

KNO-012 · Experimentos y control de variables

KNO · Formas de conocer

Juan Seguro

v0.1 · ES

Todos los derechos reservados

Resumen del artículo

En los diseños comparativos con asignación aquí tratados, un experimento es una comparación que alguien dispuso, y de esa disposición vienen su ventaja sobre mirar y las dificultades que siguen. Muchos experimentos no asignan nada a nadie, y nada de lo que sigue define el experimento en general. Este artículo sigue la disposición y no el resultado. Separa las dos tareas que se le piden a la palabra aleatorización, generar una secuencia impredecible y ocultar esa secuencia a quien recluta participantes, y registra el precio medido de fallar en la segunda. Distingue las protecciones que terminan en la asignación de las que deben continuar después, razón por la que el enmascaramiento se evalúa aparte. Pasa a los casos en que la intervención es imposible y la comparación hay que encontrarla. Cierra donde el resultado mismo se discute y juzgar la competencia del trabajo exige saber cuál debía ser la respuesta.

En 60 segundos

En un diseño comparativo con asignación, un experimento compra una sola cosa: el investigador decide quién recibe qué, de modo que la razón por la que un grupo recibió el tratamiento tiene respuesta conocida en lugar de desconocida. Los experimentos sin asignación quedan fuera de lo que sigue. Todo lo demás del método protege esa propiedad. La decisión misma es por tanto el punto vulnerable, y la palabra aleatorización cubre ahí dos tareas distintas: generar una secuencia que nadie pueda predecir y ocultar esa secuencia a quien recluta participantes. Las normas de publicación las numeran por separado, y los ensayos juzgados con ocultación inadecuada declaran efectos mayores que los juzgados con ocultación adecuada. La ocultación se detiene en el momento de la asignación. El enmascaramiento es otro instrumento, no siempre está disponible, e importa más quién quedó enmascarado que si aparece la palabra. Donde nadie puede asignar nada, la comparación hay que encontrarla, y afirmar que su variación es como si fuera aleatoria es un supuesto sobre el mundo y no un hecho sobre el diseño. Lo más difícil es un resultado en disputa, donde los juicios sobre competencia y sobre la respuesta se mueven juntos.

Por qué importa esta pregunta

Las protecciones que aquí se tratan pueden ser la diferencia entre una comparación cuya garantía es defendible y otra cuyo resultado aparente desborda su diseño. Y son además invisibles en el resultado: un ensayo declara un tamaño de efecto igual tanto si su secuencia estuvo oculta como si no, y la cifra que sale no lleva marca alguna de la disposición que la produjo. Leer bien un experimento consiste por eso en leer lo que se hizo antes de que los datos existieran, que es la parte que un resultado publicado menos suele mostrar y de la que más suele depender.

1. Lo que un experimento añade

La observación registra lo que ocurrió. Un experimento dispone antes la situación y registra después lo que ocurrió, y la disposición es toda la diferencia.

Esa diferencia compra una sola cosa. **En un experimento es el investigador quien decide quién recibe qué**, de modo que la pregunta de por qué un grupo recibió el tratamiento y otro no tiene una respuesta conocida en lugar de una desconocida. Todo lo demás del método experimental existe para proteger esa única propiedad, y cada caso de las secciones siguientes es un caso de protegerla, de perderla o de trabajar sin ella.

Qué cubre este artículo y qué no. Muchos experimentos no asignan nada a nadie: varían un voltaje, una temperatura, un reactivo o un retardo, sin participantes, sin grupos y sin aleatorización. Lo que sigue trata de **diseños comparativos en los que se asignan unidades a condiciones**, porque esa es la familia en la que las protecciones aquí discutidas hacen su trabajo. Nada de ello debe leerse como una definición de experimento en general.

La palabra control hace aquí dos trabajos y conviene separarlos desde el principio. **Mantener fija una variable** es disponer que no varíe a lo largo de la comparación. **Un grupo de control** es el grupo de comparación frente al que se juzga la intervención: puede recibir un placebo, la atención habitual, otra intervención o nada en absoluto, y lo que lo hace control es el papel que desempeña en la comparación y no la ausencia de tratamiento en él. Lo primero es una propiedad del montaje; lo segundo es una parte del diseño. Un estudio puede tener grupo de control y no mantener nada fijo, y puede mantener mucho fijo y no tener nada con qué comparar.

