

Scientific Inference and Experimental Reasoning: Master Index and Scope

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. All Rights Reserved.

Correspondence: info@3Dresyns.com · www.3Dresyns.com

INF-000 · Scientific Inference and Experimental Reasoning · Version v0.2

Abstract

Scientific conclusions are not determined by data alone. The evidential force of an observation depends on the claim being made, the process that generated and selected the data, the causal or mechanistic structure assumed, the rival explanations still compatible with the observations, the capability of proposed discriminating tests, the uncertainty carried by both parameters and model structure, the robustness of the conclusion to defensible analytical choices, and the validity of any transfer from the studied conditions to the intended target. This article establishes the scope and architecture of the Scientific Inference and Experimental Reasoning (INF) series. INF is a constructive methodological series concerned with the logic by which evidence supports a justified scientific claim. It is neither a catalogue of experimental failure modes nor a general statistics or philosophy-of-science course. The series is organised around a canonical sequence: first define the claim type; then identify admissible evidence; specify competing explanations; seek discriminating evidence; characterise uncertainty and robustness; and only then state the conclusion justified by the evidence. INF-000 defines the terminology, ownership boundaries and article map for INF-001 through INF-013. Substantive development of each inferential problem remains with its owning article.

Keywords: scientific inference; experimental reasoning; causal inference; identifiability; rival explanations; evidence; uncertainty; robustness; analytic multiplicity; triangulation; generalisability; transportability

1. Purpose and scope

The purpose of INF is to make explicit the inferential structure that lies between an observation and a scientific claim. A measurement can be precise without identifying a mechanism; an association can be reproducible without being causal; a model can fit observed data while remaining non-unique; a result can be statistically stable while depending on a selected analysis path; and a conclusion valid in one study population can fail when transported to another target. These are not variations of a single problem. They are distinct inferential questions that require different evidence.

INF therefore asks one central question: how should a scientific claim be justified when observations are incomplete, indirect, model-dependent, selected, uncertain or compatible with more than one explanation? The series treats measurement validity and experimental validity as prerequisites, then concentrates on the organisation of evidence from claim definition to justified conclusion.

The architecture is intentionally finite for this phase. INF-000 defines the series; INF-001 through INF-012 own distinct inferential problems; INF-013 is the synthesis and is written last. A future INF-014 or later article would require a demonstrated architecture-level gap rather than thematic proximity alone.

2. The canonical inference chain

The series uses the following canonical sequence, expressed in words rather than as a deterministic pipeline: define the claim type; determine what evidence is admissible for that claim; identify competing explanations or models; obtain evidence capable of discriminating among those alternatives; characterise structural and parametric uncertainty and test robustness to defensible assumptions and specifications; then state the conclusion that the evidence actually justifies.

The order matters because later operations cannot repair an earlier category error. No amount of statistical refinement converts a descriptive association into a causal effect without causal assumptions and design. No goodness-of-fit statistic makes a model mechanistically unique when rival structures remain observationally compatible. No robustness analysis rescues a conclusion if every analysed specification shares the same false assumption. Conversely, a conclusion may remain justified even when uncertainty is substantial, provided the claim is correspondingly narrow and the uncertainty is made explicit.

3. Claim classes

INF-001 develops the claim taxonomy in detail. INF-000 uses four broad classes because the evidence required for one class cannot automatically be substituted for another:

- Descriptive claims concern what was observed or measured in a defined population, specimen set, condition or time window.
- Predictive claims concern the ability of a model or procedure to predict outcomes under specified conditions, including explicit evaluation outside the data used to fit or choose the model.
- Causal claims concern the effect of an intervention, exposure or change, requiring a target causal contrast and assumptions or design features sufficient to identify it.

Scientific Inference and Experimental Reasoning · INF-000

- Mechanistic claims concern the process or structure that produces an outcome; predictive success or causal effect estimation alone need not identify that mechanism.

These classes can coexist in one investigation, but they should not be collapsed. A model may predict well without representing the true mechanism; an intervention may establish that a factor changes an outcome without identifying every intermediate mechanism; a mechanistic model may explain observed behaviour while still requiring independent validation of its quantitative predictions.

4. Mandatory terminology

INF uses ordinary scientific terms in deliberately non-interchangeable ways. Observation, measured quantity or measurand, data, model, hypothesis, mechanism, prediction, explanation, claim, evidence, uncertainty and inference should be distinguished whenever the distinction affects what can be concluded.

Term	Operational definition in INF
Confounder	Under a specified causal structure, a variable that is a common cause of the exposure or intervention and outcome, or otherwise lies on an open non-causal path between them that must be blocked for the target causal contrast. A variable is not a confounder merely because adjustment changes an estimate.
Effect modifier	A variable across whose levels the magnitude or form of a target effect differs. Effect modification is a property of a defined effect contrast and target context; it is not synonymous with confounding.
Collider	A variable that is a common effect of two or more variables on a specified path. Conditioning on a collider, or on its descendant, can open a non-causal association; collider status is path- and graph-relative rather than an intrinsic label of the variable.
Identifiability	Whether a target causal, statistical or mechanistic quantity is uniquely determined by the stated model, assumptions and idealised observable distribution. Practical identifiability concerns whether finite, noisy data can estimate an identifiable target with useful precision.
Generalisability / transportability	Generalisability concerns extending an inference from a study sample to a target population drawn from the same underlying setting; transportability concerns extending it across settings or populations whose relevant causal or distributional structure may differ. Both require an explicit target and assumptions about what is invariant and what may change.

