
Bradford Hill Criteria for Causal Inference

Julian King

Director, Julian King & Associates – a member of the Kinnect Group

www.julianking.co.nz | www.kinnect.co.nz

We think we're good at determining causality, but we suck at it

One of the great challenges in evaluation is determining whether the results we're seeing are because of the program we're evaluating, some other influences out there in the big world, or random chance.

At one level, this is an everyday, common sense task. As a species we've been making judgments about causation for a million years or so.

Unfortunately, though, the way we are wired does not predispose us to logical thinking. We are inclined to be led astray by all sorts of biases and heuristics.



Along came logic



Eventually, after a very long time, we evolved into philosophers who invented formal logic. Thanks to scientific method, our species has recently triumphed to the extent that we now have cars that drive themselves and flying drones that deliver pizza (don't confuse this with progress – we still suck at ethics but that's a story for another day).

But we're still not great at causation

Over time, the rocket science for dealing with causation has become more sophisticated – a key example being the experimental study design or randomised controlled trial (RCT). And our evidence base about what works has been enriched as a result.

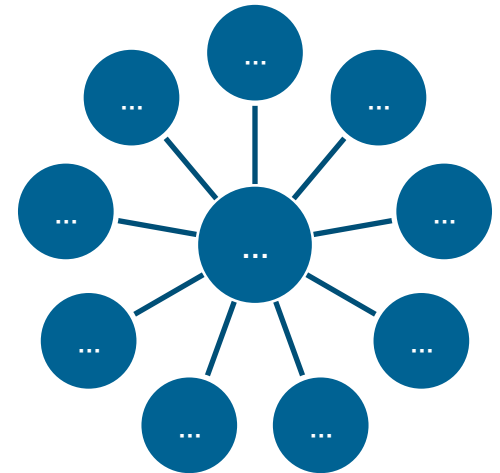
But deep down we're still biased, heuristical beings and not very good at thinking things through. We've become so enthusiastic about experimental designs we're a little inclined to think they are the only way to determine whether A causes B.

Such a rigid view is not much use in the real world, where there are all sorts of ethical, conceptual, practical and economic barriers that mean we can't always conduct RCTs. Even where technically possible, they are not necessarily the best tool for every job.

What's an evaluator to do then?

I will leave the deep thought to others. All I offer here is a small upgrade to your heuristical software, a practical tool to help make reasoned judgments about causation.

This approach may be good enough (robust enough, feasible enough, flexible enough, affordable enough) for many purposes, from a rapid assessment to an in-depth investigation.



Ta-da!

Behold, the Bradford Hill Criteria.

Well, not quite criteria actually.

Behold, the Bradford Hill Really Useful Things to Keep in Mind When Thinking About Causation.



Background

The Bradford Hill Non-Criteria

In 1965, Sir Austin Bradford Hill, an epidemiologist, presented an essay to the Royal Society of Medicine.

In it, he presented what he called, not criteria, but *“nine different viewpoints from all of which we should study association before we cry causation”*

He argued that these viewpoints were not hard-and-fast rules of evidence. They could not provide indisputable evidence for or against a cause-and-effect hypothesis.

But they could help us to weigh the evidence for or against various possible interpretations of cause and effect.

Using Bradford Hill in evaluation

The Bradford Hill Criteria have been used extensively in epidemiology, their primary intended purpose – e.g., to determine whether or not a particular chemical might be considered carcinogenic on the basis of observations in a free living population.

They may have also have broad application to other settings (like evaluation) where we have to consider the balance of real-world, messy evidence to make well reasoned judgments about causation.



Getting started

The first step is to identify possible causal explanations...

The key is logical reasoning

Michael Scriven says good sound reasoning is the real gold standard. In particular, this is about identifying competing explanations for an effect and assessing the evidence for and against each alternative.

I think the Bradford Hill Criteria are a good checklist for applying this reasoning in evaluation practice.

So: Get started by identifying your alternative explanations.

Could it be our program?

Could it be another intervention that was going on at the same time?

Could it be that the change was going to happen anyway?

Rating guide

A simple tool you can
use

The following rating guide
has been developed with
program evaluation in mind.
It has not been validated. I
welcome any feedback on
its use.

It is the totality of support for a cause-effect interpretation that matters.