Ningún montaje mantiene todo fijo. La igualdad completa de condiciones a lo largo de una comparación no es alcanzable, y por eso el argumento a favor de una comparación versa siempre sobre qué diferencias se controlaron y cuáles se dejaron al azar, y nunca sobre su ausencia. **Este es un enunciado de encuadre, declarado como tal en la sección 4 de KNO-012_UNCERTAINTIES_RECONSTRUCTION.md.**

En qué punto estamos

La observación registra lo que ocurrió; un experimento dispone antes la situación, y la disposición es toda la diferencia. En los diseños comparativos con asignación tratados aquí compra una propiedad: el investigador decide quién recibe qué, y el resto del método la protege. Se separan dos sentidos de control: mantener fija una variable pertenece al montaje; un grupo de control, al diseño. Un estudio puede tener uno sin el otro.

2. Por qué la asignación es el paso que soporta la carga

Si el investigador decide quién recibe qué, la decisión misma se convierte en el punto vulnerable.

Fisher dio la respuesta clásica, y le puso un título que enuncia el argumento entero: la sección 9 de *The Design of Experiments* se titula «Randomisation; the Physical Basis of the Validity of the Test» [1]. Su afirmación allí es que no hace falta más aparato, porque la simple precaución de la aleatorización basta para garantizar la validez de la prueba de significación [2].

Dos rasgos de esa frase merecen marcarse, y un tercero pertenece a su contexto. Es una afirmación sobre la **validez de una prueba**, no sobre el equilibrio de ningún experimento concreto, y este artículo no la extiende más allá de lo que dice. Es además condicional y no incondicional: la garantía se ofrece sobre un fondo en el que los demás errores evitables se han evitado, no como un aval que sobreviva a ellos. **La afirmación se trata aquí como condicionada por el contexto experimental circundante; no se interpreta como si la aleatorización eliminara todas las demás fuentes evitables de error.** Y describe la aleatorización como una **precaución**, que es la categoría correcta: es un procedimiento adoptado de antemano, cuyo valor no depende del aspecto que los grupos resultantes tengan después.

En qué punto estamos

El punto vulnerable es la decisión misma, en cuanto alguien la toma. La respuesta de Fisher atañe a la validez de una prueba y no al equilibrio de un experimento cualquiera. El artículo trata la afirmación como condicionada por un contexto en el que se evitaron otros errores evitables, y no como una garantía de que la aleatorización los elimine. Llama precaución a la aleatorización.

3. Dos tareas, una palabra

En la práctica se le piden a la aleatorización dos cosas distintas, y solo una de ellas es aritmética.

La primera es **generar** una secuencia de asignación que nadie pueda predecir. La segunda es **ocultar** esa secuencia a quien recluta a los participantes, de modo que saber a qué rama caerá la siguiente persona no pueda influir en si esa persona llega a ser reclutada. Las normas de publicación tratan estas obligaciones por separado: la lista de comprobación CONSORT 2010 recoge la generación de la secuencia, el mecanismo de ocultación de la asignación y el enmascaramiento como tres puntos numerados por separado, 8a, 9 y 11 [3].

La separación no es pulcritud burocrática, y lleva un número asociado. Schulz y colaboradores compararon ensayos según lo adecuado de su ocultación y hallaron que las razones de momios quedaron exageradas en un 41 % en los ensayos con ocultación inadecuada y en un 30 % en los de ocultación poco clara, ajustando por otros aspectos de la calidad [4].

Ese resultado hay que leerlo con cuidado, porque es fácil exagerarlo. Se refiere al conjunto de ensayos examinados en aquel estudio, es una asociación entre un rasgo metodológico declarado y el tamaño de los efectos declarados, y la exageración se mide frente a ensayos juzgados con ocultación adecuada, no frente a una verdad conocida. Lo que establece es estrecho y grave: **una aleatorización bien generada y mal ocultada no queda protegida por haber sido bien generada.** Lo que no establece es nada sobre la otra mitad del par. La comparación ordena los ensayos según lo adecuado de su ocultación, de modo que no muestra que las secuencias estuvieran correctamente generadas en aquellos cuya ocultación se juzgó inadecuada. Solo una de las dos propiedades es aquella por la que este estudio ordena.

En qué punto estamos

Una palabra cubre dos tareas. Generar una secuencia que nadie pueda predecir es aritmética; ocultarla a quien recluta participantes no lo es, y las normas de publicación las numeran aparte. El precio medido de fallar en la segunda es una asociación dentro de un conjunto de ensayos, medida frente a los juzgados con ocultación adecuada y no frente a una verdad conocida. Nada dice de la otra mitad del par.

4. Antes de la asignación y después

La asignación protege un momento. Casi todo el experimento ocurre después de ese momento, y para eso hacen falta instrumentos distintos.

