

KNO-012 · Experiments and Variable Control

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

In the assigned comparative designs treated here, an experiment is a comparison that someone arranged, which is its advantage over watching and the source of the difficulties that follow. Many experiments assign nothing to anybody, and none of what follows defines experiment in general. This article follows the arrangement rather than the result. It separates the two jobs the word randomisation is asked to do, generating an unpredictable sequence and concealing that sequence from whoever enrolls participants, and records the measured price of failing at the second. It distinguishes protections that end at assignment from those that must continue afterwards, which is why blinding is assessed separately. It turns to the cases where intervention is impossible and the comparison has to be found rather than made. It closes where the result itself is disputed and judging the competence of the work requires knowing what the answer should have been.

In 60 Seconds

In an assigned comparative design, an experiment buys one thing: the investigator decides who receives what, so the reason one group got the treatment has a known answer instead of an unknown one. Experiments that assign nothing lie outside what follows. Everything else in the method protects that property. The decision itself is therefore the vulnerable point, and the word randomisation covers two different jobs at it: generating a sequence nobody can predict, and concealing that sequence from whoever enrolls participants. Reporting standards number them separately, and trials judged inadequately concealed report larger effects than trials judged adequately concealed. Concealment stops at the moment of assignment. Blinding is a different instrument, it is not always available, and who was blinded matters more than whether the word appears. Where no one can assign anything, the comparison has to be found instead, and the claim that its variation is as if randomised is an assumption about the world rather than a fact about the design. Hardest of all is a contested result, where judgements about competence and judgements about the answer move together.

Why This Question Matters

The protections discussed here can be the difference between a comparison whose warrant is defensible and one whose apparent result outruns its design. They are also invisible in the result: a trial reports an effect size the same way whether its sequence was concealed or not, and the number that comes out carries no mark of the arrangement that produced it. Reading an experiment well therefore means reading what was done before the data existed, which is the part a published result is least likely to show and most likely to depend on.

1. The thing an experiment adds

Observation records what happened. An experiment arranges the situation first and then records what happened, and the arrangement is the whole of the difference.

That difference buys one thing. **In an experiment the investigator decides who receives what**, so the question of why one group got the treatment and another did not has a known answer instead of an unknown one. Everything else in experimental method exists to protect that single property, and every case in the following sections is a case of protecting it, losing it, or working without it.

What this article covers, and what it does not. Many experiments assign nothing to anybody: they vary a voltage, a temperature, a reagent or a delay, with no participants, no groups and no randomisation. What follows is about **comparative designs in which units are assigned to conditions**, because that is the family in which the protections discussed here do their work. None of it should be read as a definition of experiment in general.

The word control does two jobs in this field and they are worth separating at the start. **Holding a variable fixed** means arranging that it does not vary across the comparison. **A control group** is the comparison group against which the intervention is judged: it may receive a placebo, usual care, a different intervention or nothing at all, and what makes it a control is the role it plays in the comparison rather than the absence of treatment in it. The first is a property of the setup; the second is a part of the design. A study can have a control group and hold nothing fixed, and it can hold a great deal fixed and have nothing to compare.

No setup holds everything fixed. Complete equality of conditions across a comparison is not attainable, which is why the case for a comparison is always about which differences were controlled and which were left to chance, and never about their absence. **This is a framing statement, declared as such in section 4 of KNO-012_UNCERTAINTIES_RECONSTRUCTION.md.**

Where That Leaves Us

Observation records what happened; an experiment arranges the situation first, and the arrangement is the whole difference. In the assigned comparative designs treated here it buys one property, that the investigator decides who receives what, and the rest of the method protects it. Control has two senses: holding a variable fixed belongs to the setup, a control group to the design. A study can have one without the other.

2. Why allocation is the load bearing step

If the investigator decides who receives what, the decision itself becomes the vulnerable point.

Fisher gave the classical answer, and gave it a heading that states the whole argument: section 9 of *The Design of Experiments* is titled "Randomisation; the Physical Basis of the Validity of the Test" [1]. His claim there is that no further apparatus is required, because "the simple precaution of randomisation will suffice to guarantee the validity of the test of significance" [2].

Two features of that sentence are worth marking, and a third belongs to its context. It is a claim about the **validity of a test**, not about the balance of any particular experiment, and this article does not extend it beyond what it says. It is also conditional rather than unqualified: the guarantee is offered against a background in which other avoidable errors have been avoided, not as a warrant that survives them. **The statement is treated here as conditional on the surrounding experimental context; it is not interpreted as a claim that randomisation removes every other avoidable source of error.** And it describes randomisation as a **precaution**, which is the right category: it is a procedure adopted in advance, whose value does not depend on how the resulting groups happen to look afterwards.

Where That Leaves Us

The vulnerable point is the decision itself, once someone is making it. Fisher's answer concerns the validity of a test and not the balance of any one experiment. The article treats the statement as conditional on a background where other avoidable errors were avoided, rather than as a guarantee that randomisation removes them. He calls randomisation a precaution.

3. Two jobs, one word

In practice randomisation is asked to do two different things, and only one of them is arithmetic.

