

CT.gov Structural Missingness

Field-level information loss across the full ClinicalTrials.gov registry

Zaki Khan

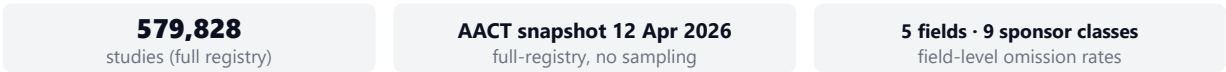
Ahmadiyya Hospital, Kenema, Sierra Leone
Corresponding author: zakikhan101@gmail.com

SYNTHESIS · METHODS NOTE

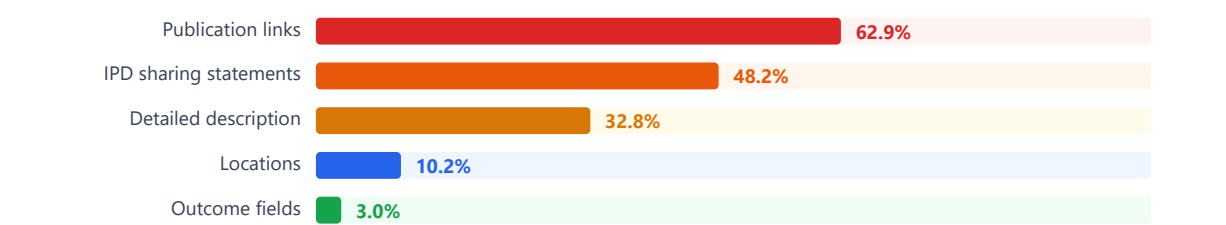
CT.gov Structural Missingness

THE QUESTION

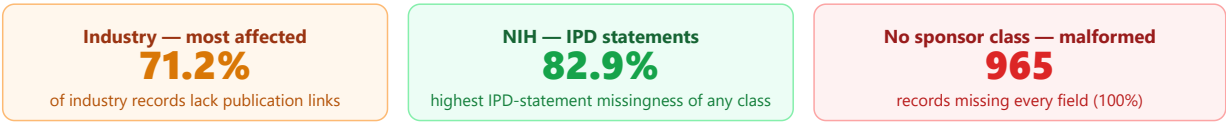
What information disappears from ClinicalTrials.gov *before* one even asks whether the results were posted?



WHAT IS MISSING — SHARE OF RECORDS LACKING EACH FIELD



SPARSITY IS UNEVEN ACROSS SPONSORS



Boundary: these capture registry-visible information loss, not proven intent to conceal.
Verified on the AACT 12 Apr 2026 full-registry snapshot (n = 579,828); four headline rates recomputed and confirmed.

Visual abstract. Field-level structural missingness across the full ClinicalTrials.gov registry (AACT snapshot 12 April 2026, n = 579,828), with the uneven distribution across sponsor classes.

PRIMARY ESTIMAND

Field-level structural missingness across the full ClinicalTrials.gov registry.

ABSTRACT (E156 · 7 SENTENCES, 151 WORDS)

What information disappears from ClinicalTrials.gov even before one asks whether the results were posted? We analysed the 12 April 2026 full-registry snapshot and quantified structural missingness in publication links, IPD statements, detailed descriptions, locations, and outcome fields. The source universe comprised 579,828 studies, allowing field-level omission rates and sponsor-specific sparsity patterns to be estimated without sampling. Across the full registry, 62.9 percent of records lacked publication links, 48.2 percent lacked IPD sharing statements, 32.8 percent lacked a detailed description, and 10.2 percent lacked locations. Structural sparsity was unevenly distributed: industry records were the most affected overall, NIH had the highest missingness of IPD sharing statements, and records carrying no sponsor class reflected malformed metadata. This missingness extends beyond results reporting into the descriptive fields needed for

interpretation, replication, and scrutiny, leaving less context for appraisal, accountability, and public oversight. These metrics capture registry-visible information loss rather than proven intent to conceal.