El manual Cochrane traza la línea de forma explícita. La ocultación de la secuencia de asignación busca evitar el sesgo en la asignación de la intervención impidiendo que el personal del ensayo y los participantes conozcan la secuencia antes de la asignación y hasta ella [5], y, de forma decisiva, puede implementarse siempre con éxito, con independencia del diseño del estudio o del área clínica [6]. El enmascaramiento es un instrumento de la clase contraria: busca evitar el sesgo después de la asignación y no siempre puede implementarse [7].

Esas dos frases juntas son lo más útil de este artículo. Una protección está disponible en los diseños y áreas clínicas que el manual describe, que son los ensayos aleatorizados, y tiene un límite temporal. Leída fuera de ese marco, la palabra siempre afirmaría más de lo que dice la fuente. La otra no siempre está disponible, porque hay intervenciones que no pueden disimularse, y donde lo está no tiene por qué extenderse a toda la duración del ensayo: ha de sostenerse en las personas, y en las etapas, en que conocer la asignación pudiera introducir sesgo. Un ensayo puede por tanto ser impecable en el punto de la asignación y quedar expuesto durante el resto de su duración, y los dos estados deben evaluarse por separado en lugar de resumirse en una única puntuación de calidad.

El mismo manual va más lejos y separa al evaluador del participante: el enmascaramiento de quienes evalúan el resultado, para evitar el sesgo en la *medición* del resultado, se considera aparte, en el dominio de sesgo en la medición de los resultados [8]. **Importa más quién quedó enmascarado que si aparece la palabra enmascarado.** Un ensayo en el que los participantes no pudieron enmascararse pero sí los evaluadores, y otro en el que ocurre lo contrario, están expuestos a fallos distintos y no deberían describirse con la misma expresión.

En qué punto estamos

La asignación protege un momento y casi todo el ensayo ocurre después. La ocultación evita el sesgo antes de la asignación y hasta ella y, dentro de los ensayos aleatorizados que el manual describe, puede implementarse siempre. El enmascaramiento actúa después y no siempre puede implementarse, y donde puede no tiene por qué cubrir toda la duración. Importa por eso más quién quedó enmascarado que si aparece la palabra.

5. Cuando la comparación hay que encontrarla y no fabricarla

Para muchas preguntas la disposición de la sección 1 sencillamente no está disponible. El documento de fundamentación del comité del premio enuncia la restricción sin rodeos: por limitaciones económicas, éticas o prácticas, muchas preguntas importantes, en particular en las ciencias sociales, no pueden estudiarse mediante un experimento aleatorizado controlado [9].

La respuesta consiste en buscar situaciones que hicieran la disposición por uno. El mismo documento define el objeto: un experimento natural es un suceso o una situación, no sometida al control de los sujetos estudiados, que genera variación en la variable de interés como si hubiera sido asignada al azar [10].

La carga de esa definición la llevan dos expresiones y ambas son condiciones, no descripciones. El suceso debe quedar fuera del control de los sujetos, lo que cierra una vía por la que las personas estudiadas podrían seleccionarse a sí mismas hacia el grupo que les conviene. No cierra por sí solo las demás. La selección ejercida por alguien distinto de los sujetos, la anticipación del suceso y las diferencias sistemáticas que el suceso acompaña en lugar de crear sobreviven todas a esa condición y hay que argumentarlas caso por caso. Que la variación resultante sea comparable a una asignación al azar es por tanto una afirmación que exige justificación, y el instrumental para evaluar esas justificaciones es la inferencia causal, que esta serie aborda en KNO-014 y no aquí. Y la variación debe ser *como si* estuviera asignada al azar, lo cual es un supuesto sobre el mundo y no un hecho sobre el diseño, porque nadie aleatorizó nada.

Esta es la diferencia con la sección 2 y no debe difuminarse. En un experimento aleatorizado la razón de la asignación se conoce porque alguien la produjo. En un experimento natural la razón se argumenta. Ese argumento puede ser fuerte, y el diseño puede ser la única vía disponible hacia una pregunta que importa, pero es una garantía de otra clase y este artículo no trata las dos como intercambiables.

En qué punto estamos

Donde la disposición no está disponible, la comparación hay que encontrarla. La definición lleva dos condiciones y no dos descripciones: el suceso queda fuera del control de los sujetos, lo que cierra una vía de autoselección y no las demás, y la variación es *como si* estuviera asignada al azar, que es un supuesto sobre el mundo. En un diseño aleatorizado la razón se conoce; aquí se argumenta.

