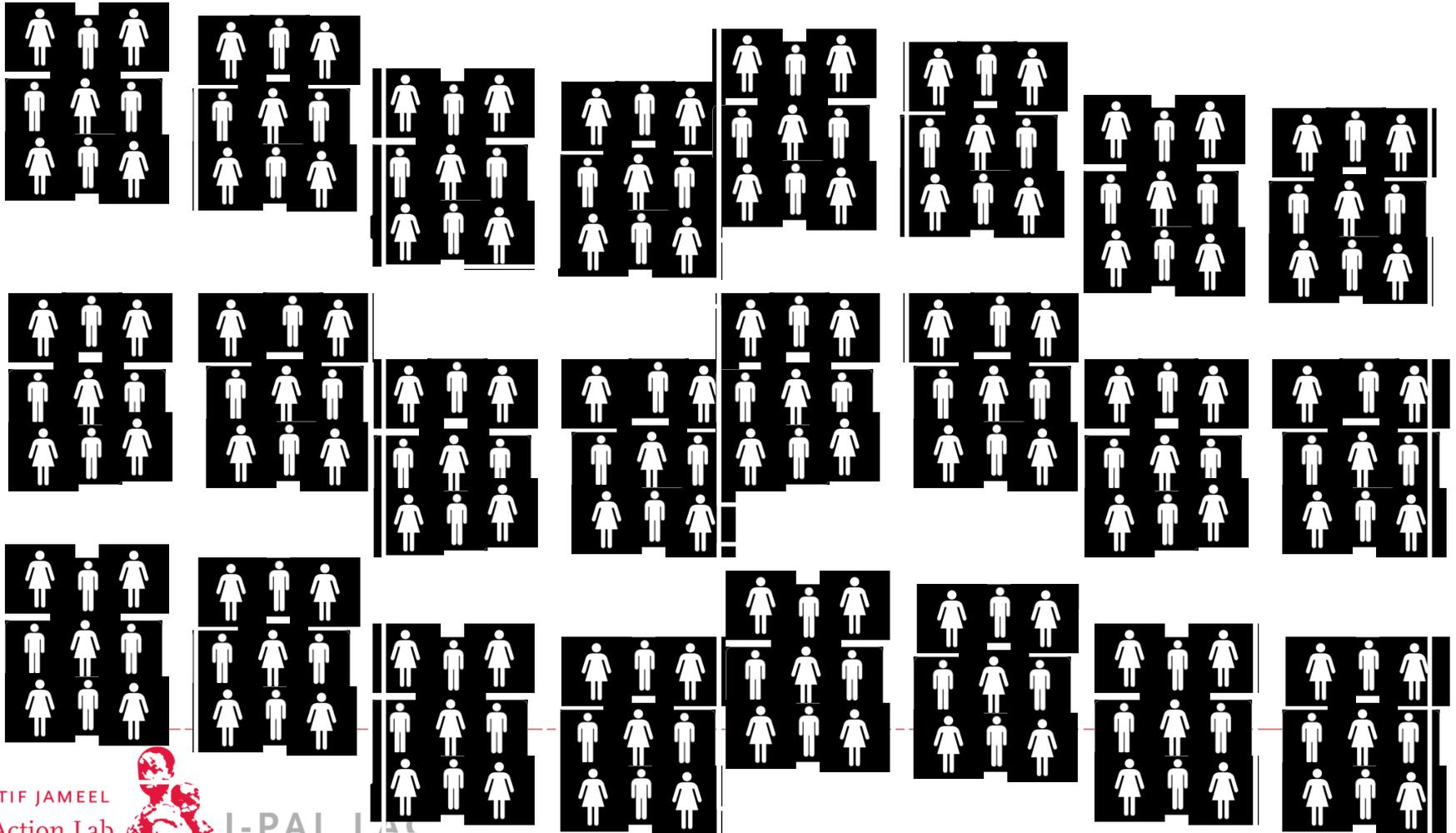


Unidad de aleatorización: ¿unidades agrupadas?

