

The Hidden Replication Problem: Repeated Measurements Are Not Independent Experiments

Juan Seguro

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Seguro. All Rights Reserved.

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

HEB-008 · Independent Technical Article · Version v0.3 · 23 September 2026

Abstract

Replication is often counted rather than defined. Five instrument readings, five positions on one specimen, five specimens from one build and five independently prepared batches can all produce the number $n = 5$ while estimating different variance components. Treating these levels as interchangeable creates pseudoreplication, overstates the amount of independent information and can make uncertainty appear smaller than the experiment warrants. This methodological article develops the hidden replication problem as a question of experimental-unit identity and variance structure. It separates repeated readings, repositioning, specimen replication, preparation replication, build replication, day effects, operator effects and laboratory effects, and shows why the level of replication must match the level of the scientific claim. The central proposition is that independence is a property of the data-generating process, not of the number of rows in a spreadsheet.

Keywords: replication; pseudoreplication; repeatability; reproducibility; hierarchical variance; independence; experimental unit; nested design

1. Purpose and evidence boundary

HEB-001 introduced the distinction between repeated readings and independent replicates [1]. HEB-008 develops that distinction into a general framework. ER-209 retains interlaboratory round-robin design and its specific statistical architecture [4]. The central proposition is: Count experimental units, not measurements.

2. The experimental unit is the unit independently assigned to treatment

If a treatment is applied once to one batch and ten aliquots from that batch are measured, the treatment was not independently replicated ten times. The ten measurements can estimate within-batch heterogeneity or measurement repeatability, but they do not estimate batch-to-batch treatment variability. The experimental unit is the unit independently assigned to treatment, whereas the observational unit is the unit on which a measurement is made. These can differ: ten aliquots measured from one independently treated batch are ten observational units but one experimental unit for the treatment contrast [5].

3. Repeated readings answer a narrow question

Repeated readings without disturbing the specimen primarily probe short-term instrument and analysis variation. Removing and reinserting the specimen adds positioning. Preparing a new specimen adds preparation variation. Repeating the build adds manufacturing variation. Repeating on another day adds temporal and environmental components. Replication becomes broader as more of the causal chain is independently regenerated. Metrological repeatability and reproducibility likewise distinguish the conditions under which measurement precision is evaluated [2].

4. Pseudoreplication creates false precision

If correlated observations are analysed as though they were independent, the nominal sample size exceeds the effective sample size. Standard errors can shrink artificially and statistical significance can appear stronger than the design supports. Averaging technical replicates before higher-level analysis can sometimes be appropriate, but the correct model depends on the hierarchy and the claim [5].

5. Nested variation is information, not inconvenience

A useful hierarchy is laboratory, with day nested within laboratory, build nested within day, specimen nested within build, position nested within specimen, and reading nested within position. Different questions live at different levels. A material-screening study may need independent batches; a manufacturing study may need independent builds; a metrology study may need repeated placements and readings [3].

6. More measurements at the wrong level do not fix missing replication

One hundred repeated readings on one specimen can estimate the mean readout of that specimen very precisely while leaving preparation-to-preparation uncertainty completely unknown. Increasing technical replication cannot substitute for independent biological, material, batch, build or laboratory replication when those levels support the claim.

7. Randomisation and replication solve different problems

Randomisation reduces systematic alignment between nuisance variation and factor settings; replication estimates variation. Neither substitutes for the other. A perfectly randomised study with no independent replication cannot estimate the relevant pure error, while a replicated study with strongly confounded run order can attribute drift to treatment.

8. Replication must match the transfer claim

ER-209 treats interlaboratory round-robin design and distinguishes within-specimen repeatability from build-to-build and laboratory-to-laboratory variation; that application-specific qualification remains outside HEB-008 [4]. HEB-015 separately owns the broader question of transfer level, while HEB-008 owns the independence structure inside the experiment.

9. Means can hide the hierarchy

Collapsing all observations into one mean and standard deviation can mix within- and between-unit variability. Two studies can report the same n and the same standard deviation while representing very different evidence structures. Reporting should therefore state what was replicated and how the observations were nested or crossed.

10. A practical replication map

Before data collection, draw the causal chain and mark every level that can vary: raw-material lot, preparation, batch, build, specimen, position, readout, day, operator, instrument and laboratory. For each scientific claim, identify the highest level across which the claim is intended to generalise. Replicate at that level or narrow the claim.

11. An illustrative replication hierarchy

Imagine a coating study with three independently prepared batches. From each batch, four specimens are cast; on each specimen, five positions are measured; at each position, the instrument records three repeated readings. The dataset contains 180 numbers, but it does not contain 180 independent batches, specimens or positions. The three readings mainly inform short-term readout repeatability; the five positions inform within-specimen spatial variation; the four specimens inform specimen-to-specimen variation conditional on batch; and the three batches inform preparation-to-preparation variation. If the scientific claim concerns formulation preparation, the strongest independent count is closer to the number of batches than to the number of raw readings. The lower-level measurements are still valuable because they help decompose the error structure. Their value disappears only when they are mislabelled as independent evidence for a higher-level claim.

12. Nested and crossed structures answer different transfer questions

Hierarchies are often nested: positions belong to specimens, specimens belong to builds and builds belong to batches. Other factors can be crossed: the same operators may measure specimens from every batch, or the same instruments may test material from several preparations. Crossing factors can separate operator or instrument effects from material effects more efficiently than assigning one operator to one batch, which confounds them. The design should therefore reflect the variance components that matter. A study intended to compare operators should give operators overlapping material states; a study intended only to estimate within-operator repeatability may intentionally keep operator fixed. Replication is not a generic request for 'more n '. It is an architecture for making particular sources of variation estimable.

13. Statistical power and generalisability are not the same objective

Adding many technical replicates can increase precision for a narrow mean while doing little for generalisation across batches, days, machines or laboratories. Conversely, spreading limited resources across more independent high-level units may increase transfer information at the cost of fewer readings per unit. Experimental design therefore involves a trade-off between characterising each unit accurately and sampling the population to which the conclusion will be applied. The optimal allocation depends on the variance structure and claim. What must not happen is to use abundant low-level replication to imply evidence about high-level variability that was never sampled.

14. Conclusions

Replication is meaningful only relative to a defined experimental unit. Repeated measurements can improve precision at one level while leaving higher-level uncertainty untouched. Pseudoreplication arises when correlated observations are counted as independent evidence. A robust experiment therefore declares the hierarchy, preserves unit identity and estimates the variance components that correspond to the intended claim.

References

- [1] Segurolo, 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. International Vocabulary of Metrology - Basic and General Concepts and Associated Terms (VIM), 3rd ed., JCGM 200:2012. DOI: 10.59161/JCGM200-2012.
- [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] Segurolo, J. Interlaboratory Round-Robin Design for Vat-Photopolymer Materials. ER-209. 3Dresyns Engineering Publications (2026). Version DOI: 10.5281/zenodo.22862020.
- [5] Hurlbert, S. H. Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs 54(2) (1984) 187-211. DOI: 10.2307/1942661.

Knowledge boundary

HEB-008 owns the general distinction between repeated observations and independent experimental replication. ER-209 retains interlaboratory design, round-robin execution and standards-specific reproducibility analysis.

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.