THE ELEVATOR PITCH

Make the work behind the finding inspectable.

Research needs a clear connection between a reported result and the work that produced it.

Grid brings computational models, explanatory notebooks, and experiment comparisons into one workspace. Make assumptions explicit, examine how results change, and preserve selected findings with their exact model source, inputs, and computational provenance.

Grid Pro adds governed evidence capture and scholarly packages so collaborators can inspect the record supporting a conclusion.

Show how the result was produced. Show how it changes. Preserve the version under review.

REPEATABILITY IS CENTRAL

Preserve the conditions of a computation so supported results can be repeated and checked. Combine that repeatability with visible methods, explicit uncertainty, and a record another researcher can examine.

INSPECT THE METHOD | REPEAT THE COMPUTATION | PRESERVE THE EVIDENCE

01 / THE RESEARCH CASE

The methodological concerns are specific and actionable.

The evidence supports a practical agenda: expose analytical choices, retain the computational context, and give reviewers more of the record behind a conclusion. These studies explain the need; they did not evaluate Grid.

Analytical flexibility

Simmons and colleagues showed how ordinary choices about outcomes, sample size, covariates, and comparisons can inflate false positives. One combined simulation reached about 61% against a nominal 5% threshold. This was a conditional demonstration, not a prevalence estimate. [2]

Results depend on analytical decisions

In *Many Analysts, One Data Set*, 20 of 29 teams found a significant positive relationship and nine did not. A later reanalysis stresses the importance of a precise question and comparable target quantities. [3] [10]

Selective reporting changes the apparent evidence

A review covered 74 FDA-registered antidepressant studies. Among published trials, 94% appeared positive; FDA assessments classified 51% of all 74 trials as positive. The published evidence overstated the overall effect size by 32% in that trial set. [4]

Replication needs methods, data, and materials

A project repeating 100 psychology studies found significant results in 36% of replications versus 97% of originals. A cancer-biology project repeated 50 experiments from 23 papers and documented obstacles involving methods and materials. [5] [6]

Precision and interpretation matter

Low-powered studies can yield exaggerated effect estimates. A p-value alone does not measure an effect's magnitude or importance. Review requires uncertainty, design assumptions, and scientific context. [7] [8]

Grid's response: make the computational record available for scrutiny. The foundational argument about power, bias, and prior plausibility comes from Ioannidis; it does not establish that most science is fabricated. [1]

02 / REPEATABILITY

A result should carry the conditions for checking it.

Terminology varies by discipline. This brief uses the following distinctions, drawing on the National Academies' treatment of computational reproducibility and replication. [9]

Practice	Question it answers
Repeatability	Can the team repeat a computation under its recorded conditions?
Computational reproducibility	Can the original inputs, methods, and code support consistent results, including when checked by another researcher?
Independent replication	Does a new study, with new data, support the scientific finding?

What Grid can check today

For an eligible completed time-simulation run, Grid reopens the saved model, restores captured effective inputs and configuration in an isolated session, and compares the full canonical result with the saved record. A mismatch is rejected. [P2, P3]

PRESERVE Model + inputs → **REPLAY** Saved conditions → **COMPARE** Full result identity → **REVIEW** Verified trajectory

Preservation and replay serve different purposes

Research Workspace seals selected outputs with the exact current source, compiled model, effective inputs, and computational provenance. That captured record supports inspection; it is not a general rerun service for every artifact. The supported simulation replay path can use its own retained historical model. [P1-P3]

Scope: trajectory replay covers successful completed time simulations with full saved evidence. Active or failed runs, discrete-event and spatial-agent rows, and legacy records without that evidence are excluded. A semantic executor identifier is not a complete operating-system, hardware, and dependency image. Universal cross-platform byte identity is not claimed. [P3]

Repeatability makes a computation checkable. The same mistaken method can still repeat consistently. Scientific confidence also depends on valid measurements, justified methods, robustness, and independent evidence. [9]

03 / FEATURES FOR ANALYSIS

Keep reasoning, calculation, and uncertainty together.

