

Ensayos aleatorizados e inferencia estadística

Juan Seguro

Pregunta central

¿Cómo cambiaron la aleatorización, los grupos explícitos de comparación y la inferencia estadística formal el estándar para decidir si una intervención causa una mejora reproducible?

Contexto histórico y evolutivo

La teoría de la probabilidad, las estadísticas demográficas y el análisis de errores del siglo XIX ya habían aportado herramientas para razonar a partir de observaciones variables. Desde la década de 1880, Galton desarrolló regresión y correlación, y Pearson formalizó el coeficiente de correlación producto-momento en 1896; R. A. Fisher formalizó después aleatorización, replicación y diseño experimental en investigación agrícola durante las décadas de 1920 y 1930. El trabajo de Neyman y Pearson de 1933 aclaró ideas de contraste de hipótesis relacionadas con errores y potencia. La medicina adoptó controles relacionados de forma desigual. El ensayo del Medical Research Council de 1948 sobre estreptomicina para tuberculosis pulmonar se convirtió en un hito porque la asignación del tratamiento utilizó números aleatorios, los resultados se compararon prospectivamente y los procedimientos de evaluación se organizaron explícitamente para limitar el sesgo.

¿Qué podían saber las comunidades antes de esta transición?

Los clínicos podían comparar resultados antes y después, series de casos y grupos tratados frente a no tratados, y PIO-065 había mostrado cómo los patrones epidemiológicos podían poner a prueba afirmaciones causales. Sin embargo, la selección de tratamiento, el pronóstico y las expectativas del observador podían diferir sistemáticamente entre grupos. Una mejoría aparente podía reflejar quién recibía el tratamiento y no lo que hacía el tratamiento.

¿Qué cambió?

La asignación aleatoria rompió el vínculo directo entre pronóstico y elección del tratamiento

La aleatorización no garantiza grupos perfectamente idénticos, pero hace la asignación del tratamiento independiente de la preferencia del investigador y proporciona una base probabilística para la comparación. El trabajo de Fisher sobre diseño experimental hizo explícita esta lógica antes de su adopción clínica sistemática.

El grupo de comparación pasó a formar parte de la afirmación sobre la intervención

En la investigación de estreptomicina de 1948, los resultados de grupos tratado y control se siguieron bajo un protocolo común. La afirmación causal dejó de ser «los pacientes mejoraron después del tratamiento» y pasó a ser «los resultados difirieron bajo condiciones asignadas prospectivamente».

La inferencia conectó diferencias observadas con incertidumbre controlada

Los contrastes de Neyman-Pearson, los conceptos de confianza y métodos frecuentistas relacionados crearon formas formales de separar variabilidad muestral de efectos suficientemente grandes para poner en cuestión un modelo nulo. La discusión de Bradford Hill de 1952 sobre el ensayo clínico ayudó a articular cómo debían interactuar estas ideas estadísticas con el juicio médico en vez de sustituirlo.

¿Por qué pudo surgir esta capacidad?

La transición dependió de probabilidad, censos, experimentos agrícolas, biometría, estadísticas de salud pública, sistemas fiables de registro y terapias como la estreptomicina cuyos efectos podían medirse prospectivamente. Requirió además capacidad administrativa para reclutar, asignar y seguir pacientes según protocolo.

¿Qué formas posteriores de conocimiento hizo posibles?

Los ensayos controlados aleatorizados se volvieron centrales para evaluación de fármacos, investigación comparativa de tratamientos y síntesis de evidencia. Su lógica apoyó después registro de ensayos, desenlaces preespecificados, revisiones sistemáticas y metaanálisis.

Nuevos límites y riesgos

La aleatorización no corrige mala medición, pérdidas de seguimiento, comunicación selectiva de resultados, falta de adherencia o reclutamiento no representativo. La significación estadística no equivale a importancia clínica, y un efecto medio del tratamiento puede ocultar beneficio y daño heterogéneos. Los ensayos pueden además ser no éticos cuando no existe incertidumbre genuina o se priva a participantes de atención eficaz establecida.

¿Qué cambió en la arquitectura del conocimiento?

La evidencia médica causal se convirtió en un **sistema prospectivo de asignación y comparación**. El diseño de quién recibe qué intervención pasó a formar parte de la afirmación evidencial y dejó de ser un detalle administrativo.

Pérdida, exclusión e incertidumbre

La expresión «primer ensayo clínico aleatorizado» puede borrar comparaciones controladas anteriores y el desarrollo gradual de la aleatorización en agricultura, psicología y medicina. También puede ocultar la contribución de pacientes y personal clínico. El ensayo de estreptomycin de 1948 se trata aquí como una consolidación importante, no como una invención aislada.

Mapa de dependencias

- DIRECTO ← PIO-056 Probabilidad, censos y error: de Pascal a Quetelet
- INDIRECTO ← PIO-058 De la variolación a la vacunación
- INDIRECTO ← PIO-065 Semmelweis, Snow y la evidencia a partir de patrones
- INDIRECTO ← PIO-069 Pasteur, Lister y Koch: de la intervención antiséptica a la demostración microbiana
- INDIRECTO ← PIO-081 Antibióticos y control experimental de la infección
- DIRECTO → PIO-086 Ética de la investigación, consentimiento y límites de la investigación permisible

Referencias

1. Fisher, R. A. *The Design of Experiments*. Oliver and Boyd, Edinburgh (1935).
2. Neyman, J. & Pearson, E. S. «On the Problem of the Most Efficient Tests of Statistical Hypotheses.» *Philosophical Transactions of the Royal Society of London. Series A* 231, 289-337 (1933). DOI: 10.1098/rsta.1933.0009.
3. Medical Research Council. «Streptomycin Treatment of Pulmonary Tuberculosis: A Medical Research Council Investigation.» *British Medical Journal* 2(4582), 769-782 (1948). DOI: 10.1136/bmj.2.4582.769.
4. Hill, A. B. «The Clinical Trial.» *New England Journal of Medicine* 247(4), 113-119 (1952). DOI: 10.1056/NEJM195207242470401.