

KNO-015 · Uncertainty and Probability

KNO · Ways of Knowing

Juan Segurola

v0.7 · EN

All Rights Reserved

Ways of Knowing

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

Article Abstract

Scientific results often come with intervals, probabilities, error bars or sensitivity analyses, but what kinds of uncertainty do those devices actually represent? This article asks how measurement uncertainty, sampling variation, parameter uncertainty, model uncertainty, structural uncertainty and expert judgement differ, and what happens when they are collapsed into a single number. It examines confidence intervals, credible intervals, p-values, propagation, prediction and sensitivity analysis while asking which statements depend on a chosen statistical framework and which limits remain outside it. It also asks how to communicate uncertainty without converting ignorance into false precision or treating every unknown as quantifiable. The central problem is methodological: what representation of uncertainty matches the claim being made, and what remains unrepresented even after the calculation is complete?

In 60 Seconds

Uncertainty is not one thing. A measurement can be uncertain because of calibration, repeatability or resolution; an estimate can vary across samples; model parameters can be imprecise; structural choices can change the result; future observations can add prediction uncertainty; and some unknowns lie outside the probability model entirely. Confidence intervals and Bayesian credible intervals answer different questions because they rely on different inferential frameworks. A p-value is not the probability that a hypothesis is true. Propagating uncertainty through a model is necessary when uncertain inputs affect outputs, but propagation cannot represent omitted mechanisms or an inappropriate model structure. Sensitivity analysis asks how conclusions change when assumptions or inputs vary, making hidden dependencies visible. Expert judgement can be structured, documented and calibrated, but it does not become observational data. Good uncertainty practice therefore matches each representation to the claim, separates quantified variability from structural ignorance, and states what the chosen analysis does not cover.

Why This Question Matters

Scientific decisions are often made under uncertainty, but different uncertainties call for different responses. A narrow confidence interval can coexist with model misspecification; a precise instrument can measure the wrong quantity; a probability can be well calculated while excluding the most important unknown. Treating every uncertainty as interchangeable produces false precision and misleading comparisons. Distinguishing what is measured, what is sampled, what is modelled and what remains outside the model is essential for interpreting evidence responsibly and for communicating how strongly a conclusion should guide action.

1. Uncertainty is not one substance

The word covers several mechanisms that should not be collapsed into one residual category.

[C01] Measurement uncertainty concerns incomplete knowledge of the value of a measurand given the measurement procedure, inputs, corrections and model. The GUM framework establishes general rules for evaluating and expressing such uncertainty, while later JCGM documents extend the framework to distribution propagation and measurement models. [S01–S03]

[C02] Sampling uncertainty arises because a finite sample is only one realization from a broader data-generating process. Its statistical treatment depends on the sampling or assignment mechanism and on the estimator being studied. It is conceptually distinct from instrument calibration uncertainty or structural model uncertainty. [S01, S08]

[C03] Model uncertainty includes uncertainty about functional form, omitted structure, approximations and alternative plausible specifications. Structural uncertainty cannot always be reduced to a single standard error estimated inside one chosen model. [S07, S09]

These categories can interact. A calibration constant may enter a statistical model; a model may transform measured inputs; a sampling design may determine which parameters are estimable. The methodological requirement is not to force each source into an isolated box, but to prevent one quantified component from being mistaken for all uncertainty.

Where That Leaves Us

The word covers several mechanisms that should not be collapsed into one residual category. Measurement, sampling and model uncertainty are conceptually distinct, yet they can interact, as when a calibration constant enters a statistical model. What matters is that no quantified component is mistaken for all uncertainty. The next section examines the proposition that a measurement result is not just its displayed digits.

2. A measurement result is not just its displayed digits

A digital instrument can display 12.345678 without establishing knowledge to six decimal places.

[C04] In the GUM/NIST approach, individual standard uncertainty components are combined, including covariance where relevant, to obtain a combined standard uncertainty. Expanded uncertainty is then obtained as a coverage factor times the combined standard uncertainty when that reporting form is appropriate. [S01, S04]

[C05] A coverage factor such as $k = 2$ does not mechanically guarantee 95% coverage in every case. The probability interpretation depends on distributional and other assumptions, and small effective degrees of freedom can require different factors. [S04, S05]

This is a general lesson. Precision printed by software is formatting. Evidential precision comes from the uncertainty model and its justified assumptions.

Where That Leaves Us

A digital instrument can display 12.345678 without establishing knowledge to six decimal places. This is a general lesson. Precision printed by software is formatting. Evidential precision comes from the uncertainty model and its justified assumptions. The next section examines the proposition that probability is a representation tool, not a solvent for ignorance.

