

KNO-016 · How Evidence Is Built: Putting the Tools Together

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

Evidence is not a substance that appears automatically when data accumulate. What has to be connected for an observation to bear on a scientific claim? This article follows that question from claim formulation and measurement through study design, modelling, causal identification, uncertainty, replication, reproducibility, triangulation, directness, transparency and collective scrutiny. It asks which links are empirical, which depend on assumptions, which can be tested independently and which remain vulnerable to shared error. It also asks why more studies do not necessarily mean stronger evidence if they repeat the same weaknesses, and why evidence can become more reliable without becoming certain. The aim is to make the architecture of evidential support explicit enough that another reader can see not only what conclusion is proposed, but how each step is supposed to support it.

In 60 Seconds

Evidence becomes stronger when the links between claim, observation, design, model and inference are explicit enough to test. A precise claim must be vulnerable to failure. Measurements are produced by instruments and procedures, so their calibration and transformation matter. Study design determines which alternative explanations can be excluded; models connect observations to quantities or mechanisms; causal claims require assumptions beyond association; uncertainty belongs inside the result rather than after it. Replication and reproducibility probe different links in the chain. Triangulation is most informative when methods fail in different ways rather than sharing the same bias. Direct evidence is usually easier to interpret than a long chain of proxies, but indirect evidence can still be strong when each link is independently supported. Transparency makes those links inspectable; peer review and consensus can challenge them but cannot replace them. Evidence is therefore an architecture of support whose strength depends on the weakest relevant links, not simply on how many studies or citations exist.

Why This Question Matters

Scientists rarely observe a conclusion directly. They observe traces, measurements or outcomes and connect them to a claim through design, models, assumptions and background knowledge. When those links remain implicit, disagreement becomes difficult to diagnose and replication failures become hard to interpret. Making the evidential chain explicit clarifies where a result is

strong, where it depends on assumptions and what new observation could change it. That matters not only for research practice but also for anyone deciding how much confidence to place in a scientific claim.

1. Evidence begins with a claim precise enough to fail

Before asking whether evidence is strong, one must know what claim is being supported.

[C01] A scientific claim should identify its target sufficiently clearly that observations or analyses could count against it. Broad narratives that change meaning after each result are difficult to test because the evidential target moves. [S01, S02]

Operational definitions are not mere paperwork. They connect theoretical language to observable procedures. The same word can refer to different estimands, measurements or classifications. If the target shifts between design, analysis and interpretation, the apparent evidence may answer a different question from the one originally posed.

Where That Leaves Us

Before asking whether evidence is strong, one must know what claim is being supported.

Operational definitions are not mere paperwork. They connect theoretical language to observable procedures. The same word can refer to different estimands, measurements or classifications. The next section examines the proposition that observations are produced, not simply found.

2. Observations are produced, not simply found

KNO-008 showed that instruments extend the senses; KNO-011 showed that numbers depend on measurement conventions. Here those lessons become part of the evidence chain.

[C02] The evidential value of a measurement depends on how the measurand or variable was defined, how the instrument or procedure was calibrated, what corrections were applied and what uncertainty remains. A numerical output is not self-validating. [S03]

A photograph, spectrum, assay result or survey response is already the result of choices about sampling, resolution, processing and classification. Evidence begins before statistical analysis.

Where That Leaves Us

KNO-008 showed that instruments extend the senses; KNO-011 showed that numbers depend on measurement conventions. Here those lessons become part of the evidence chain. A photograph, spectrum, assay result or survey response is already the result of choices about sampling, resolution, processing and classification. The next section examines the proposition that design determines which alternatives can be excluded.

3. Design determines which alternatives can be excluded

Different questions require different designs.

[C03] Experimental controls, randomization, blinding, repeated measurement and observational comparison structures address different potential biases. None is a universal marker of quality independent of the scientific question. [S04]

For causal questions, design determines which counterfactual contrasts are plausibly identified. For descriptive questions, representative sampling may matter more than random treatment allocation. For historical sciences, convergent traces and mechanistic consistency can carry the evidential load when intervention is impossible.