The first is **generating** an allocation sequence that nobody can predict. The second is **concealing** that sequence from whoever is enrolling participants, so that knowing which arm the next person will fall into cannot influence whether that person is enrolled at all. Reporting standards treat these as separate obligations: the CONSORT 2010 checklist carries sequence generation, allocation concealment mechanism and blinding as three separately numbered items, 8a, 9 and 11 [3].

The separation is not bureaucratic tidiness, and there is a number attached to it. Schulz and colleagues compared trials by the adequacy of their concealment and found that "odds ratios were exaggerated by 41% for inadequately concealed trials and by 30% for unclearly concealed trials (adjusted for other aspects of quality)" [4].

Read that result carefully, because it is easy to overstate. It concerns the set of trials examined in that study, it is an association between a reported methodological feature and the size of reported effects, and the exaggeration is measured against trials judged adequately concealed rather than against a known truth. What it establishes is narrow and serious: **a randomisation that is generated properly and concealed poorly is not protected by having been generated properly.** What it does not establish is anything about the other half of the pair. The comparison sorts trials by the adequacy of their concealment, so it does not show that sequences were correctly generated in the trials whose concealment was judged inadequate. Only one of the two properties is what this study sorts on.

Where That Leaves Us

One word covers two jobs. Generating a sequence nobody can predict is arithmetic; concealing it from whoever enrolls participants is not, and reporting standards number them apart. What failing at the second costs was measured in one set of trials, against those judged adequately concealed and not against any known truth. It says nothing about the other half of the pair.

4. Before assignment and after

Allocation protects one moment. Most of an experiment happens after that moment, and different instruments are needed for it.

The Cochrane Handbook draws the line explicitly. Allocation sequence concealment "seeks to prevent bias in intervention assignment by preventing trial personnel and participants from knowing the allocation sequence before and until assignment" [5], and, crucially, "it can always be successfully implemented, regardless of the study design or clinical area" [6]. Blinding is the opposite kind of instrument: it "seeks to prevent bias after assignment and cannot always be implemented" [7].

Those two sentences together are the most useful thing in this article. One protection is available across the designs and clinical areas the handbook is describing, which are randomised trials, and it is time limited. Read outside that setting, the word always would claim more than the

source does. The other is not always available at all, because some interventions cannot be disguised, and where it is available it need not run the whole length of the trial: it has to hold for the people, and at the stages, at which knowing the assignment could introduce bias. A trial can therefore be impeccable at the point of allocation and exposed for the rest of its duration, and the two states have to be assessed separately rather than summarised as a single quality score.

The same handbook goes further and separates the assessor from the participant: "blinding of outcome assessors, to avoid bias in *measuring* the outcome, is considered separately, in the 'Bias in measurement of outcomes' domain" [8]. **Who was blinded matters more than whether the word blinded appears.** A trial in which participants could not be blinded but outcome assessors were, and a trial in which the reverse holds, are exposed to different failures and should not be described by the same phrase.

Where That Leaves Us

Allocation protects one moment and most of a trial happens after it. Concealment prevents bias before and until assignment and, within the randomised trials the handbook describes, can always be implemented. Blinding works after assignment and cannot always be implemented at all, and where it can it need not run the whole length. Who was blinded therefore matters more than whether the word appears.

5. When the comparison has to be found rather than made

For many questions the arrangement in section 1 is simply unavailable. The Nobel Committee's background document states the constraint plainly: "due to financial, ethical, or practical constraints, many important questions — particularly in the social sciences — cannot be studied using a controlled randomized experiment" [9].

The response is to look for situations that did the arranging for you. The same document defines the object: "a natural experiment is an event or a situation, that is not under the control of the subjects under study, which generates variation in the variable of interest that is as if it had been randomly assigned" [10].

The load in that definition is carried by two phrases and both are conditions, not descriptions. The event must be outside the control of the subjects, which closes one route by which the people studied could select themselves into the group that suits them. It does not on its own close the others. Selection by some party other than the subjects, anticipation of the event, and systematic differences that the event tracks rather than creates all survive that condition and have to be argued away case by case. Whether the resulting variation is comparable to random assignment is therefore a claim requiring justification, and the machinery for assessing such justifications is causal inference, which this series takes up in KNO-014 rather than here. And the variation must be *as if* randomly assigned, which is an assumption about the world rather than a fact about the design, because nobody randomised anything.

This is the difference from section 2 and it should not be blurred. In a randomised experiment the reason for the assignment is known because someone produced it. In a natural experiment the reason is argued for. That argument can be strong, and the design can be the only available route to a question that matters, but it is a different kind of warrant and this article does not treat the two as interchangeable.

Where That Leaves Us

Where the arrangement is unavailable the comparison has to be found instead. The definition carries two conditions rather than two descriptions: the event lies outside the subjects' control, which closes one route to self-selection and not the others, and the variation is *as if* randomly assigned, which is an assumption about the world. In a randomised design the reason is known; here it is argued.

6. When the result itself is what is in dispute

The hardest case is not a badly run experiment. It is a well run experiment whose result is contested, because then the question of whether it was well run cannot be settled independently of the result.