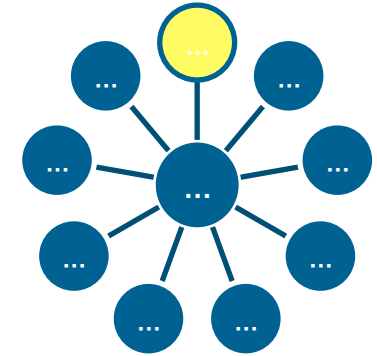
Use the nine viewpoints together

Non-causality	Weak causality	Stronger causality
Factors that collectively suggest non-causality	Factors that collectively, support a weak cause-effect interpretation	Factors that collectively, support a stronger cause-effect interpretation
The more of these factors are present, the stronger the evidence against causality	The more of these factors are present, the more confident we can be in suggesting possible causality	The more of these factors are present, the stronger our causal inference

The following pages step through the nine viewpoints in turn. Although they are presented separately, they need to be considered collectively.

Strength

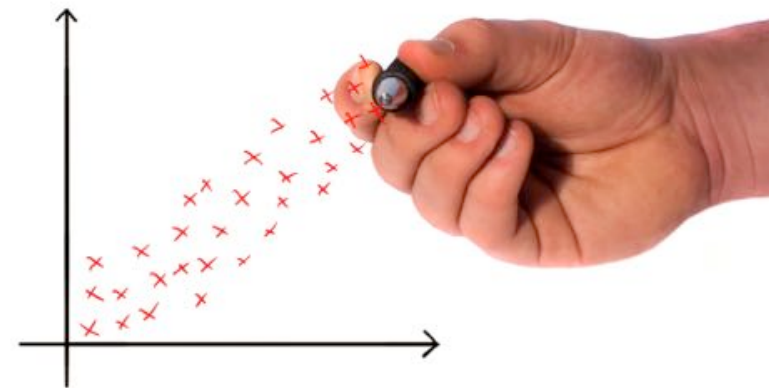
1



Association is not causation. However, a strong association (e.g., correlation, between-group difference, or within-group change), in combination with other viewpoints, can lend support toward a strong cause-effect interpretation.

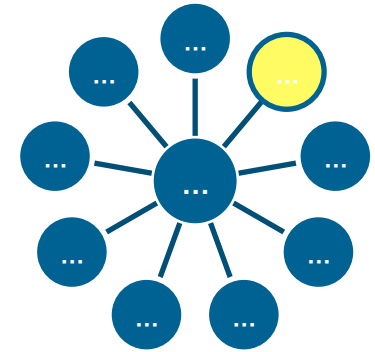
A weak association would contribute weaker support to a causal interpretation.

No association would tend to contribute toward a judgment of non-causality.



Consistency

2



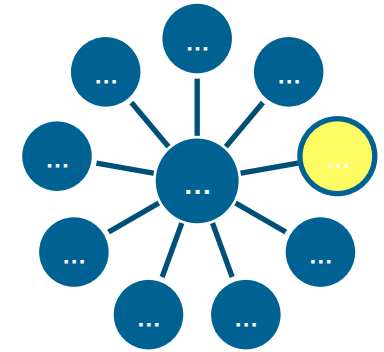
Similar results across different studies, circumstances, places or times contribute toward a cause-effect interpretation. If different types of evidence from multiple sources all support our cause-effect interpretation, this also increases our confidence in the interpretation.

Similar results across *some* studies, circumstances, places or times; or, consistent feedback from multiple stakeholders, contribute weaker evidence.

Inconsistent results across studies, circumstances, places or times would tend to contribute toward a judgment of non-causality.

Specificity

3



A causal inference is strengthened if the effect is observed only in association with the suspected cause, and not seen in the absence of the suspected cause.

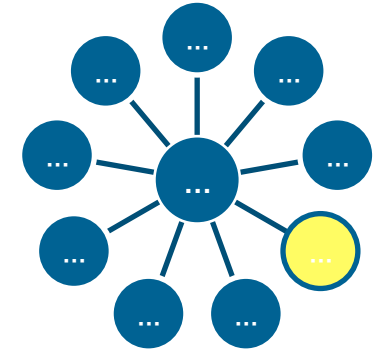
If the effect is seen more often in association with the suspected cause than without it, this might lend weaker support to a causal inference.

Absence of specificity does not rule out a possible causal relationship. Causality can be complex!



Temporality

4



If results are ambiguous, keep in mind that in complex systems like social programs, feedback loops might mean causality is bi-directional and possibly multi-factorial.

By definition, causes happen before effects. If the effect happens after the suspected cause, that lends support to our cause-effect interpretation.

In the real world it might be that the effect *generally* happens after the suspected cause, with some exceptions that can be plausibly explained by other factors.