Where That Leaves Us

Different questions require different designs. For descriptive questions, representative sampling may matter more than random treatment allocation. For historical sciences, convergent traces and mechanistic consistency can carry the evidential load when intervention is impossible. Its value lies in making this particular inferential step inspectable rather than treating the method as self-validating. The next section examines the proposition that models are bridges between observations and claims.

4. Models are bridges between observations and claims

Data rarely speak directly to the proposition of interest. A model translates observations into quantities, mechanisms or predictions.

[C04] Every model-based evidential claim is conditional on modelling choices to some degree. Good practice therefore distinguishes what is directly observed from what is inferred through the model and tests sensitivity to plausible alternatives. [S06]

This is not a defect unique to complex simulations. Even a straight-line fit, a calibration curve or a diagnostic threshold is a model.

Where That Leaves Us

Data rarely speak directly to the proposition of interest. This is not a defect unique to complex simulations. Even a straight-line fit, a calibration curve or a diagnostic threshold is a model. What matters here is the stated relation between evidence, method and inference, not a generic promise of certainty. The next section examines the proposition that causal evidence requires more than association.

5. Causal evidence requires more than association

KNO-014 established the central causal boundary.

[C05] An observed association can support a causal claim only through assumptions and design features that address confounding, reverse causation, selection and other competing explanations. The strength of causal evidence therefore depends on the identification strategy, not simply on the size or statistical significance of the association. [S04, S07]

Different causal methods can be combined, but multiple analyses that share the same hidden bias are not independent confirmation.

Where That Leaves Us

KNO-014 established the central causal boundary. Different causal methods can be combined, but multiple analyses that share the same hidden bias are not independent confirmation. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines the proposition that uncertainty belongs inside the claim.

6. Uncertainty belongs inside the claim

KNO-015 treated uncertainty as a methodological object rather than a decorative interval.

[C06] Evidence assessment should distinguish uncertainty arising from measurement, sampling, modelling and other recognized sources. A narrow within-model interval does not establish that unmodelled systematic error is small. [S03, S08]

A strong claim therefore states not only an estimate but also the conditions under which the estimate is expected to be reliable.

Where That Leaves Us

KNO-015 treated uncertainty as a methodological object rather than a decorative interval. A strong claim therefore states not only an estimate but also the conditions under which the estimate is expected to be reliable. Its value lies in making this particular inferential step inspectable rather than treating the method as self-validating. The next section examines the proposition that replication and reproducibility test different links.

7. Replication and reproducibility test different links

The terms are often blurred.

[C07] The National Academies distinguishes computational reproducibility, obtaining consistent computational results using the same data, code and methods, from replicability, obtaining consistent results across studies addressing the same scientific question with new data. [S01]

[C08] A successful replication increases confidence but does not prove an original claim universally correct; a failed replication does not by itself conclusively refute it because differences can arise from genuine heterogeneity, uncontrolled variables, methodological defects or other causes. [S01]

This asymmetry matters. Repetition is evidence about robustness, not a logical proof operator.

Where That Leaves Us

The terms are often blurred. Repetition is evidence about robustness, not a logical proof operator. What matters here is the stated relation between evidence, method and inference, not a generic promise of certainty. [S01] A successful replication increases confidence but does not prove an original claim universally correct. The next section examines the proposition that triangulation is strongest when errors differ.

8. Triangulation is strongest when errors differ

Independent evidence does not mean simply “more studies.”

[C09] Triangulation gains force when different methods address the same question while having different key sources of bias. Convergence is more informative when the biases would not be expected to drive all approaches toward the same result. [S07]

Ten studies using the same instrument, same selection mechanism and same model may reproduce a shared bias. A smaller set using genuinely different designs can be more diagnostically valuable.

Where That Leaves Us

Independent evidence does not mean simply “more studies.” Ten studies using the same instrument, same selection mechanism and same model may reproduce a shared. A smaller set using genuinely different designs can be more diagnostically valuable. The next section examines the proposition that evidence is cumulative, but accumulation needs structure.

9. Evidence is cumulative, but accumulation needs structure

Meta-analysis, systematic review and evidence frameworks can combine studies, yet combination is only meaningful when the underlying questions are sufficiently compatible.