Collins put the circularity in the form that gave it its name, in a passage Franklin quotes from him: "we will have no idea whether we can do it until we try to see if we obtain the correct outcome. But what is the correct outcome? What the correct outcome is depends upon whether or not there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried it and obtained the correct outcome!" [11].

That is Collins's diagnosis, and it is recorded here as his rather than adopted. On his account competence is judged by the outcome while the outcome is trusted according to the competence. In a settled field this is invisible, because everyone already agrees what a working instrument produces. It becomes visible precisely when a field is not settled, which is when it matters.

It does not follow that the circle cannot be broken, and this article does not claim that it cannot. Franklin, who reports the passage, argues against exactly that conclusion and holds that "the regress can be broken by reasoned argument" [12]. The two positions are recorded here and neither is adopted, and the reader should note that they are not symmetrical: one says a difficulty exists, the other says it is surmountable, and both can be true.

The scope limit fixed earlier in this series is preserved here. The regress is a real problem in some contested cases, not a general law that every calibration is circular. A thermometer checked against a national standard is not in a regress, and describing it as though it were would empty the concept of the content that makes it useful.

Where That Leaves Us

A well run experiment whose result is contested is the hard case, because competence cannot then be judged independently of the result. Collins states that circle and Franklin, who quotes him, holds it can be broken by reasoned argument. Both are recorded and neither adopted, and they are not symmetrical. The regress belongs to contested cases and not to calibration in general.

7. Reading an experiment well

Was anything assigned, or only observed? If nobody assigned, the reason people are in the groups they are in has to be argued rather than known.

Who knew what, and when? Sequence generation and allocation concealment are different failures. So are concealment and blinding, and so are blinding of participants and blinding of outcome assessors.

Which protections ended at assignment? Concealment is available throughout the randomised trials the handbook describes, and stops at assignment. Blinding is not always available, and where it is, it has to hold for the people and stages at which knowing the assignment could bias what happens or what is recorded.

What is the comparison group, and what makes it comparable? In a randomised design the answer is procedural. In a natural experiment it is an argument about the world, and the argument should be visible.

Is the result contested? If it is, expect judgements about competence and judgements about the answer to move together, and look for the reasons offered rather than for a decisive experiment.

What was held fixed, and was that the same thing as having a control group? They are separate, and a study can have one without the other.

An experiment is not a device that produces truth. It is an arrangement that replaces one unknown, why these people and not those, with a procedure, and then spends the rest of its length defending that substitution.

Where That Leaves Us

The article ends in questions a reader can carry to an experiment: whether anything was assigned, who knew what and when, which protections stopped at assignment, what makes the comparison group comparable, whether the result is contested, and whether what was held fixed is the same thing as a control group. An arrangement, not a device producing truth. What is established, and how, follows.

What We Know, and How

Fisher titled section 9 of *The Design of Experiments* Randomisation; the Physical Basis of the Validity of the Test, and argued there that the simple precaution of randomisation suffices to guarantee the validity of the test of significance. That is a claim about the validity of a test and not about the balance of any particular experiment, and the surrounding context also makes clear that validity is not a claim that randomisation removes every other avoidable source of error. The CONSORT 2010 checklist carries sequence generation, allocation concealment mechanism and blinding as three separately numbered items. Schulz and colleagues, comparing trials by the adequacy of their concealment, report odds ratios exaggerated by 41 per cent for inadequately concealed trials and by 30 per cent for unclearly concealed ones. The Cochrane Handbook says concealment prevents bias before and until assignment and, within the randomised trials the handbook describes, can always be implemented, that blinding prevents bias after assignment and cannot, and that blinding of outcome assessors is considered separately. The Nobel Committee's background document states that many important questions cannot be studied by controlled randomised experiment, and defines a natural experiment. Collins states the regress in a passage Franklin quotes, and Franklin holds it can be broken by reasoned argument.

What We Still Don't Know

Whether sequences were correctly generated in the trials whose concealment was judged inadequate, since the Schulz comparison sorts on concealment and not on generation. How far the exaggeration figures travel beyond the set of trials examined, being an association between a reported feature and reported effect sizes rather than a distance from a known truth. Whether a

particular natural experiment's variation really is as if randomly assigned, which is an assumption about the world and not a fact about the design. And whether the regress can be broken in a given contested case, on which Collins and Franklin are recorded here and neither is adopted.

What Could Change Our Minds

A comparison that sorted trials by the adequacy of sequence generation rather than of concealment would say something about the half of the pair this evidence leaves untouched. Concealment studies outside the randomised trials the Cochrane Handbook describes would test how far its always extends. For a natural experiment, evidence closing the routes the definition leaves open, selection by a party other than the subjects, anticipation of the event, and systematic differences the event tracks rather than creates, would move the argument towards the procedural warrant it is standing in for.

If You Remember Only One Thing

In the assigned comparative designs treated here, assignment replaces the unknown reason why these units and not those with a procedure, and the rest defends that substitution.

Related Articles

- KNO-014 · *Causality and Correlation*

References

The bracketed numbers in the body are claim numbers, not reference numbers. Each resolves to a row of the claims ledger, which names the source below that supports it.

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