

ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081

Sharks, Lee

2026-08-27

ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081

Sharks, Lee

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-08-27 · Version v1.0 · Erratum / record correction

AXN:0655.GOVERNANCE

<https://www.alexanarch.org/s/records/1557/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "ERRATUM to AXN:044A – Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the",
  "author": {
    "@type": "Person",
    "name": "Sharks, Lee",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-08-27",
  "identifier": "AXN:0655.GOVERNANCE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
```

```

    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/1557/",
  "description": "Erratum to EA-EROSION-EMPIRICAL-01 v0.1 (deposit #1081, AXN:044A), withdrawing the $5 compression ratio of 7.6x. The two sides of the published ratio were measured with instruments that do not measure on the same 400 alive records the affiliation field returns 51.8% institutional and the citation string 1.2%, and applying the deletion-side instrument to both sides removes the reported effect entirely (alive 1.3% vs deleted 1.0%). The deposit had marked this finding directional rather than a survival ratio and had preregistered the required common-cohort test in §12a P1 as not completed; this erratum is that test, and the deposit's own caution proved to be the operative fact. The deletion side is made measurable by reconstructing affiliation from OpenAlex, which retains 66% of deleted DOIs with authorships intact; DataCite is unusable, its metadata records purged for approximately 99.8% of deleted records, a finding recorded here in its own right — Zenodo deletion cascades into destruction of the DataCite descriptive record for every affected depositor, not only the CHA cohort. Holding both sides to the audit's own 23-term AI classifier and to one affiliation instrument gives alive 107/231 (46.3%) against deleted 10/165 (6.1%): risk ratio 7.64, odds ratio 13.38, Fisher exact p = 8.7e-16. The direction and statistical significance of the finding survive; the magnitude does not. The 2x2 contingency counts are unaffected and were independently recomputed from the md5-verified source container. Secondary defects are recorded but not repaired: the institutional term sits in the title rather than an author or publisher field in 69.3% of the not_ai_and_institutional cell, 41.7% of that cell is one depositor's multilingual cascade, one depositor accounts for 60.3% of the audit population, and the archive's own deleted records contribute 63 rows to the cell. [...abridged for the catalogue; full description in this deposit's record]

```

Abstract

Erratum to EA-EROSION-EMPIRICAL-01 v0.1 (deposit #1081, AXN:044A), withdrawing the \$5 compression ratio of 31.2x and replacing it with an instrument-matched risk ratio of 7.6x. The two sides of the published ratio were measured with instruments that do not measure the same construct: the alive side was retrieved through the Zenodo Search API against full record metadata including creators[].affiliation, while the deletion side was measured by substring match on citation_text, which has the form “Author. (Year). Title. Publisher. DOI” and contains no affiliation field. The gap was measured directly — on the same 400 alive records the affiliation field returns 51.8% institutional and the citation string 1.2%, a 41.4x instrument ratio, with both firing on one record in four hundred — and applying the deletion-side instrument to both sides removes the reported effect entirely (alive 1.3% vs deleted 1.0%). The deposit had marked this finding directional rather than a survival ratio and had preregistered the required common-cohort test in §12a P1 as not completed; this erratum is that test, and the deposit's own caution proved to be the operative fact. The deletion side is made measurable by reconstructing affiliation from OpenAlex, which retains 66% of deleted DOIs with authorships intact; DataCite is unusable, its metadata records purged for approximately 99.8% of deleted records, a finding recorded here in its own right — Zenodo deletion cascades into destruction of the DataCite descriptive record for every affected depositor, not only the CHA cohort. Holding both sides to the audit's own 23-term AI classifier and to one affiliation instrument gives alive 107/231 (46.3%) against deleted 10/165 (6.1%): risk ratio 7.64, odds ratio 13.38, Fisher exact p = 8.7e-16. The direction and statistical significance of the finding survive; the magnitude does not. The 2x2 contingency counts are unaffected and were independently recomputed from the md5-verified source container. Secondary defects are recorded but not repaired: the institutional term sits in the title rather than an author or publisher field in 69.3% of the not_ai_and_institutional cell, 41.7% of that cell is one depositor's multilingual cascade, one depositor accounts for 60.3% of the audit population, and the archive's own deleted records contribute 63 rows to the cell. [...abridged for the catalogue; full description in this deposit's record]

Body