These capabilities support the working researcher while an analysis is being constructed and examined. The shared model gives each view a common computational basis. [P2, P4, P5]

01 Computational notebooks

Combine narrative, formulas, live values, inputs, and provenance over one model. Execution receipts retain revision and status; captured output is marked stale when the model advances.

Use it to explain a result and reveal when an old output no longer describes the current model.

Captured output is a record; live values continue to follow model state.

02 Declared experiments and comparisons

Enumerate explicit parameter values, ranges, and scenarios in supported time-simulation sweeps. Inspect matrix size, returned summaries, failures, and eligible trajectories.

Use it to show whether a conclusion survives the alternatives the researcher declared.

The current sweep is full-factorial; it is not an automatic search over all defensible analyses.

03 Uncertainty and sensitivity

Propagate authored input distributions using supported first-order analysis or seeded Monte Carlo. Inspect local sensitivity drivers and their contributions.

Use it to show how measurement or input uncertainty reaches the reported quantity.

Distributions and correlations remain researcher assumptions; local derivatives are not global robustness proofs.

04 Checks during model authoring

Use declared units, strictness controls, and supported matrix-shape and value-bounds diagnostics to expose specific inconsistencies.

Use it to catch a unit mismatch or incompatible calculation before it enters the narrative.

These checks do not establish causal validity or appropriate study design; floating-point bounds are author guidance.

Related predictive capability: calibrated XGBoost regression provides supported upper prediction endpoints and calibration records. These are marginal predictive bounds, not confidence intervals for scientific effect estimates. [P6]

04 / GRID PRO RESEARCH WORKSPACE

Turn selected work into an inspectable research record.

One model-scoped workspace connects five workflows. Capture and publication are deliberate actions over exact artifacts, with corrections expressed through new records and status changes. [P1]

Evidence & Reproducibility

Preserve a selected current model state and outputs, including exact source, effective inputs, and computational provenance. Verify the retained artifact graph.

A reviewer can identify the computational record under discussion.

Research Use Policies

Bind authorized operations to an exact purpose, recipient class, artifact scope, and policy status. Preserve policy successors and withdrawals.

Teams can make permitted sharing and reuse explicit.

Bibliography & Standards

Import CSL-JSON, BibTeX, and RIS with source retention and reconciliation receipts. Supported exports report transformations and omissions.

Review can include the reference record and its conversion history.

Specialist Validation

Prepare exact specialist inputs and execution requests with declared identities and conditions. Execution depends on the selected supported contract.

External-tool work can be represented by a bounded, attributable record.

Scholarly Release

Compose immutable scholarly objects and RO-Crate packages. Use governed release, status history, and format-specific exports.

Evidence can accompany a citable object and retain correction history.

Sharing has a defined scope. A self-contained package embeds its bounded artifact closure; linked packages still depend on retained artifacts. Policies are not legal consent, and integrity verification is not scientific approval. DataCite and citation metadata do not mint a DOI or deposit a release. [P1]

Specialist execution is narrow: the separately selected native-Linux GROMACS worker currently admits a pinned single-argon reference contract. Ordinary Pro does not provide a general GROMACS or arbitrary-code executor. [P1, P7]

05 / WORKED EXAMPLE 1

From measurements to an inspectable uncertainty claim.

Question: how precisely have we estimated the density of a sample? A methods lab or research team can put measurements, assumptions, calculation, and interpretation in the same Notebook.

Illustrative inputs
Mass: 10.0 g with standard deviation 0.2 g
Volume: 5.0 cm^3 with standard deviation 0.1 cm^3
Assume independent, approximately normal measurement errors.

Density = mass / volume = **2.00 g/cm^3**

For a first-order approximation, the relative output variance is the sum of the relative input variances. Both inputs have 2% relative uncertainty, so each contributes half of the estimated variance.

Quantity	Illustrative calculation
Propagated standard deviation	2.00 x sqrt[$(0.2 / 10.0)^2$ + $(0.1 / 5.0)^2$] = **0.0566 g/cm^3**
Approximate central 95% interval	2.00 +/- 1.96 x 0.0566 = **1.889 to 2.111 g/cm^3**
If volume SD doubles to 0.2 cm^3	Output SD becomes **0.0894**; the same approximation gives **1.825 to 2.175 g/cm^3**.