6. Cuando lo que está en disputa es el resultado mismo

El caso más difícil no es un experimento mal hecho. Es un experimento bien hecho cuyo resultado se discute, porque entonces la cuestión de si estuvo bien hecho no puede zanjarse con independencia del resultado.

Collins formuló la circularidad de la manera que le dio nombre, en un pasaje que Franklin cita de él: no tendremos idea de si podemos hacerlo hasta que intentemos ver si obtenemos el resultado correcto; pero ¿cuál es el resultado correcto?; cuál sea el resultado correcto depende de si hay o no ondas gravitatorias alcanzando la Tierra en flujos detectables; para averiguarlo hemos de construir un buen detector de ondas gravitatorias y mirar; pero no sabremos si hemos construido un buen detector hasta que lo hayamos probado y hayamos obtenido el resultado correcto [11].

Ese es el diagnóstico de Collins, y aquí queda registrado como suyo y no adoptado. Según él, la competencia se juzga por el resultado mientras que el resultado se acepta según la competencia. En un campo asentado esto es invisible, porque todo el mundo sabe ya qué produce un instrumento que funciona. Se vuelve visible precisamente cuando un campo no está asentado, que es cuando importa.

De ahí no se sigue que el círculo no pueda romperse, y este artículo no afirma que no pueda.

Franklin, que recoge el pasaje, argumenta en contra de esa conclusión y sostiene que el regreso puede romperse mediante argumentación razonada [12]. Las dos posiciones quedan registradas y no se adopta ninguna, y conviene notar que no son simétricas: una dice que existe una dificultad, la otra que es superable, y ambas pueden ser ciertas.

El límite de alcance fijado antes en esta serie se conserva aquí. El regreso es un problema real en algunos casos disputados, no una ley general según la cual toda calibración sea circular. Un termómetro contrastado con un patrón nacional no está en un regreso, y describirlo como si lo estuviera vaciaría el concepto del contenido que lo hace útil.

En qué punto estamos

Un experimento bien hecho cuyo resultado se discute es el caso difícil, porque entonces la competencia no puede juzgarse con independencia del resultado. Collins formula ese círculo y Franklin, que lo cita, sostiene que puede romperse mediante argumentación razonada. Ambas posiciones quedan registradas sin adoptar ninguna, y no son simétricas. El regreso pertenece a los casos disputados y no a la calibración en general.

7. Leer bien un experimento

¿Se asignó algo, o solo se observó? Si nadie asignó, la razón por la que cada persona está en el grupo en que está hay que argumentarla en lugar de conocerla.

¿Quién sabía qué, y cuándo? Generación de la secuencia y ocultación de la asignación son fallos distintos. También lo son ocultación y enmascaramiento, y también el enmascaramiento de los participantes y el de quienes evalúan el resultado.

¿Qué protecciones terminaron en la asignación? La ocultación está disponible en los ensayos aleatorizados que el manual describe, y se detiene en la asignación. El enmascaramiento no siempre está disponible, y donde lo está, ha de sostenerse en las personas y etapas en que conocer la asignación pudiera sesgar lo que ocurre o lo que se registra.

¿Cuál es el grupo de comparación, y qué lo hace comparable? En un diseño aleatorizado la respuesta es procedimental. En un experimento natural es un argumento sobre el mundo, y el argumento debería estar a la vista.

¿Está en disputa el resultado? Si lo está, cabe esperar que los juicios sobre la competencia y los juicios sobre la respuesta se muevan juntos, y conviene buscar las razones ofrecidas en lugar de un experimento decisivo.

¿Qué se mantuvo fijo, y era eso lo mismo que tener un grupo de control? Son cosas separadas, y un estudio puede tener una sin la otra.

Un experimento no es un dispositivo que produce verdad. Es una disposición que sustituye una incógnita, por qué estas personas y no aquellas, por un procedimiento, y que dedica el resto de su duración a defender esa sustitución.

En qué punto estamos

El artículo termina en preguntas para llevarse a un experimento: si se asignó algo, quién sabía qué y cuándo, qué protecciones se detuvieron en la asignación, qué hace comparable al grupo de comparación, si el resultado está en disputa y si lo que se mantuvo fijo era lo mismo que un grupo de control. Una disposición, no un dispositivo que produce verdad. Qué sabemos, y cómo, es lo que sigue.