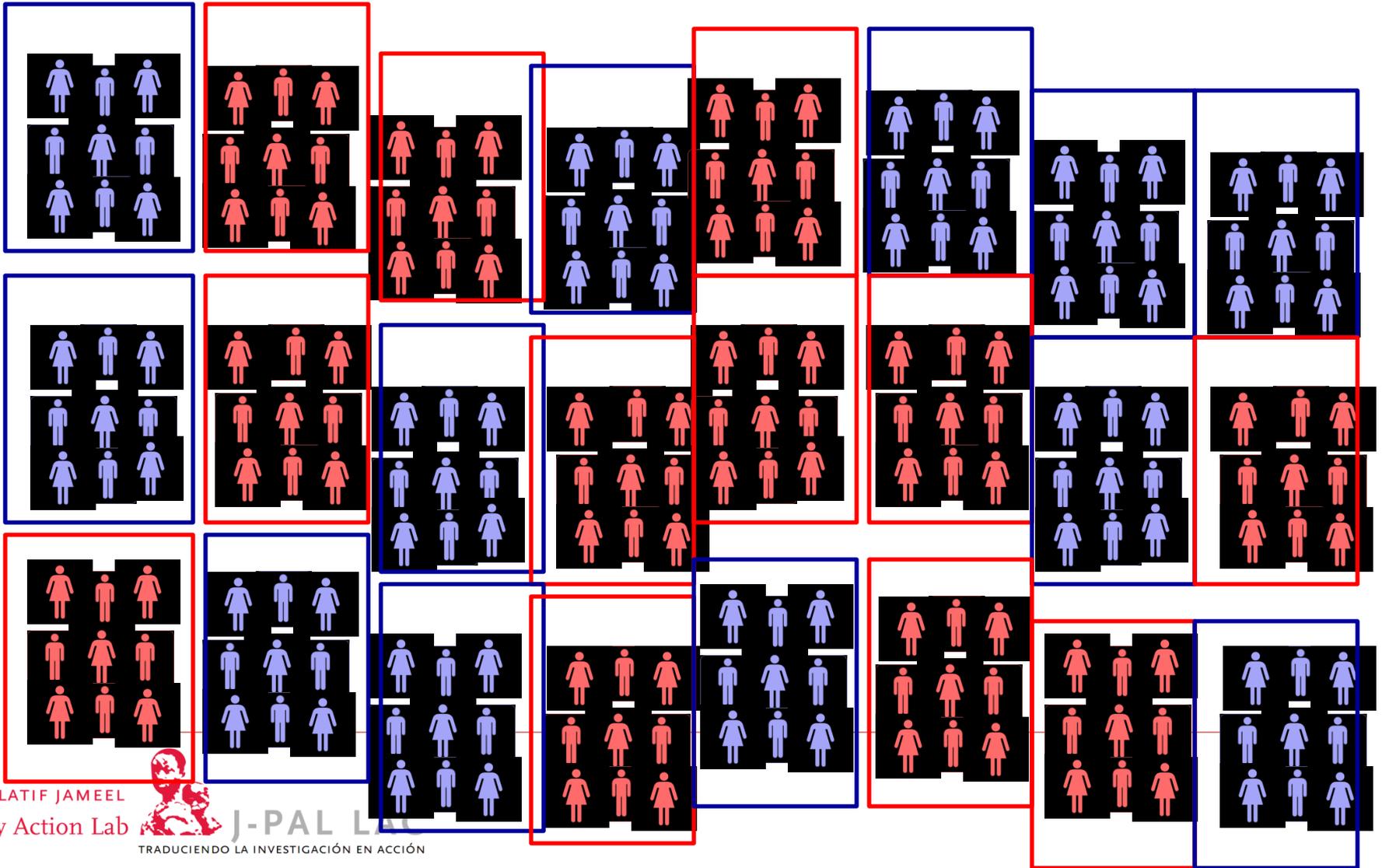


“Grupos de personas”: Estudio aleatorio en unidades