How Grid extends the example

Explain: document the measurement source, distributions, and independence assumption.
Inspect: show the derived quantity, propagated interval, and local sensitivity drivers.
Compare: change the volume uncertainty and observe the revised interval.
Preserve: capture the model state and selected calculation outputs; retain uncertainty settings and the reported interval in the accompanying research record. [P1, P2, P4]

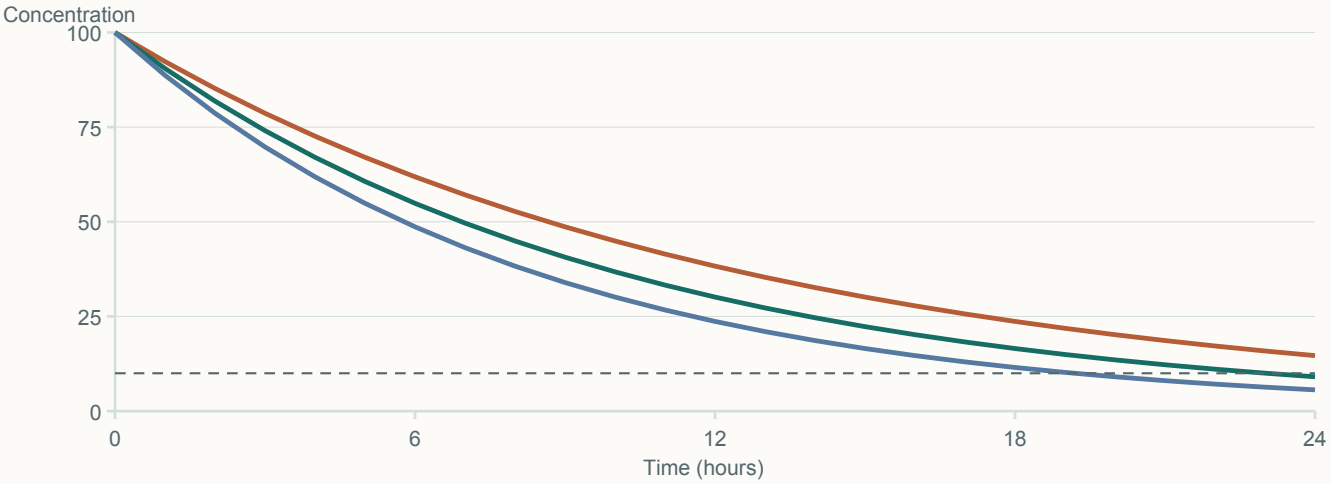
The methodological lesson: a repeated point estimate can conceal materially different uncertainty. These are propagated uncertainty intervals under stated assumptions, not evidence that the measurements are unbiased or a general confidence interval for a population parameter.

Illustrative first-order arithmetic with a normal approximation; not a recorded Grid execution.

06 / WORKED EXAMPLE 2

Repeat one run.
Challenge the conclusion across nine.

Consider an illustrative tracer-decay model: **$C(t) = C_0 \exp(-kt)$**. The question is whether concentration falls below 10 units after 24 hours. Declare three initial concentrations and three decay rates before comparing results.



C_0 = 100: k = 0.08 | k = 0.10 | k = 0.12 h^{-1}. Dashed line: threshold 10.

Decay rate (h^{-1})	C_0 = 90	C_0 = 100	C_0 = 110
0.08	**13.19** above	**14.66** above	**16.13** above
0.10	**8.16** below	**9.07** below	**9.98** below
0.12	**5.05** below	**5.61** below	**6.17** below

Interpretation: the baseline (100, 0.10) is below the threshold at 9.07. All three slow-decay cases remain above it. The threshold conclusion therefore depends on the declared decay-rate assumption; repeating the baseline alone would not reveal this.

In Grid: construct the nine-run time-simulation matrix, inspect every returned summary and failure, and replay an eligible completed trajectory. Publish the experiment evidence explicitly to retain the declared specification, counts, failures, results, and exact result Frame. [P2, P3]

Illustrative analytic reference values, rounded to two decimals; curves use the equation directly. No Grid run is claimed. A numerical simulation requires declared integration settings and comparison with a justified reference tolerance. The parameter matrix is not comprehensive statistical multiverse analysis.