5. Evidence architecture

Each substantive INF article develops part of a discipline that requires an analysis to state seven things explicitly: the kind of claim being assessed; the evidence relevant to that claim; the alternative explanation or model that remains live; the observation or intervention that would discriminate among alternatives; which uncertainty is parametric and which is structural; which assumptions the conclusion depends on; and what the evidence does not establish.

This requirement is deliberately stricter than citation density. A reference supports only the sentence or clause it actually addresses. A large bibliography cannot compensate for an unsupported inference. General methodological claims should be anchored in the broader methodological literature rather than supported solely by the author's own corpus. Where material competing interpretations exist, credible alternatives belong in the source set.

The distinction between evidence and decision threshold is also important. A p value, confidence interval, Bayes factor, likelihood ratio, posterior probability, prediction error or model-selection score can contribute information under a specified inferential framework; none is a universal replacement for defining the scientific claim and the assumptions that connect the statistic to that claim. Misinterpretations of significance tests and intervals are well documented [8], while flexible analysis choices can materially change apparent support [4-6].

6. Rival explanations and discriminating evidence

Scientific inference becomes most informative when alternatives are stated strongly enough that different observations would count differently against them. Platt's formulation of strong inference emphasised explicit alternative hypotheses and experiments capable of excluding alternatives [9]. Modern causal and statistical frameworks make the same requirement more precise in different domains: the target must be identified under a stated model and assumptions, and the proposed test must have the capability to distinguish the alternatives it is invoked to separate [1-3].

INF does not treat falsification as a universal one-step rule. Auxiliary assumptions, measurement capability, model misspecification and limited test severity can make an apparent failure ambiguous. INF-004 therefore owns rival explanations and identifiability, while INF-007 owns the design of discriminating experiments and operational severe tests. INF-008 then owns qualification of the conclusion when more than one structure remains live.

7. Robustness, multiplicity and evidence integration

Robustness and analytic multiplicity are related but not identical. INF-009 asks whether a conclusion survives a declared set of defensible alternatives in assumptions, models, preprocessing and specifications. Multiverse and specification-curve approaches are examples of procedures that expose how conclusions vary across reasonable analytical choices [5,6]. Agreement across variants is informative only to the extent that the variants do not all share the same material false assumption.

INF-010 instead owns undisclosed or unplanned selection among analysis paths: researcher degrees of freedom, multiple testing, flexible stopping and forking paths. These processes can inflate apparent support when only favourable paths are reported [4]. Preregistration and transparent analysis plans can reduce some forms of this flexibility, but they do not themselves establish that the scientific model or measurement system is valid.

Scientific Inference and Experimental Reasoning · INF-000

INF-011 addresses a different problem: integration of heterogeneous evidence. Convergence across measurements can strengthen an inference when the evidence streams depend on different assumptions and failure modes. Apparent convergence is weaker when supposedly independent measurements share the same calibration, selection mechanism, model assumption or data source.

8. Generalisation and the target of inference

A justified statement must specify where it is intended to hold. Internal validity for a study does not automatically provide external validity for another population, scale, environment or operating condition. INF-012 owns the explicit target, effect modification, generalisability and transportability. Contemporary methodological work distinguishes extension from a study sample to a target population in the same underlying setting from transport across settings whose relevant structures may differ [7]. The inferential burden therefore includes stating which relationships are assumed invariant and which may change.