agrupadas

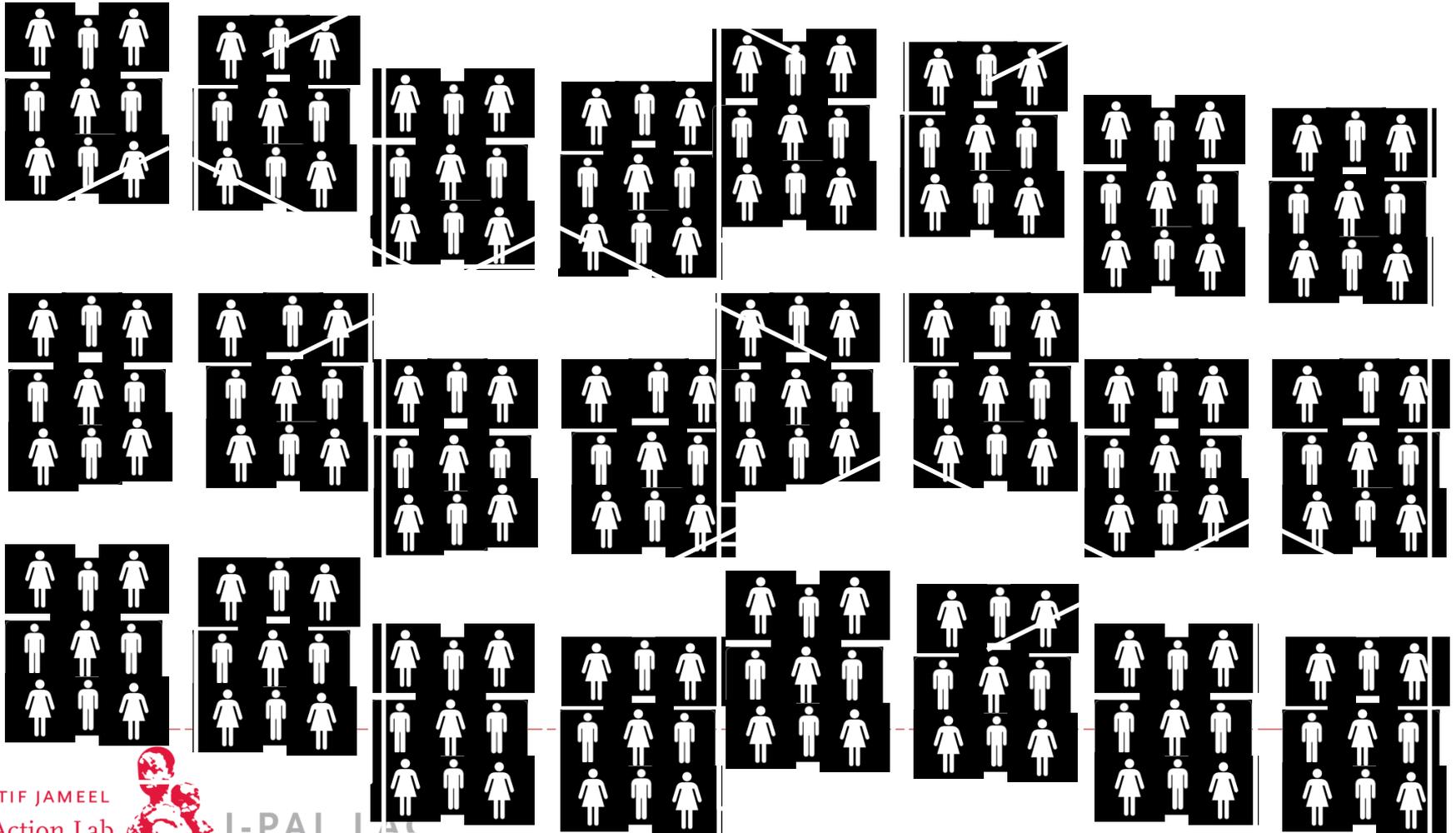
Unidad de aleatorización: ¿clase?