[C10] Evidence frameworks such as GRADE explicitly examine risk of bias, inconsistency, indirectness, imprecision and dissemination/publication bias rather than treating study count alone as certainty. [S09]

The point is not that GRADE governs every scientific field. It demonstrates a general architecture: bodies of evidence must be evaluated for systematic error, coherence, relevance and precision, not merely aggregated.

Where That Leaves Us

Meta-analysis, systematic review and evidence frameworks can combine studies, yet combination is only meaningful when the underlying questions are sufficiently compatible. The point is not that GRADE governs every scientific field. It demonstrates a general architecture: bodies of evidence must be evaluated for systematic error, coherence, relevance and precision, not merely aggregated. The next section examines the proposition that directness matters.

10. Directness matters

Evidence can be accurate yet answer the wrong question.

A laboratory surrogate may not predict a clinical outcome. A result in one population may not transport to another. A proxy variable may measure only one component of the intended construct.

[C11] The concept of indirectness formalizes mismatches between the question of interest and the populations, interventions or exposures, comparators and outcomes actually studied. [S09, S10]

External validity and transportability therefore belong to evidence assessment, not to an afterthought after internal validity has been established.

Where That Leaves Us

Evidence can be accurate yet answer the wrong question. A laboratory surrogate may not predict a clinical outcome. A result in one population may not transport to another. A proxy variable may measure only one component of the intended construct. The next section examines the proposition that transparency changes what others can test.

11. Transparency changes what others can test

Evidence is stronger when the path from raw material to conclusion can be inspected.

[C12] Clear reporting of methods, data-processing steps, analytic choices and uncertainty supports computational checking and helps other investigators identify errors or test robustness. Reproducibility is facilitated by access to the relevant digital artifacts and sufficiently complete methods. [S01, S02]

Transparency does not make a flawed analysis correct. It makes flaws more discoverable.

Where That Leaves Us

Evidence is stronger when the path from raw material to conclusion can be inspected. Transparency does not make a flawed analysis correct. It makes flaws more discoverable. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines the proposition that peer review and consensus are filters, not substitutes for evidence.

12. Peer review and consensus are filters, not substitutes for evidence

Peer review can detect errors, request clarification and improve reporting, but it does not independently reproduce the underlying experiment. Consensus can summarize a mature evidential landscape, but shared assumptions can also propagate through a field.

These institutional processes belong fully to KNO-020. Here the boundary is simpler: social validation can strengthen quality control, but the evidential chain still has to connect observations to claims.

Where That Leaves Us

Peer review can detect errors, request clarification and improve reporting, but it does not independently reproduce the underlying experiment. Consensus can summarize a mature evidential landscape, but shared assumptions can also propagate through a field. These institutional processes belong fully to KNO-020. The next section examines a practical evidence audit.

13. A practical evidence audit

Before accepting a strong scientific conclusion, ask:

1. **What exactly is the claim?** Descriptive, predictive, causal, mechanistic or historical?
2. **What observations anchor it?** What was directly measured or recorded?
3. **How were variables constructed?** What classifications, thresholds or transformations intervened?
4. **What design addresses alternatives?** Controls, randomization, matching, temporal ordering, natural variation, comparative cases?
5. **What model connects data to claim?** Which assumptions are essential?
6. **What uncertainties are quantified?** Which remain recognized but unquantified?
7. **What biases could move the answer?** In what direction?
8. **Has the computation been reproduced?** Can the analytic path be rerun?
9. **Has the finding been replicated or triangulated?** Are the new failure modes genuinely different?
10. **How direct is the evidence for this target population, quantity or mechanism?**

11. **Is the result consistent with a broader body of evidence?** If not, is the disagreement informative?
12. **Which conclusion is actually warranted?** Avoid upgrading “compatible with” into “proved.”

Where That Leaves Us

Before accepting a strong scientific conclusion, ask: What uncertainties are quantified? Which remain recognized but unquantified? What observations anchor it? What was directly measured or recorded? How were variables constructed? What classifications, thresholds or transformations intervened? The next section examines the proposition that evidence can become stronger without becoming certain.