9. Series map and ownership

Code	Title	Primary ownership
INF-000	Scientific Inference and Experimental Reasoning: Master Index and Scope	Series architecture, terminology, claim classes, boundaries and article map.
INF-001	The Claim Comes First: Descriptive, Causal, Mechanistic and Predictive Claims Need Different Evidence	Claim taxonomy and evidential adequacy by claim type.
INF-002	Causal Structure: Confounders, Colliders, Mediators and What Intervention Adds	Causal roles, identification logic, observational versus interventional evidence, and measurement-error distortion of effect estimates.
INF-003	Which Data Reached the Inference: Selection, Missingness, Survivorship and Collider Bias	Selection into the analysed dataset, including missingness, survivorship and selection on a collider.
INF-004	Rival Explanations: Identifiability, Underdetermination and Auxiliary Hypotheses	Which rival structures remain live and what evidence would separate them.
INF-005	Strength of Evidence: Likelihood and How Much a Result Should Change a Belief	Evidential weight between claims or hypotheses without reducing inference to binary significance.
INF-006	Prior Knowledge and Base Rates: Necessary Constraint or Hidden Bias	Background knowledge, base rates and prior constraints in inference.
INF-007	Discriminating Experiments: Severe Tests and Operational Falsification	Measurements or interventions whose possible outcomes differ under rival explanations.
INF-008	Structural versus Parametric Uncertainty	Qualification of conclusions while multiple structures remain live; distinction from parameter uncertainty within one structure.
INF-009	Robustness of Conclusions: Sensitivity to Assumptions, Models and Specifications	Declared robustness analyses across defensible variants, including multiverse and specification-curve approaches.
INF-010	Analytic Multiplicity: Many Paths from the Same Data	Forking paths, researcher degrees of freedom, multiple testing, flexible stopping and selective analysis.
INF-011	Evidence Integration: Triangulation Across Measurements That See Different Things	Integration of heterogeneous evidence and dependence among evidence streams.
INF-012	Generalisation: From the Conditions Studied to the Target	External validity, generalisability, transportability and effect modification.
INF-013	From Observation to a Justified Claim: Synthesis of Scientific Inference and Experimental Reasoning	Final synthesis linking the full evidence-to-claim architecture; drafted last.

10. Boundaries with existing series

10.1 Hidden Error Budget (HEB)

HEB is diagnostic: it identifies ways in which apparently valid experimental knowledge can fail because measurement, reference, observer, replication, correction, interaction, time, scale, model, classification or transfer were mishandled. INF is constructive: it begins once measurement and experimental validity are treated as prerequisites and asks how evidence should be organised to support a claim. HEB-016 remains the synthesis and closure of HEB; INF does not reopen that catalogue under new names.

10.2 Ways of Knowing (KNO)

KNO-012 through KNO-016 own broad, accessible treatments of experiments, models, causality, uncertainty and evidence construction. INF operates at a more formal methodological level. It may analyse identifiability, causal variable roles, evidential strength, discriminating tests, structural uncertainty, robustness, multiplicity and transportability, but it does not replace KNO's explanatory or historical function.

10.3 Fundamental Ideas (IDE)

IDE owns foundational concepts such as causality, probability, measurement, model and uncertainty as concepts. INF uses those concepts operationally inside an inference architecture and does not redefine their general philosophical or scientific meaning.

10.4 Engineering Review ER-184

ER-184 owns diagnostic inference applied specifically to vat photopolymerisation, including competing defect mechanisms and discriminating measurements in a VPP failure-analysis context. INF-004 and INF-007 own only the discipline-general logic of rival explanations and discriminating tests. INF must not reproduce VPP-specific diagnostic branches, process thresholds or engineering fault mappings.

11. What constitutes a justified claim in INF

INF does not define a justified claim as a claim that is certain, statistically significant, widely believed, compatible with one fitted model or supported by a large number of citations. A claim is justified to the extent that its type is explicit, the evidence is relevant to that type, material alternatives have been represented fairly, the data-selection and causal structure are understood well enough for the target inference, discriminating evidence has adequate capability, structural and parametric uncertainty are not conflated, sensitivity to defensible assumptions is reported, analytical flexibility has not been hidden, evidence streams are integrated with their dependencies exposed, and any move from the studied conditions to a target is separately justified.

The permitted endpoint is therefore not always a unique explanation. A scientifically adequate conclusion can be that one explanation is retained, several remain compatible with the evidence, one or more alternatives are excluded under stated assumptions, a quantitative estimate is too uncertain for the intended claim, or the available evidence is insufficient. Preserving those distinctions is the central discipline of the INF series.

References

- [1] Pearl, J. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511803161.
- [2] Hernán, M.A.; Robins, J.M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. DOI: 10.1201/9781315374932.
- [3] Mayo, D.G. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press, 2018. DOI: 10.1017/9781107286184.
- [4] Simmons, J.P.; Nelson, L.D.; Simonsohn, U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science* 22(11), 1359-1366 (2011). DOI: 10.1177/0956797611417632.
- [5] Steegen, S.; Tuerlinckx, F.; Gelman, A.; Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science* 11(5), 702-712 (2016). DOI: 10.1177/1745691616658637.
- [6] Simonsohn, U.; Simmons, J.P.; Nelson, L.D. Specification curve analysis. *Nature Human Behaviour* 4, 1208-1214 (2020). DOI: 10.1038/s41562-020-0912-z.
- [7] Degtiar, I.; Rose, S. A Review of Generalizability and Transportability. *Annual Review of Statistics and Its Application* 10, 501-524 (2023). DOI: 10.1146/annurev-statistics-042522-103837.
- [8] Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. *European Journal of Epidemiology* 31, 337-350 (2016). DOI: 10.1007/s10654-016-0149-3.
- [9] Platt, J.R. Strong Inference. *Science* 146(3642), 347-353 (1964). DOI: 10.1126/science.146.3642.347.

Data and materials statement

No new experimental dataset was generated for INF-000. The article defines the architecture and scope of the INF series and relies on the cited methodological literature together with the controlled ownership boundaries of the existing 3Dresyns publication corpus.