07 / WORKED EXAMPLE 3 & NEXT STEPS

Make the research handoff a checkable record.

A collaborator asks, months later, which assumptions produced a submitted figure. The live model has changed. Grid's value is a concrete record of the submitted work and a way to examine supported computations against it.

1. Preserve at submission

Explicitly publish the completed experiment evidence and capture the selected current model outputs while that is the version being reported. Retain the explanation and bind the exact use policy.

2. Package the exact record

Compose the scholarly object from those artifacts. Use the supported governed release and export; state which inputs are embedded and which remain external dependencies.

3. Reopen and check later

Months later, inspect the preserved record. For an eligible time-simulation row, replay its saved model, conditions, and inputs against the recorded result, even if the live model has advanced.

4. Correct with a new version

If the analysis changes, create new evidence and a successor release or status transition. Preserve the original record so reviewers can distinguish the versions. [P1-P3]

Academic use: teach the method alongside the answer

The same pattern supports a methods assignment: students state assumptions, compare scenarios, explain uncertainty, and submit the relevant computational record. An instructor can assess both the result and the reasoning. This is a proposed teaching use of the demonstrated model workflow.

Extensions that would strengthen the argument

Prospective plans: hypotheses, outcomes, exclusions, stopping rules, and amendment history.
Complete study accounting: declared scope, attempted analyses, null outcomes, failures, and external-work gaps.
Methodological checks: power and precision planning, justified alternative analyses, and appropriate multiplicity handling.

These are proposed product directions, not current feature claims or delivery commitments. Current capture and experiment preservation do not establish preregistration or universal analysis logging.

Test the practical benefit. In a lab pilot, compare the same handoff task in Grid and the existing workflow. Measure correct source identification, detection of changed assumptions or stale outputs, completion of supported replay, and the distribution of review time. Report observed results before claiming time savings or improved research accuracy.

08 / REFERENCES & SCOPE

The evidence behind this brief.

[1] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLOS Medicine 2(8): e124.

[2] Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science 22(11): 1359-1366.

[3] Silberzahn, R., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. Advances in Methods and Practices in Psychological Science 1(3): 337-356.

[4] Turner, E. H., et al. (2008). Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. New England Journal of Medicine 358: 252-260.

[5] Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science 349(6251): aac4716.

[6] Errington, T. M., et al. (2021). Investigating the replicability of preclinical cancer biology. eLife 10: e71601.

[7] Button, K. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience 14: 365-376.

[8] Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician 70(2): 129-133.

[9] National Academies of Sciences, Engineering, and Medicine (2019). Reproducibility and Replicability in Science. National Academies Press. Chapters 3-4.

[10] Auspurg, K., & Brüderl, J. (2021). Has the Credibility of the Social Sciences Been Credibly Destroyed? Reanalyzing the "Many Analysts, One Data Set" Project. Socius 7.

Grid product basis

Product references refer to the reviewed Core documentation:
[P1] Research Workspace - docs/product/research-workspace.md
[P2] Notebook and Experiments - docs/product/surfaces.md
[P3] Experiment replay - docs/specs/simulation_phase_3_experiments.md
[P4] Authored uncertainty - docs/specs/gir/uncertainty_lane.md
[P5] Units, shapes, and bounds - docs/language/reference.md
[P6] Predictive calibration - docs/guides/calibrated-predictions.md
[P7] Implementation status - docs/roadmap/executable-research.md

Review scope. Feature descriptions were checked against Core source 074a9dd5c373df9cef67c05fb6664ec13a64c5c6 (package version 0.66.4). They describe source implementation, not verified availability in every installed or hosted edition. Research Workspace is Pro only; other features depend on their supported configuration. No live product session was executed for this brief.

Evidence boundary. Worked examples are transparent illustrations. The cited research does not endorse Grid or establish that it reduces false findings. Captured integrity, successful replay, and scientific validity require distinct evidence.