3. Probability is a representation tool, not a solvent for ignorance

Probability enters science in more than one role. In repeated-sampling models it characterizes the behaviour of random variables under repetitions. In Bayesian analysis it is also used to represent uncertainty about unknown quantities conditional on a model, prior information and observed data. In metrology, probability distributions are used to represent values that could reasonably be attributed to uncertain quantities and to propagate those distributions through measurement models. [S02, S03]

[C06] The same numerical probability can therefore arise from different inferential constructions. A scientifically adequate report must state what random object or uncertain quantity the probability refers to, and under which assumptions. [S02, S06]

KNO-015 does not decide the philosophical meaning of probability. That belongs to IDE-008. Its concern is operational: do not move between interpretations without saying so.

Where That Leaves Us

Probability enters science in more than one role. [S02, S03] KNO-015 does not decide the philosophical meaning of probability. In Bayesian analysis it is also used to represent uncertainty about unknown quantities conditional on a model, prior information and observed data. The next section examines the proposition that confidence intervals and credible intervals answer different questions.

4. Confidence intervals and credible intervals answer different questions

The phrase “95% interval” is incomplete.

[C07] A frequentist confidence procedure is calibrated by its repeated-sampling properties under the assumed model. The usual 95% confidence interpretation concerns the long-run coverage of the procedure, not a posterior probability that the fixed parameter lies in the particular observed interval. [S10]

[C08] A Bayesian credible interval is a probability statement about a parameter or quantity conditional on the specified Bayesian model, including the prior and likelihood. Its interpretation is therefore not interchangeable with that of a frequentist confidence interval, even when the numerical endpoints are similar. [S10]

Neither label makes assumptions disappear. The relevant question is what data, model and inferential construction justify the interval.

Where That Leaves Us

The phrase “95% interval” is incomplete. Neither label makes assumptions disappear. The relevant question is what data, model and inferential construction justify the interval. 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 a p-value is not the probability that a hypothesis is true.

5. A p-value is not the probability that a hypothesis is true

This error is so common that it deserves a separate boundary.

[C09] The American Statistical Association states that p-values can indicate incompatibility between data and a specified statistical model, but do not measure the probability that the studied hypothesis is true, do not measure effect size or scientific importance, and should not by themselves determine scientific or policy conclusions. [S06]

A p-value is calculated under a specified model, usually including a null hypothesis and other assumptions. It is not a posterior probability of that hypothesis unless an additional Bayesian construction has been supplied, in which case the resulting posterior probability is a different object.

Where That Leaves Us

This error is so common that it deserves a separate boundary. It is not a posterior probability of that hypothesis unless an additional Bayesian construction has been supplied, in which case the resulting posterior probability is a different object. The next section examines the proposition that uncertainty propagates through models.

6. Uncertainty propagates through models

If an output depends on uncertain inputs, uncertainty in those inputs propagates to the output.

[C10] The GUM provides analytical propagation rules under its conditions of applicability, while JCGM 101 specifies propagation of probability distributions through a measurement model using Monte Carlo methods. [S01, S02]

Monte Carlo propagation is not magic. It inherits the input distributions, dependence structure and measurement model supplied to it. If those are poor, a large simulation can produce a very precise answer to the wrong uncertainty model.

Where That Leaves Us

If an output depends on uncertain inputs, uncertainty in those inputs propagates to the output. Monte Carlo propagation is not magic. It inherits the input distributions, dependence structure and measurement model supplied to it. If those are poor, a large simulation can produce a very precise answer to the wrong uncertainty model. The next section examines the proposition that structural uncertainty deserves its own line of sight.

7. Structural uncertainty deserves its own line of sight

Some uncertainties live outside the parameter covariance matrix of the selected model: alternative mechanisms, boundary conditions, unrepresented feedbacks, unknown dependence, or unresolved expert disagreement.

[C11] The IPCC uncertainty guidance explicitly separates assessments of evidence and agreement, confidence, likelihood or probability, and asks authors to state assumptions and assess structural uncertainties where relevant. [S07]

This separation prevents a category mistake. A probability distribution can summarize uncertainty conditional on a model while leaving uncertainty about the model itself only partially quantified.

Where That Leaves Us

Some uncertainties live outside the parameter covariance matrix of the selected model: alternative mechanisms, boundary conditions, unrepresented feedbacks, unknown dependence, or unresolved expert disagreement. A probability distribution can summarize uncertainty conditional on a model while leaving uncertainty about the model itself only partially quantified. The next section examines the proposition that expert judgment can be disciplined without pretending it is data.