Absence of temporality strongly suggests lack of a causal relationship.

Extra: Google *Granger Causality*

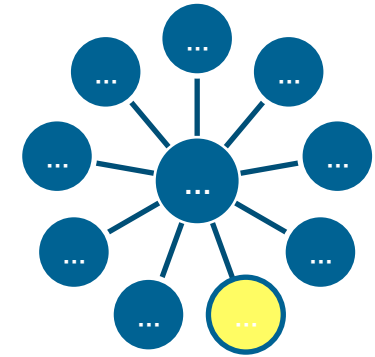
Dose-response

5

A strong and consistent dose-response gradient would contribute strong support to a cause-effect interpretation.

A reasonably consistent dose-response gradient might lend weaker support to a cause-effect interpretation.

Absence of a dose-response relationship does not rule out causality.

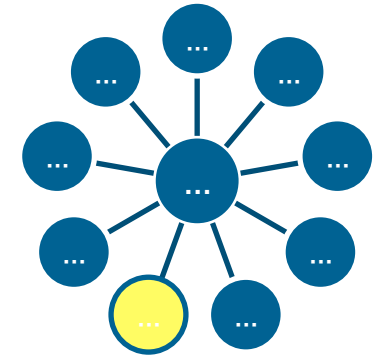


Dose-response gradients do not have to be linear.

For example, some interventions may be beneficial in small doses but harmful in higher doses (e.g., dietary iron).

Plausibility

6



What is biologically plausible depends on the biological knowledge of the day.

- Sir Austin Bradford Hill

A known mechanism linking the cause to the effect supports a causal interpretation.

If our cause-effect interpretation is consistent with a plausible theory of change, that would contribute weaker support to an interpretation of causality.

Absence of a plausible mechanism does not rule out a possible causal relationship.

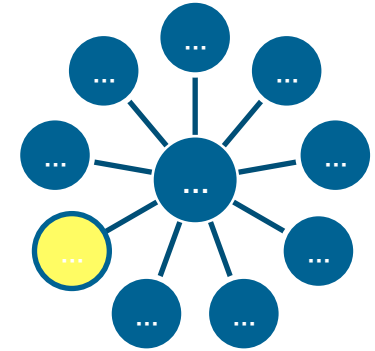
Coherence

7

This is the flipside to the plausibility criterion. A cause-effect interpretation that seriously conflicts with the existing base of high quality evidence would be called into question in the first instance, and would require a high standard of evidence to contribute to a change in the existing body of knowledge.

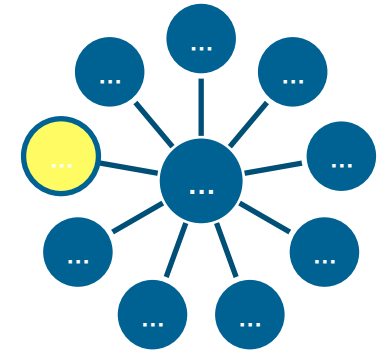
(Of course, we may have just made a groundbreaking discovery that turns science on its head. But here we can apply a Dilbert heuristic called the “which is more likely test” to help keep our feet on the ground).

A lack of conflict with the existing evidence base leaves the door open to a causal inference but does not strengthen it.



Experiment

8



In this context we are talking about experimentation in its broadest sense – e.g., not only experimental and quasi-experimental designs but also just changing something to see what happens, etc.

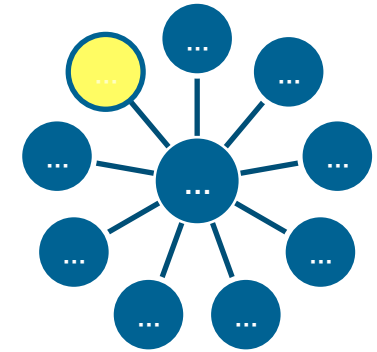
High quality RCT evidence would contribute strong support for our cause-effect interpretation. Note, however, that RCTs do not prove causality; they seek to minimise certain types of bias.

Other experimental evidence could lend weaker support for our causal interpretation.

If experimental evidence exists, and does not support our interpretation, that would contribute toward a judgment of non-causality.