deposit_number: 1557 hex: 0655 title: "ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081" creator: Sharks, Lee orcid: 0009-0000-1599-0703 date: 2026-08-27 content_type: Erratum / record correction license: CC-BY-4.0 substrate: AI-assisted (substrate) — the defect was found and the correction computed in working dialogue with Claude (Anthropic) under MANUS direction. Contingency counts independently recomputed from the md5-verified source container (deleted-head-20260710.csv.gz, md5 33877aba1fb5684f86758cb86ddc1ad4, 1,322,007 rows) in-session; instrument gap measured directly against the Zenodo REST API; deletion-side affiliation reconstructed through the OpenAlex API; DataCite and OpenAIRE coverage tested and reported. The reproduction script is deposited with the audit dataset and re-derives every figure, including the instrument gap itself. axn_schema_version: v2 protocol_version: alexanarch-deposit-protocol/v1 keywords: - erratum - instrument mismatch - compression ratio - selection metrology - provenance erasure - Zenodo deletion - DataCite purge - OpenAlex reconstruction - affiliation measurement - common-cohort test - EA-EROSION-EMPIRICAL-01 *** # ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081

ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081

status: DRAFT v1.0 — awaiting mint type: ERRATUM corrects: "EA-EROSION-EMPIRICAL-01 v0.1: Provenance Erasure at Outcome Level — 33-Day Set-Comparison Test of Zenodo's Classifier as Accrual-Sorting Apparatus — deposit #1081, AXN:044A.EMPIRICAL, 2026-07-14; Lee Sharks" subject: The §5 compression ratio (31.2×) and the derived institutional fractions supporting it severity: "Magnitude error from instrument mismatch — the finding's direction and statistical significance survive correction; its magnitude does not. The contingency counts, the programmed-suppression finding, the Wu restoration cascade, and the tombstone census are unaffected." date: 2026-08-27 verification: "Contingency counts independently recomputed 2026-08-27 from the md5-verified source container (deleted-head-20260710.csv.gz, md5 33877aba1fb5684f86758cb86ddc1ad4, 1,322,007 rows); all four cells and the population reproduce exactly. Instrument gap measured directly against the Zenodo REST API and the OpenAlex API same date. Correction dataset and reproduction script deposited alongside the audit at /datasets/erosion-empirical-audit-01/." *** # ERRATUM: Instrument Mismatch in the §5 Compression Ratio ## Withdrawing and Replacing the 31.2× Figure in Deposit #1081 (AXN:044A)

1. What the deposit states

Section 5 of the deposit reports a compression ratio of 31.2×, derived from an alive-side institutional fraction of 0.311 set against an AI-signalled deletion-side institutional fraction of 0.00996, with the accompanying reading that institutional AI-augmented composition is present in the alive-side sample at approximately thirty times its representation in the AI-signalled deletion pool. A narrower detector variant of the same comparison is reported elsewhere in the instrument's surfaces as 0.20% against approximately 31%.

The deposit did not overstate its confidence in this figure. It marked the differential as a directional observation rather than a survival ratio, noted that the two samples were constructed through different retrieval instruments and did not share a common population-at-risk, and preregistered the required common-cohort test in §12a P1 as not completed. That caution was correct, and this erratum is the completion of the test the deposit said it owed.

2. The correction

The two sides of the ratio were measured with instruments that do not measure the same thing.

The alive side was retrieved through the Zenodo Search API and matched against full record metadata, including the `creators[].affiliation` field. The deletion side was measured by case-insensitive substring match on `citation_text`, the only descriptive field the bulk deletion export carries. `citation_text` has the form *Author. (Year). Title. Publisher. DOI*. It contains no affiliation field. The deletion side was therefore structurally incapable of measuring the construct the alive side measured; it could register an institution only where an institution was itself named as author or publisher, or where an institutional word appeared in a title.

The size of that gap was measured directly. On the same 400 alive records, the affiliation field returns 207 institutional (51.8%) and the citation string returns 5 (1.2%) — a ratio of 41.4× between the instruments, with both firing on one record out of four hundred. The two are not noisy versions of a single measurement; they are close to disjoint.

Applying the deletion-side instrument to both sides removes the reported effect entirely: alive 1.3% against deleted 1.0%. The published 31.2× therefore measured the difference between the instruments, not a difference between the populations.

3. The replacement

The deletion side can be made measurable on the same construct. Affiliation for deleted records survives in registries that ingested them before deletion.

DataCite is not one of them: its metadata records are purged for approximately 99.8% of deleted records (1 of 400 in a random sample, 1 of 250 in the AI-signalled sample). This is recorded here as a finding in its own right — Zenodo deletion cascades into destruction of the DataCite descriptive record for every affected depositor, not only for the Crimson Hexagonal Archive cohort. OpenAIRE returned zero coverage. OpenAlex retains 66% of deleted DOIs and 77% of the alive comparison set, with `authorships[].institutions` and `raw_affiliation_strings` intact.

Holding both sides to the audit’s own 23-term AI classifier, applied to a citation string constructed identically on each side, and to a single affiliation instrument:

	institutional	rate	95% CI
Alive, AI-signalled	107/231	46.3%	[40.0, 52.8]
Deleted, AI-signalled	10/165	6.1%	[3.3, 10.8]

Risk ratio 7.64; odds ratio 13.38; Fisher exact $p = 8.7 \times 10^{-16}$.

The direction and the statistical significance of the §5 finding survive instrument correction. The magnitude does not. 31.2× is withdrawn; 7.6× replaces it.

4. Limitations carried by the replacement