Lo que sabemos y cómo lo sabemos

Fisher tituló la sección 9 de *The Design of Experiments* «Randomisation; the Physical Basis of the Validity of the Test», y sostuvo allí que la simple precaución de la aleatorización basta para garantizar la validez de la prueba de significación. Es una afirmación sobre la validez de una prueba y no sobre el equilibrio de ningún experimento concreto, y el contexto deja claro además que la validez no equivale a afirmar que la aleatorización elimine todas las demás fuentes evitables de error. La lista CONSORT 2010 recoge la generación de la secuencia, el mecanismo de ocultación y el enmascaramiento como tres puntos numerados por separado. Schulz y colaboradores, comparando ensayos según lo adecuado de su ocultación, informan de razones de momios exageradas en un 41 % con ocultación inadecuada y en un 30 % con ocultación poco clara. El manual Cochrane dice que la ocultación evita el sesgo antes de la asignación y hasta ella y, dentro de los ensayos aleatorizados que el manual describe, puede implementarse siempre, que el enmascaramiento lo evita después y no siempre puede, y que el de quienes evalúan el resultado se considera aparte. El documento del comité del premio enuncia que muchas preguntas importantes no pueden estudiarse con experimento aleatorizado controlado, y define el experimento natural. Collins formula el regreso en un pasaje que Franklin cita, y Franklin sostiene que puede romperse mediante argumentación razonada.

Lo que todavía no sabemos

Si las secuencias estaban correctamente generadas en los ensayos cuya ocultación se juzgó inadecuada, ya que la comparación de Schulz ordena por ocultación y no por generación. Hasta dónde viajan las cifras de exageración más allá del conjunto de ensayos examinados, siendo una asociación entre un rasgo declarado y los tamaños de efecto declarados y no una distancia respecto de una verdad conocida. Si la variación de un experimento natural concreto es de veras como si estuviera asignada al azar, que es un supuesto sobre el mundo y no un hecho sobre el diseño. Y si el regreso puede romperse en un caso disputado dado, donde Collins y Franklin quedan registrados sin que se adopte ninguno.

Qué podría hacernos cambiar de opinión

Una comparación que ordenara los ensayos según lo adecuado de la generación de la secuencia y no de la ocultación diría algo sobre la mitad del par que esta evidencia deja intacta. Estudios de ocultación fuera de los ensayos aleatorizados que el manual Cochrane describe pondrían a prueba hasta dónde llega ese siempre. Para un experimento natural, evidencia que cerrara las vías que la definición deja abiertas, la selección ejercida por alguien distinto de los sujetos, la anticipación del suceso y las diferencias sistemáticas que el suceso acompaña, acercaría el argumento a la garantía procedimental que sustituye.

Si solo recuerdas una cosa

En los diseños comparativos con asignación aquí tratados, la asignación sustituye la razón desconocida de por qué estas unidades y no aquellas por un procedimiento, y el resto la defiende.

Artículos relacionados

- KNO-014 · *Causalidad y correlación*

Referencias

Los números entre corchetes del cuerpo son números de afirmación y no de referencia. Cada uno remite a una fila del registro de afirmaciones, que nombra la fuente de abajo que la sostiene.

Comité del Premio en Ciencias Económicas. 2021. *Answering Causal Questions Using Observational Data. Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021*. The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel, The Royal Swedish Academy of Sciences, 11 de octubre de 2021. URL estable: <https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf>

Fisher, R. A. 1935. *The Design of Experiments*. Edimburgo y Londres: Oliver and Boyd. **Sin ISBN: la obra es anterior al sistema ISBN**. URL estable del ejemplar digitalizado consultado: <https://archive.org/details/in.ernet.dli.2015.502684>

Franklin, Allan. 1994. «How to Avoid the Experimenters' Regress». *Studies in History and Philosophy of Science* 25 (3): 463–491. **Sin DOI mostrado sobre página consultada; identificador impreso 0039-3681/94**. URL estable: https://physics.yale.edu/sites/default/files/files/3_1%20franklin.pdf

Higgins, Julian PT, Jelena Savović, Matthew J Page, Roy G Elbers, y Jonathan AC Sterne. 2024. «Chapter 8: Assessing risk of bias in a randomized trial». En *Cochrane Handbook for Systematic Reviews of Interventions*, versión 6.5. Cochrane. URL estable: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-08>

Schulz, Kenneth F., Douglas G. Altman, y David Moher, for the CONSORT Group. 2010. «CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials». *PLoS Medicine* 7 (3): e1000251. DOI 10.1371/journal.pmed.1000251

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, y Douglas G. Altman. 1995. «Empirical Evidence of Bias: Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials». *JAMA* 273 (5): 408–412. DOI 10.1001/jama.1995.03520290060030 URL estable: https://www.jameslindlibrary.org/wp-data/uploads/2012/02/Schulz_Chalmers_Hayes_Altman_1995.pdf