8. Expert judgment can be disciplined without pretending it is data

In some domains direct frequencies are sparse, models disagree, or future conditions are not repetitions of a stable historical process. Expert judgment may then contribute to uncertainty assessment.

[C12] IPCC guidance uses calibrated language and traceable evaluation of evidence and agreement to discipline such judgments, rather than presenting every confidence statement as a frequency observed in data. [S07]

Expert judgment is not automatically weaker or stronger than statistical output. Its evidential role depends on transparency, calibration, independence, elicitation design and the evidence on which experts rely. The crucial rule is provenance: readers should know when a number came from observations, a model ensemble, formal elicitation or a mixture.

Where That Leaves Us

In some domains direct frequencies are sparse, models disagree, or future conditions are not repetitions of a stable historical process. Expert judgment may then contribute to uncertainty assessment. Expert judgment is not automatically weaker or stronger than statistical output. The next section examines the proposition that unknown unknowns are not inside every interval.

9. Unknown unknowns are not inside every interval

An uncertainty budget can only include sources represented in the analysis.

[C13] Scientific reports should therefore distinguish quantified uncertainty from recognized but unquantified limitations and from genuinely unrecognized error. The National Academies recommends reporting recognized uncertainties and potential sources across stochastic, measurement, computational, knowledge, modelling and analysis dimensions rather than treating a single statistical interval as exhaustive. [S08]

This does not make quantified intervals useless. It defines their jurisdiction.

Where That Leaves Us

An uncertainty budget can only include sources represented in the analysis. This does not make quantified intervals useless. 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 sensitivity analysis tests the dependence of conclusions on assumptions.

10. Sensitivity analysis tests the dependence of conclusions on assumptions

If a conclusion changes sharply under plausible alternative specifications, the uncertainty is not captured by quoting only the standard error under one specification.

Sensitivity analysis can vary priors, functional forms, nuisance parameters, missing-data assumptions, calibration inputs or plausible structural choices. It is not a universal cure: the analyst must still decide which alternatives are plausible. But it exposes where conclusions depend on choices that would otherwise remain hidden.

Where That Leaves Us

If a conclusion changes sharply under plausible alternative specifications, the uncertainty is not captured by quoting only the standard error under one specification. It is not a universal cure: the analyst must still decide which alternatives are plausible. The next section examines the proposition that prediction uncertainty is not parameter uncertainty.

11. Prediction uncertainty is not parameter uncertainty

Estimating a mean response and predicting a new observation are different tasks. A model parameter can be well estimated while individual future outcomes remain highly variable.

[C14] A complete uncertainty statement must identify the target quantity: a parameter, a latent state, a measurand, a future observation, a population average, or a decision-relevant outcome. Different targets generally require different uncertainty constructions. [S01, S03, S10]

This is why “the uncertainty is $\pm x$ ” is often underspecified.

Where That Leaves Us

Estimating a mean response and predicting a new observation are different tasks. A model parameter can be well estimated while individual future outcomes remain highly variable. This is why “the uncertainty is $\pm x$ ” is often underspecified. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines a practical uncertainty workflow.

12. A practical uncertainty workflow

For scientific work, the following sequence is more robust than attaching an interval at the end:

1. Define the target quantity and estimand or measurand.
2. Identify the data-generating or measurement process.
3. List material uncertainty sources and dependencies.
4. Separate stochastic variability from incomplete knowledge where the distinction matters.
5. Specify the model and inferential framework.
6. Quantify components that can be justified quantitatively.
7. Propagate them through the model, analytically or computationally as appropriate.
8. Test sensitivity to plausible alternative assumptions.
9. Record important limitations that remain unquantified.
10. Report the meaning of every interval, probability and coverage statement.

The workflow does not guarantee truth. It makes the path from evidence to uncertainty claim auditable.

Where That Leaves Us

Rather than attaching an interval at the end, the workflow defines the target quantity and its data or measurement process, lists uncertainty sources, separates variability from incomplete knowledge, specifies model and framework, quantifies and propagates justified components, tests sensitivity, records what stays unquantified and reports what each interval means. This makes the path auditable without guaranteeing truth. The next section asks what uncertainty does not license.

13. What uncertainty does not license

Uncertainty is not permission to say that every claim is equally plausible. Nor does a narrow interval prove a model correct. A result may be statistically precise and systematically wrong. A wide interval may faithfully report weak information rather than methodological failure.

