

Robustness of Conclusions: Sensitivity to Assumptions, Models and Specifications

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. All Rights Reserved.

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

INF-009 · Scientific Inference and Experimental Reasoning · Version v0.2

Abstract

A conclusion is robust only relative to a declared set of plausible assumptions, models, data-processing choices and analytical specifications. Robustness is therefore not a synonym for truth, replication or statistical significance. This article defines robustness as the stability of a scientific conclusion across a prospectively or transparently declared family of defensible alternatives. It distinguishes parameter sensitivity, model sensitivity, specification-curve analysis and multiverse analysis, and explains why hidden researcher flexibility belongs instead to INF-010. The central requirement is that the alternative specifications be scientifically admissible rather than generated merely to make a result look stable. Robustness can strengthen confidence when materially different but justified specifications converge; instability can reveal that the scientific conclusion depends on choices whose uncertainty must be reported. Neither outcome establishes the mechanism or causal structure by itself.

Keywords: robustness; sensitivity analysis; specification curve; multiverse analysis; model dependence; assumptions; scientific inference

1. Robustness is conditional stability

Every empirical conclusion depends on assumptions: how variables are defined, which observations are included, which functional form is fitted, which covariates are adjusted for, how missing values are handled, and which model class is treated as plausible. A robustness analysis asks whether the scientific conclusion survives across a declared family of alternatives that are each defensible for the stated claim. It does not ask whether every imaginable analysis produces the same answer.

2. Robustness is not truth

A conclusion can be stable across many related models and still be wrong if the entire family shares the same structural defect. Conversely, a scientifically real effect can appear unstable when the available data are weak, measurement error is large or reasonable models target different estimands. INF therefore prohibits the synonym collapse robustness = truth. Robustness is evidence about dependence on assumptions, not an ontological guarantee.

3. Four distinct robustness questions

- Parameter sensitivity: does the conclusion change when uncertain parameter values or nuisance inputs vary over scientifically plausible ranges?
- Model sensitivity: does the conclusion change across materially different model structures that are each compatible with the scientific question?
- Specification sensitivity: does the conclusion change across defensible choices of coding, transformation, adjustment, inclusion or functional form?
- Target sensitivity: does the conclusion change when the estimand, target population or time horizon changes? A change here may reflect a different question rather than fragility.

4. Specification curves and multiverse analysis

Specification-curve analysis makes a prespecified or transparently justified set of non-redundant specifications visible and evaluates how results vary across them [1]. Multiverse analysis similarly exposes consequences of alternative data-processing and analytical choices by performing analyses across a set of reasonable scenarios [2]. These tools are valuable because they reveal dependence that a single preferred analysis hides. Their validity still depends on how the admissible specification set is defined.

5. What belongs in the admissible set

An alternative belongs in a robustness analysis when it is scientifically defensible before looking at whether it supports the preferred conclusion. Alternatives that change the claim, estimand or target population should be labelled as different questions. Alternatives known to violate assumptions should not be counted merely to enlarge the curve. The analyst should state why each branch is admissible and which assumptions differ across branches.

6. Distinction from INF-010

INF-009 owns transparent robustness analyses in which the analyst deliberately exposes dependence on declared alternatives. INF-010 owns analytic multiplicity when many possible paths exist but only a subset is selected, emphasised or reported after seeing the results. The same set of analytical choices can therefore be used constructively in INF-009 or

become a bias mechanism in INF-010 depending on whether the multiplicity is prospectively controlled or transparently disclosed.

7. Interpretation discipline

- Convergence across plausible specifications strengthens the claim only against the particular choices varied.
- Instability identifies a dependency that belongs in the uncertainty statement; it does not automatically invalidate every specification.
- A robustness analysis should report the distribution and pattern of results, not only the proportion crossing a significance threshold [3].
- Robustness to statistical specification does not establish causal identification, mechanism or generalisability.
- If all specifications share the same measurement defect or structural assumption, robustness cannot diagnose that common failure.

8. Minimum reporting

- State the scientific claim and estimand held fixed.
- Define the admissible specification family before interpreting its outcomes.
- Explain why each major branch is scientifically plausible.
- Report which decisions drive material changes.
- Separate changes caused by different estimands from changes within one estimand.
- State explicitly which assumptions were not varied.

References

[1] 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.

[2] 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.

[3] 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.

Data and materials statement

No new experimental dataset was generated for this article. The article is a methodological analysis based on the cited literature and the frozen ownership architecture of the INF series.