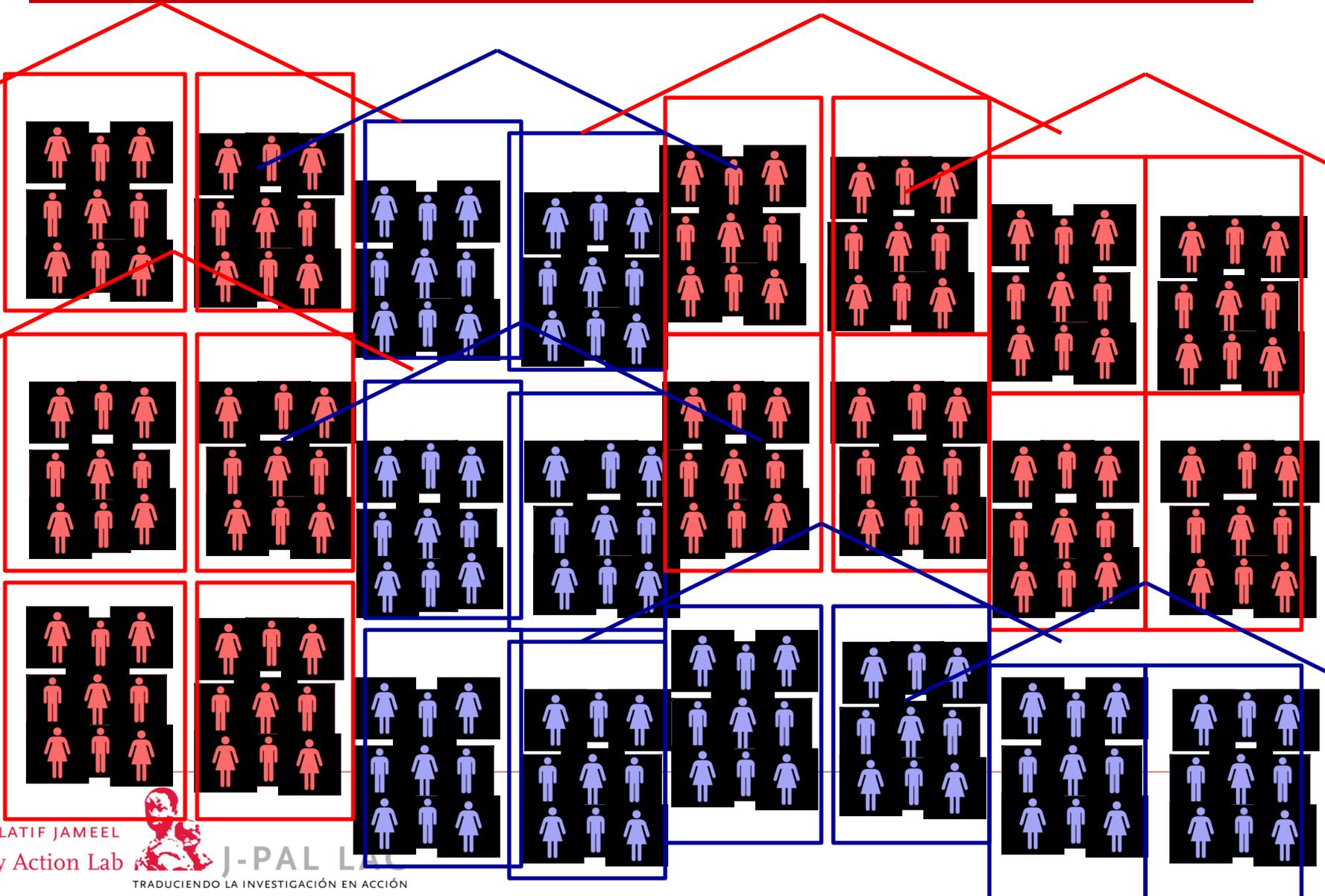
Unidad de aleatorización: ¿clase?



