

The Hidden Correction Problem: When Data Treatment Creates Confidence Instead of Information

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. All Rights Reserved.

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

HEB-009 · Independent Technical Article · Version v0.3 · 23 September 2026

Abstract

Data treatment is necessary in many experiments, but every correction contains a model of what the unwanted component is and how it enters the measurement. Baseline subtraction, normalization, smoothing, blank correction, background removal, deconvolution and outlier deletion can reduce known artefacts when their assumptions are valid. They can also suppress real structure, amplify noise, manufacture apparent precision or hide the fact that the experiment never separated signal from interference. This methodological article develops the hidden correction problem: the risk that a cleaner result becomes a more persuasive representation without becoming a more informative measurement. The central proposition is that preprocessing can recover information only when the information was present in the raw measurement in a separable form; it cannot create distinctions the experiment never resolved.

Keywords: data correction; baseline; smoothing; normalization; subtraction; deconvolution; outlier; preprocessing; uncertainty; raw data

1. Purpose and evidence boundary

HEB-001 states: statistics cannot manufacture information that the experiment never generated [1]. HEB-009 applies that principle to preprocessing and correction. Application-specific analytical workflows remain with their ER owners; for example ER-253 retains blank subtraction and threshold logic for extractables [4].

2. Every correction is a measurement model

Subtracting a blank assumes that blank and sample share the same background contribution. Normalising by a reference assumes that the reference captures the multiplicative change affecting the sample. Smoothing assumes that the removed high-frequency component is noise rather than signal. Baseline fitting assumes a shape for the background. A correction is therefore not neutral arithmetic; it is a model embedded in the analysis [2].

3. Physical removal is stronger than mathematical removal when feasible

HEB-001 describes thin-film interference introduced by polyethylene films in a double-beam RTIR configuration. The broader lesson is that an artefact should preferably be prevented or physically eliminated before a numerical correction is designed around it [1]. A mathematical correction can be useful, but it inherits uncertainty from both the original measurement and the correction model.

4. Subtraction can increase uncertainty

If a corrected result is the difference between two uncertain measurements, both uncertainty contributions matter. When sample and blank are similar in magnitude, the corrected value can have poor relative precision even if each input appears stable. Reporting only the corrected number conceals the information loss created by subtraction.

5. Normalization can move the error into the denominator

A ratio is only as stable as its denominator. If several samples are divided by one drifting reference, the normalization can move a common error into every reported result and create correlation across the corrected series. The apparent reduction of scatter therefore does not demonstrate improved validity; the denominator belongs inside the measurement model [2].

6. Smoothing trades resolution for variance

Smoothing can make a trace visually coherent and improve estimation of broad trends, but it also attenuates narrow features and changes correlation between neighbouring points. The method, window and parameters should therefore be declared, and scientific conclusions should remain traceable to unsmoothed data. A smooth curve is not automatically a more precise experiment [3].

7. Deconvolution is constrained inference

Resolving overlapping signals by model fitting can be powerful when component shapes and response physics are justified. It is weak when multiple decompositions fit the same data similarly well. The output may contain more numerical parameters than the raw experiment independently constrained. Parameter identifiability must therefore be separated from visual goodness of fit.

8. Outlier removal can erase mechanism

An extreme value can be transcription error, instrument fault or a genuine observation from a different state. Deleting it solely because it is inconvenient makes the model determine the data. Investigation should begin with identity, raw data, calibration, protocol deviation and physical plausibility. Analyses with and without a consequential exclusion can expose whether the conclusion depends on the decision.

9. Correction chains accumulate hidden assumptions

A pipeline that performs blank subtraction, normalization, baseline fitting, smoothing and selective exclusion can contain several individually plausible steps whose combined effect is difficult to audit. Each transformation should retain a link to the raw data, parameters and rationale. The longer the chain, the more important it is to test the complete pipeline with known or simulated inputs.

10. The complementary-failure test for data treatment

For every correction ask two questions: What failure is this intended to suppress? What new failure can the correction introduce? This is the complementary-failure principle from HEB-001 applied to analysis [1]. A correction should be challenged using controls for overcorrection, undercorrection, parameter sensitivity and preservation of known features.

11. Correction uncertainty should be propagated, not hidden

A correction factor, background model or reference value has uncertainty of its own. If that term is estimated from limited data, its uncertainty can be correlated across many corrected observations. Treating the corrected values as independent can therefore underestimate shared uncertainty. For example, normalising an entire series by one uncertain reference introduces a common denominator. Every corrected point can move together if the reference is wrong. The apparent precision of point-to-point differences and the uncertainty of the absolute scale are then different questions. A defensible analysis states which uncertainties are shared, which are independent and which transformation created the dependence.

12. Sensitivity to preprocessing choices is evidence about robustness

When several reasonable windows, baselines or fitting constraints are available, the stability of the scientific conclusion across those choices is informative. If a peak, trend or group difference exists only for one narrow preprocessing setting, the conclusion may be an analysis artefact. If the result survives a declared range of reasonable settings and remains visible in the raw data, confidence increases. This does not justify trying many pipelines and selecting the one that gives the preferred result. The choices and robustness checks should be defined by the measurement problem, ideally before the final comparison is inspected.

13. Some information losses are irreversible

No post-processing can recover a saturated detector value, an unresolved spatial average, signal content outside the instrument's effective bandwidth or a specimen state that was never recorded. Interpolation can estimate a missing point only under a model; deconvolution can separate components only to the degree that the measurement contains distinguishable structure. The hidden correction problem becomes severe when the analysis treats an irreversible information loss as though it were a removable nuisance. The correct scientific output may be a limitation or a repeat experiment rather than a more elaborate algorithm.

14. Conclusions

Data treatment is justified when its measurement model is explicit, testable and proportionate to the information in the raw data. Corrections should preserve raw observations, propagate uncertainty and remain sensitive to alternative plausible parameter choices. A visually improved dataset is not necessarily an epistemically improved dataset. The correct endpoint is not the cleanest curve but the most defensible inference.

References

- [1] Segurola, J. The Hidden Error Budget: Measurement-System Failure, False Precision and Experimental Design in Laboratory Science. HEB-001. 3Dresyns Independent Technical Articles (2026). DOI: 10.5281/zenodo.22862642.
- [2] Joint Committee for Guides in Metrology. Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008 with Amendment 1:2026, Nonlinearity in measurement models. Base DOI: 10.59161/JCGM100-2008E; amendment DOI: 10.59161/PPDI3267.
- [3] Heckert, N. A.; Filliben, J. J.; Croarkin, C. M.; Hembree, B.; Guthrie, W. F.; Tobias, P.; Prinz, J. Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods. National Institute of Standards and Technology, 2002.
- [4] Segurola, J. Analytical Evaluation Thresholds for Photopolymer Extractables: Coverage, Identification, Quantitation and Uncertainty. ER-253. 3Dresyns Engineering Publications (2026).

Knowledge boundary

HEB-009 owns the general epistemic consequences of preprocessing and correction. It does not prescribe an application-specific analytical pipeline or replace ER-253 and other method-specific qualification reviews.

Limitations

This methodological article is a bounded synthesis of the controlled HEB and Engineering Review corpus. It does not report a new experimental dataset, does not establish universal numerical thresholds, and does not convert examples from one domain into application-wide specifications. Claims that require application-specific qualification remain with the corresponding technical review or primary study.