The strongest uncertainty statement is therefore not always the smallest number. It is the one whose scope, assumptions and construction survive inspection.

Where That Leaves Us

Uncertainty is not permission to say that every claim is equally plausible. A result may be statistically precise and systematically wrong. A wide interval may faithfully report weak information rather than methodological failure. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

Boundary note

KNO-015 is about **the methods by which scientific work identifies, represents, propagates and communicates uncertainty**. It does not attempt to settle what probability fundamentally is; that conceptual question belongs to **IDE-008 · Probability**. It does not treat uncertainty as a general ontological or epistemological concept; that belongs to **IDE-017 · Uncertainty**. It builds directly on **KNO-011 · Measuring Before Theorising**, **KNO-012 · Experiments and Variable Control**, **KNO-013 · Models, Maps and Reality**, and **KNO-014 · Causality and Correlation**.

Declared framing statements not represented as numbered claims

This list is a control aid, **not a guarantee of semantic exhaustiveness**. It records substantive framing sentences intentionally left outside the numbered claim ledger and the reason for doing so.

- “Uncertainty is not permission to say that every claim is equally plausible.” Normative framing of calibrated uncertainty, not a distinct empirical claim.
- “A result may be statistically precise and systematically wrong.” Compressed methodological warning already bounded by the sections on systematic/model uncertainty.
- “A wide interval may faithfully report weak information rather than methodological failure.” Interpretive framing of interval width, not an independent quantitative rule.
- “The strongest uncertainty statement is therefore not always the smallest number.” Synthesis of the preceding uncertainty distinctions, not a new source-level proposition.

What We Know, and How

Uncertainty can arise from measurement, sampling, parameters, prediction, model structure and judgement, and these sources are not interchangeable. Statistical intervals and probabilities have meanings fixed by their inferential framework, while sensitivity analysis exposes dependence on assumptions.

What We Still Don't Know

No finite interval automatically contains every relevant unknown. Structural omissions, unmodelled mechanisms and genuinely unforeseen possibilities can remain outside a probability model even when the arithmetic inside that model is correct.

What Could Change Our Minds

A conclusion should be revised when calibration changes, new data widen or narrow sampling uncertainty, alternative model structures fit comparably well, sensitivity analyses reveal fragile assumptions, or previously ignored mechanisms become plausible enough to affect the target quantity.

If You Remember Only One Thing

Represent uncertainty in the form that matches the claim, and state explicitly which important unknowns the calculation does not contain.

Related Articles

- KNO-011 · Measuring Before Theorising
- KNO-014 · Causality and Correlation
- KNO-016 · How Evidence Is Built: Putting the Tools Together
- KNO-021 · Open Science and Reproducibility

References

[S01] 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.

[S02] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method*. JCGM 101:2008. DOI: 10.59161/JCGM101-2008.

[S03] Joint Committee for Guides in Metrology (JCGM). *Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models*. JCGM GUM-6:2020. DOI: 10.59161/JCGMGUM-6-2020.

[S04] Taylor, B. N.; Kuyatt, C. E. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST Technical Note 1297. NIST web manifestation consulted 5 September 2026; page updated 6 May 2026. <https://www.nist.gov/pml/nist-technical-note-1297>

- [S05] National Institute of Standards and Technology. *NIST TN 1297: Appendix B. Coverage Factors*. NIST web page, updated 11 September 2025; consulted 5 September 2026. <https://www.nist.gov/pml/nist-technical-note-1297/nist-tn-1297-appendix-b-coverage-factors>
- [S06] Wasserstein, Ronald L.; Lazar, Nicole A. “The ASA’s Statement on p-Values: Context, Process, and Purpose.” *The American Statistician* 70(2), 2016, 129–133. DOI: 10.1080/00031305.2016.1154108.
- [S07] 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.
- [S08] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S09] Draper, David. “Assessment and Propagation of Model Uncertainty.” *Journal of the Royal Statistical Society: Series B* 57(1), 1995, 45–70. DOI: 10.1111/j.2517-6161.1995.tb02015.x.
- [S10] Possolo, Antonio; Gelabert, Maria Victoria; Hibbert, David; Stohner, Jürgen; Bodnar, Olha; Meija, Juris. *Guía Breve sobre la Incertidumbre de Medición*. NIST Technical Note 2330, 2025. DOI: 10.6028/NIST.TN.2330.
- [S11] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Amendment 1: Nonlinearity in measurement models*. JCGM 100:2008/Amd.1:2026. DOI: 10.59161/PPDI3267.