

Randomised trials and statistical inference

Juan Segurola

Core question

How did randomisation, explicit comparison groups and formal statistical inference change the standard for deciding whether an intervention causes a reproducible improvement?

Historical and evolutionary context

Probability theory, demographic statistics and nineteenth-century error analysis had already supplied tools for reasoning from variable observations. From the 1880s, Galton developed regression and correlation, and Pearson formalised the product-moment correlation coefficient in 1896; R. A. Fisher later formalised randomisation, replication and experimental design in agricultural research during the 1920s and 1930s. Neyman and Pearson's 1933 work clarified hypothesis-testing ideas involving errors and power. Medicine adopted related controls unevenly. The Medical Research Council's 1948 streptomycin trial for pulmonary tuberculosis became a landmark because treatment allocation used random numbers, outcomes were compared prospectively and assessment procedures were explicitly organised to limit bias.

What could communities know before this transition?

Clinicians could compare before-and-after outcomes, case series and treated versus untreated groups, and PIO-065 had shown how epidemiological patterns could test causal claims. Yet treatment selection, prognosis and observer expectations could differ systematically between groups. An apparent improvement could therefore reflect who received treatment rather than what treatment did.

What changed?

Random allocation broke the direct link between prognosis and treatment choice

Randomisation does not guarantee perfectly identical groups, but it makes treatment assignment independent of investigator preference and supplies a probabilistic basis for comparison. Fisher's experimental design work made this logic explicit before its systematic clinical adoption.

The comparison group became part of the intervention claim

In the 1948 streptomycin investigation, outcomes in treated and control groups were followed under a common protocol. The causal claim was no longer “patients improved after treatment” but “outcomes differed under prospectively assigned conditions.”

Inference connected observed differences to controlled uncertainty

Neyman-Pearson testing, confidence concepts and related frequentist methods created formal ways to separate sampling variability from effects large enough to challenge a null model. Bradford Hill's 1952 discussion of the clinical trial helped articulate how these statistical ideas should interact with medical judgment rather than replace it.

Why could this capability emerge?

The transition depended on probability, censuses, agricultural experiments, biometrics, public-health statistics, reliable record systems and therapies such as streptomycin whose effects could be prospectively measured. It also required administrative capacity to enrol, allocate and follow patients according to protocol.

What later forms of knowledge did it enable?

Randomised controlled trials became central to drug evaluation, comparative treatment research and evidence synthesis. Their logic later supported trial registration, prespecified endpoints, systematic reviews and meta-analysis.

New limits and risks

Randomisation does not correct poor measurement, attrition, selective reporting, nonadherence or unrepresentative recruitment. Statistical significance is not clinical importance, and an average treatment effect may conceal heterogeneous benefit and harm. Trials can also be unethical when genuine uncertainty is absent or participants are denied established effective care.

What changed in the architecture of knowledge?

Causal medical evidence became a **prospective allocation-and-comparison system**. The design of who receives what intervention became part of the evidential claim rather than an administrative detail.

Loss, exclusion and uncertainty

The phrase “first randomised clinical trial” can erase earlier controlled comparisons and the gradual development of randomisation across agriculture, psychology and medicine. It can also obscure the contribution of patients and clinical staff. The 1948 streptomycin trial is treated here as a major consolidation, not an isolated invention.

Dependency map

- DIRECT ← PIO-056 Probability, censuses and error: Pascal to Quetelet
- INDIRECT ← PIO-058 From variolation to vaccination
- INDIRECT ← PIO-065 Semmelweis, Snow and evidence from patterns
- INDIRECT ← PIO-069 Pasteur, Lister and Koch: from antiseptic intervention to germ proof
- INDIRECT ← PIO-081 Antibiotics and experimental control of infection
- DIRECT → PIO-086 Research ethics, consent and the limits of permissible inquiry

References

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.