Unidad de aleatorización: ¿escuela?



Unidad de aleatorización: ¿escuela?



Unidad de aleatorización

¿A que nivel aleatorizamos: Alumno, salón, año escolar, escuela, municipio? Depende de varias cosas:

Restricciones institucionales: puede ser percibido como injusto que un alumno tenga libros de texto gratis y el de enseguida no, por lo que podemos querer, por ejemplo, hacerlo a nivel escuela. Puede ser también que por la forma de operar el programa sea más fácil hacerlo por zonas.

Restricciones de contagio/contaminación: usando el ejemplo anterior podemos no quererlo hacer a nivel alumno porque podría compartir el libro con el vecino, contaminando al control

A que nivel se implementa el programa en la vida real.

Unidad de aleatorización

Si esperamos que el efecto total se refleje en unidades grandes, deberíamos aleatorizar a este nivel de unidades grandes.

Por ejemplo un programa de apoyo a proveedores le ayuda a varias empresas de una localidad, por lo que habría que aleatorizar localidades .

De esta forma no solo evitamos el problema de contaminación, sino que lo estudiamos en sí.

A veces por razones operativas o políticas nos vemos forzados a aleatorizar a niveles mayores.

Por ejemplo: si en un programa de crédito tratan a algunas personas de forma diferente algunos se podrían molestar.

Unidad de aleatorización

Noten que si las personas tienden a comportarse de forma parecida a los demás en su grupo, ya sea porque les afectan las mismas cosas o porque hay imitación entonces las observaciones dentro del grupo NO son estadísticamente independientes (una unidad mas no agrega “muchas” información).

Es decir, si comparamos 10 personas dentro de un grupo vs otras 10 personas de otro grupo, es como si tuvieramos información adicional de MENOS de 20 personas.

Por lo tanto necesitaremos tener muchas unidades cuando comparamos entre grupos en vez de DENTRO de cada grupo.

Tratamientos múltiples

A veces la pregunta central es decidir entre diferentes intervenciones posibles.

¿Libros de texto, vs clases extras?