14. Evidence can become stronger without becoming certain

Science accumulates constraint. Measurements narrow possible values; experiments eliminate alternatives; models organize implications; independent methods test different vulnerabilities; replication probes robustness; synthesis shows where results cohere and where they do not.

The endpoint is rarely certainty. It is a claim whose remaining alternatives have become progressively harder to sustain.

Where That Leaves Us

Measurements narrow possible values; experiments eliminate alternatives; models organize implications; independent methods test different vulnerabilities; replication probes robustness; synthesis shows where results cohere and where they do not. The endpoint is rarely certainty. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

Boundary note

KNO-016 is deliberately a **synthesis article**. It does not replace the detailed methods in KNO-008 through KNO-015, and it does not pre-empt KNO-020 on peer review and consensus or KNO-021 on open science. Its role is to show how measurement, experimental design, modelling, causality, uncertainty, reproducibility, replication and triangulation combine into an auditable evidential argument.

What We Know, and How

Evidence is built by linking precise claims to observations through measurement, design, models and inferential assumptions. Replication, reproducibility, triangulation, directness and transparency strengthen different links, especially when their failure modes are not shared.

What We Still Don't Know

No audit can guarantee that every relevant alternative has been imagined, that every source of bias is independent or that a well-supported model is the only possible explanation. Evidence can therefore become stronger without becoming logically final.

What Could Change Our Minds

The evidential assessment should change when a key measurement fails calibration, a design permits an alternative explanation, a model prediction breaks, a replication fails under genuinely comparable conditions, or a new method with different error structure contradicts the existing result.

If You Remember Only One Thing

Strong evidence is not a pile of results; it is a chain of support whose assumptions, error modes and tests can be inspected.

Related Articles

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-014 · Causality and Correlation
- KNO-015 · Uncertainty and Probability
- KNO-020 · Review, Consensus and Disagreement

References

- [S01] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S02] National Academies of Sciences, Engineering, and Medicine. “Understanding Reproducibility and Replicability,” chapter 3 in *Reproducibility and Replicability in Science*, pp. 39–54. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S03] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. JCGM 100:2008(E). DOI: 10.59161/JCGM100-2008E.
- [S04] Hernán, Miguel A. “A Definition of Causal Effect for Epidemiological Research.” *Journal of Epidemiology and Community Health* 58(4), 2004, 265–271. DOI: 10.1136/jech.2002.006361.
- [S05] Schulz, Kenneth F.; Altman, Douglas G.; Moher, David; CONSORT Group. “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.” *BMJ* 340, 2010, c332. DOI: 10.1136/bmj.c332.
- [S06] Frigg, Roman; Hartmann, Stephan. “Models in Science.” *The Stanford Encyclopedia of Philosophy*, Summer 2025 Edition; substantive revision 2 April 2025. <https://plato.stanford.edu/archives/sum2025/entries/models-science/>
- [S07] Lawlor, Debbie A.; Tilling, Kate; Davey Smith, George. “Triangulation in aetiological epidemiology.” *International Journal of Epidemiology* 45(6), 2016, 1866–1886. DOI: 10.1093/ije/dyw314.
- [S08] Mastrandrea, Michael D.; Field, Christopher B.; Stocker, Thomas F.; Edenhofer, Ottmar; Ebi, Kristie L.; Frame, David J.; Held, Hermann; Kriegler, Elmar; Mach, Katharine J.; Matschoss, Patrick R.; Plattner, Gian-Kasper; Yohe, Gary W.; Zwiers, Francis W. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC, 2010. <https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf>

[S09] Neumann, Ignacio; Brennan, Sue; Meerpohl, Joerg; Davoli, Marina; Alonso Coello, Pablo; Akl, Elie; Dahm, Philipp; Skoetz, Nicole; Xia, Jun; Brozek, Jan; Schünemann, Holger J. “Overview of the GRADE approach.” *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/overview-of-the-grade-approach>

[S10] “Indirectness.” *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/indirectness>. Author display currently contains an apparent duplicated Holger J. Schünemann entry; author chain therefore remains P pending human confirmation and is not normalised in this bibliography.