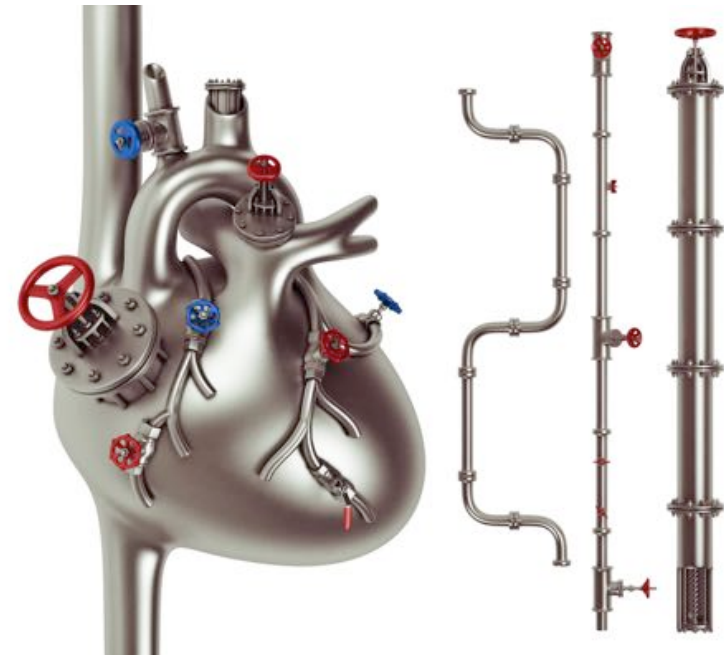
Analogy

9



Analogy contributes weak support for a causal inference. For example, if we know how a process works in one context we can use this knowledge as an explanatory ‘model’ to propose how it might work in a new context.

Absence of analogy does not rule out a possible causal relationship.



Rating guide (complete one of these for each competing causal interpretation)

	Strong support for causality	Weak support for causality	No support for or against causality	Support for non-causality
Strength				
Consistency				
Specificity				
Temporality				
Dose-response				
Plausibility				
Coherence				
Experiment				
Analogy				

This is how we used the Bradford Hill Criteria in a recent evaluation

Example

“The increase in participation satisfies Bradford Hill criteria of causation for: strength (a large shift in participation following the introduction of the program), consistency (the increase occurred in every region the program was introduced), plausibility (the increase in participation was an explicit outcome in the theory of change), and temporality (in each region, the increases in participation occurred after the program was introduced). The only other substantive initiative that might have contributed to the increase in participation was itself significantly influenced by this program.”



Closing remarks

Caution: Please read the operating manual carefully before using your new Bradford Hill Criteria. The manufacturer will not be held responsible for any risks to personal or public safety that arise from misapplication of the criteria.

Statistical significance doesn't really come into this

As Hill said,

No formal tests of significance can answer these questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of these effects. Beyond that they contribute nothing to the “proof” of our hypothesis.



Be cautious

Convincingly, though somewhat depressingly, Ioannidis (2005) argues that most research findings are probably false.

Causal inferences are, well... inferences.
Not real proof.

But the effects of our judgments can be very, very real.

Be careful out there.



Recommended reading

Be nerdy

Davidson, E.J., (2006).
*The RCTs only doctrine:
brakes on the acquisition
of knowledge?*

[http://journals.sfu.ca/
jmde/index.php/jmde_1/
article/viewFile/35/45](http://journals.sfu.ca/jmde/index.php/jmde_1/article/viewFile/35/45)

Scriven, M. (2008). A
*summative evaluation of
RCT methodology: & an
alternative approach to
causal research*

[http://journals.sfu.ca/
jmde/index.php/jmde_1/
article/viewFile/160/186](http://journals.sfu.ca/jmde/index.php/jmde_1/article/viewFile/160/186)

Lucas, R.M., McMichael,
A.J. (2005). *Association or
causation: evaluating links
between “environment
and disease”*

[http://www.who.int/
bulletin/volumes/
83/10/792.pdf](http://www.who.int/bulletin/volumes/83/10/792.pdf)

See also

Höfler, M. (2005). *The Bradford Hill Considerations on causality: a counterfactual perspective*

<http://www.ete-online.com/content/2/1/11>

Ward, A.C. (2009). *The role of causal criteria in causal inferences: Bradford Hill's "Aspects of Causation"*

<http://www.biomedcentral.com/1742-5573/6/2>

Davidson, E.J. (2005). *Evaluation methodology basics: the nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Chapter 7 offers additional strategies for dealing with causality.



© Julian King, 2015

Director, Julian King & Associates Limited

– a member of the Kinnect Group

Suggested citation: King, J. (2015). *Bradford Hill Criteria for Causal Inference*.

Auckland: Julian King & Associates. www.julianking.co.nz