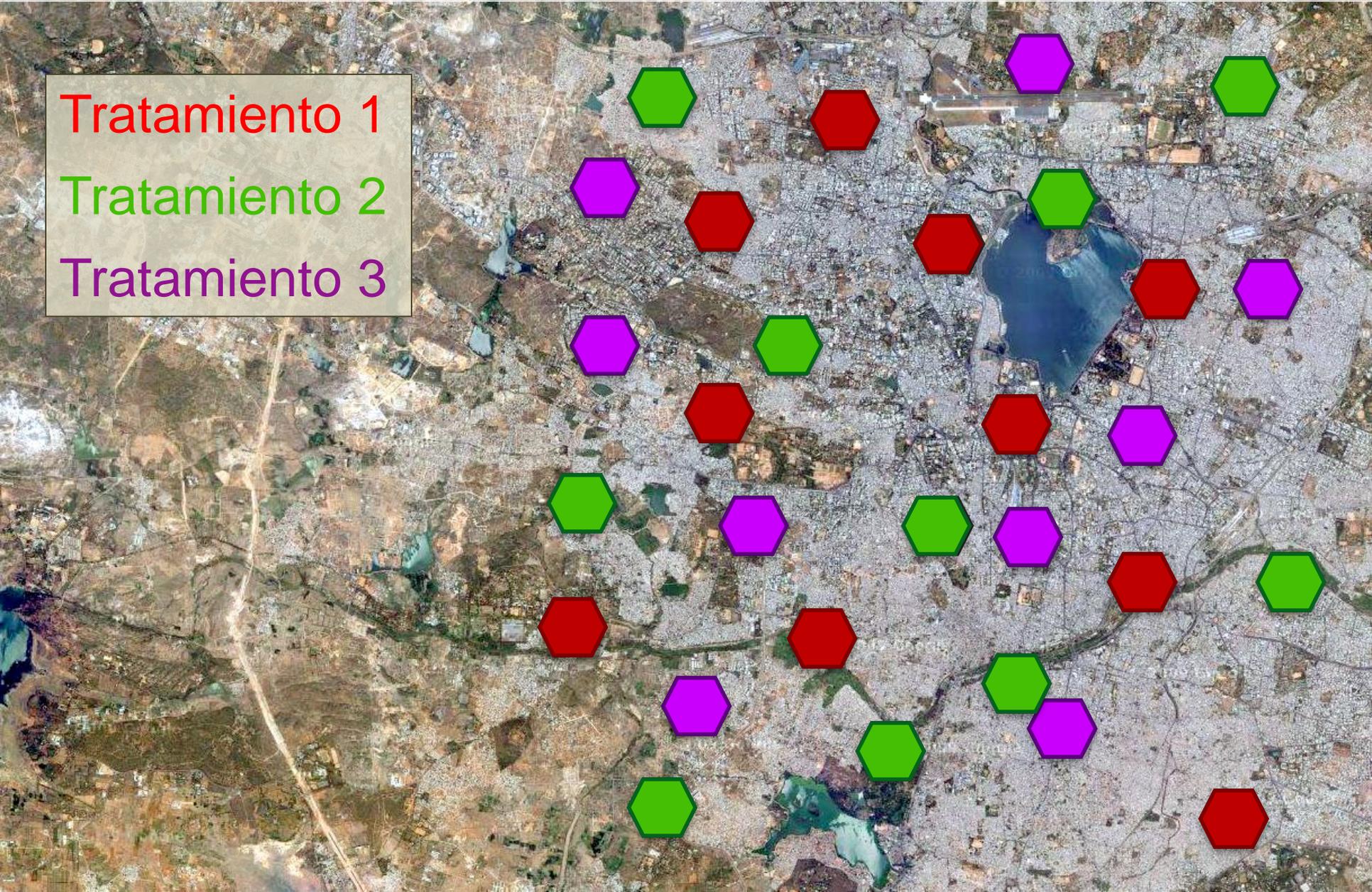
¿Vacuna A vs Vacuna B?

Tratamientos múltiples

Tratamiento 1

Tratamiento 2

Tratamiento 3



Interacción de tratamiento

Prueba diferentes componentes de tratamiento en diferentes combinaciones

Prueba si los componentes sirven como sustitutos o complementos

¿Cuál es la combinación más económica?

Ventaja: situación gana-gana para operaciones, pueden ser útiles para responder preguntas para ellos, ¡más allá del simple “impacto”!

Variando intensidad del tratamiento

A algunas escuelas se les asigna tratamiento completo

- Todos los niños reciben píldoras

A algunas escuelas se les asigna tratamiento parcial

- Se designa que el 50% recibirá píldoras

En otros experimentos: cambiar el precio que se cobra por un bien o servicio

Estratificación

Objetivo: equilibrar su muestra cuando tiene una muestra pequeña.

¿Qué es?

Dividir la muestra en diferentes subgrupos
seleccionar el tratamiento y control de cada
subgrupo.

Ventaja: asegurar tener unidades de tratamiento
y control para subpoblaciones de interés.
Aumentar el poder estadístico.

Cuándo estratificar

Estratificar en variables que podrían tener un impacto importante en la variable de resultados (un poco a ojo)

Estratificar en subgrupos que le interesan particularmente (donde pudiese creer que el impacto del programa será diferente)

La estratificación es más importante cuando bases de datos son pequeñas, porque **asegura** que tengas tratamiento y control en cada estrato.

¿Preguntas, Comentarios, Dudas?

Enrique Seira

ITAM

Abdul Latif Jameel Poverty Action Lab