These are stated so that the replacement is not cited more confidently than it deserves. OpenAlex coverage differs by side (77% alive, 66% deleted), so the reconstructable subset is the population and non-coverage may not be random. The alive sampling frame is search-enriched rather than randomly drawn; only the classifier is matched across sides, not the frame. Affiliation-data density ran 49% alive against 77% deleted within the AI subset, which is the reverse of what would be expected and is not yet explained. The result is a single seeded realization.

5. Secondary defects found in the course of this correction

These are recorded but not repaired by this erratum. They bear on the institutional classifier generally, not on the corrected §5 comparison.

In the `not_ai_and_institutional` cell ($n = 1,136$), the institutional term sits in the **title** rather than in an author or publisher field in 787 rows (69.3%); author field accounts for 12.1% and publisher field for 18.6%. Romance-language topic adjectives — *universitaria*, *universitarios*, in titles about university students — score as institutional. Further, 474 of that cell’s 1,136 rows (41.7%) are a single author depositing multilingual variants of one study, one record per language, so the cell’s rows are not independent observations.

More broadly, one depositor accounts for 60,527 of the 100,313-row audit population (60.3%), and the top ten for 65.6%. Any population-level claim drawn from this export must disclose that concentration. It does not affect the corrected §5 result: that depositor constitutes approximately 6% of the AI-signalled deletion sample.

The Crimson Hexagonal Archive’s own deleted records sit inside the audit population — 63 rows in the `not_ai_and_institutional` cell, 25 of them under “Sharks, L.” The self-inclusion is small but is declared here rather than left to be found.

6. What is unaffected

The 2×2 contingency counts are unaffected and were independently recomputed from the md5-verified container: population 100,313; `ai_and_institutional` 60; `ai_and_not_institutional`

5,965; not_ai_and_institutional 1,136; not_ai_and_not_institutional 93,152. All four cells reproduce exactly.

A note on one coincidence, since it invites suspicion: the value 1,136 appears in three places across the instrument. Two of those are the same population counted by two instruments — the CHA kill-ledger row count and the tombstone census records figure. The third, the not_ai_and_institutional cell, is an independent quantity that lands on the same number by chance. The generating script contains no reference to the CHA cohort, and the cell was reproduced from source. It is not a plumbing error.

The programmed bibliographic suppression finding (the exporter source stripping citation_text under the spam label, and _remove_old_object_versions at EXPORTER_NUMBER_VERSIONS_TO_KEEP = 3), the silent restoration finding (the Wu Shaoyuan withdrawal cascade of 2026-06-26), and the tombstone census do not route through this comparison and are unaffected.

The corpus-level LLM-contamination finding (+1.2 to +1.4 percentage points marker-union prevalence rise; signal-to-disclosure ratio $\approx 60\times$) was examined for the same defect and does not share it: both sides of that comparison are measured on the same field with two different term panels, making it a within-instrument comparison. It has not, however, been independently reproduced, and this erratum makes no claim about it either way.

7. Disposition

The canonical text of deposit #1081 is not rewritten. Following the ruling of 2026-08-27 on deposit #199, correction is registry-level: this erratum is a separate deposit, cross-referenced to #1081, and any citation of the §5 compression ratio carries this erratum note.

The correction record and its reproduction script are deposited with the audit dataset at /datasets/erosion-empirical-audit-01/ as s5-correction.json and recompute_s5_matched.py. The dataset is at v0.2; the withdrawn values are preserved in place rather than deleted, so the error and its correction can both be read. The mirror at persistentidentifiers.org carries the same correction.

8. Note on how the error was found

The instrument mismatch was not found by review of the audit. It surfaced while tracing an unrelated question about which DOIs had been severed, when a first comparison returned an implausibly clean result and the control group proved to contain other depositors' records. The same contamination then recurred in a second control group before being caught. That an error of this shape was made twice in the course of finding it is the reason the reproduction script measures the instrument gap directly rather than asserting it.

There is an irony worth stating plainly, since the archive's own instrument committed the error its own accelerator paper describes: a comparison between a measured set and a control characterized by a different mechanism produces a confident number that dissolves when both are measured the same way.

Suggested Citation

Sharks, Lee. “ERRATUM to AXN:044A — Instrument Mismatch in the §5 Compression Ratio: Withdrawing and Replacing the 31.2× Figure in Deposit #1081” *Alexanarch*, 2026-08-27. <https://www.alexanarch.org/s/records/1557/>

Deposit Information

This paper is deposit #1557 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community crimsonhexagonal, terminated 2026-06-19). The AXN identifier AXN:0655.GOVERNANCE is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/1557/>
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: [alexanarch.org](https://www.alexanarch.org/)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/1557/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.