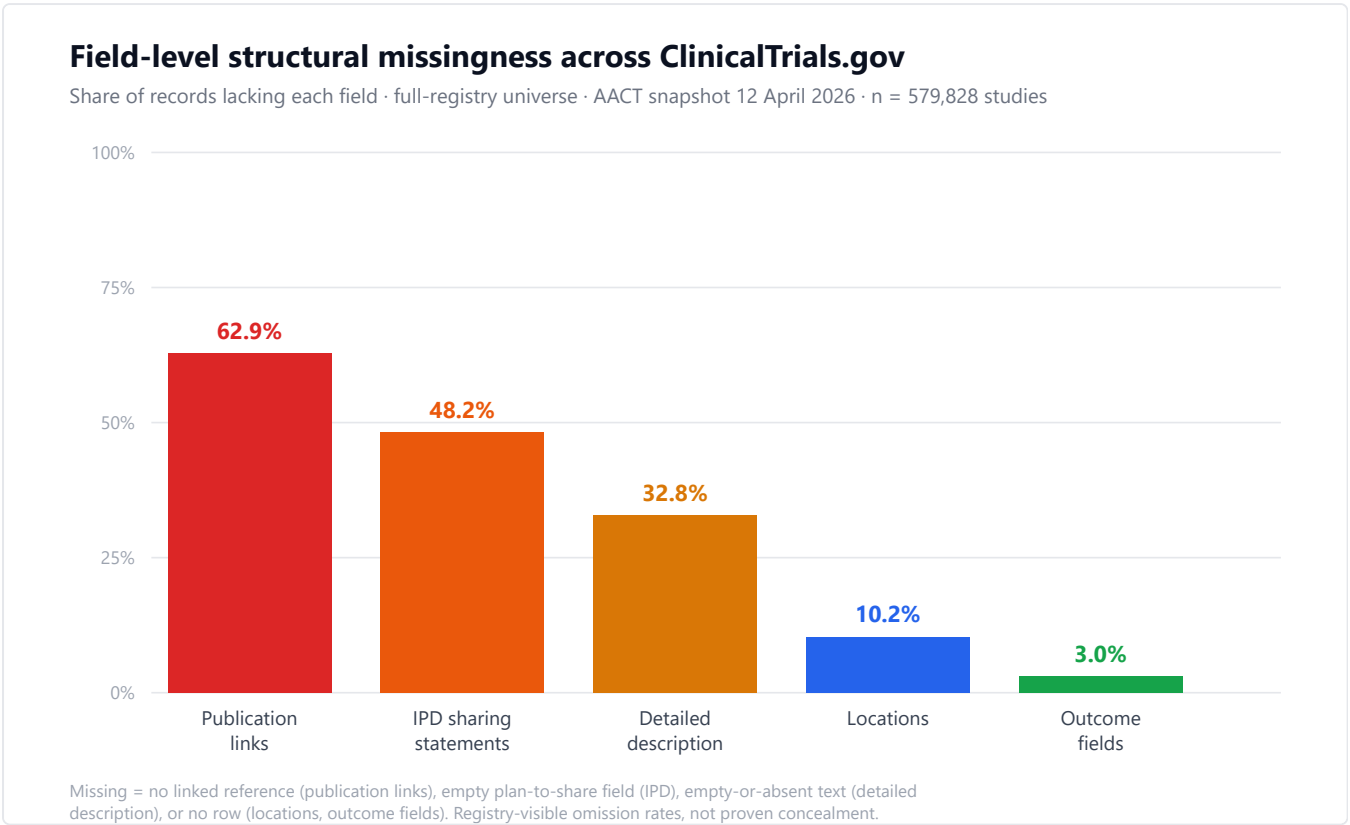


Figure 1. Share of records lacking each field across the full-registry universe (n = 579,828): publication links 62.9%, IPD sharing statements 48.2%, detailed description (empty or absent) 32.8%, locations 10.2%, and outcome fields 3.0%. Rates are registry-visible omission rates, not proven concealment.

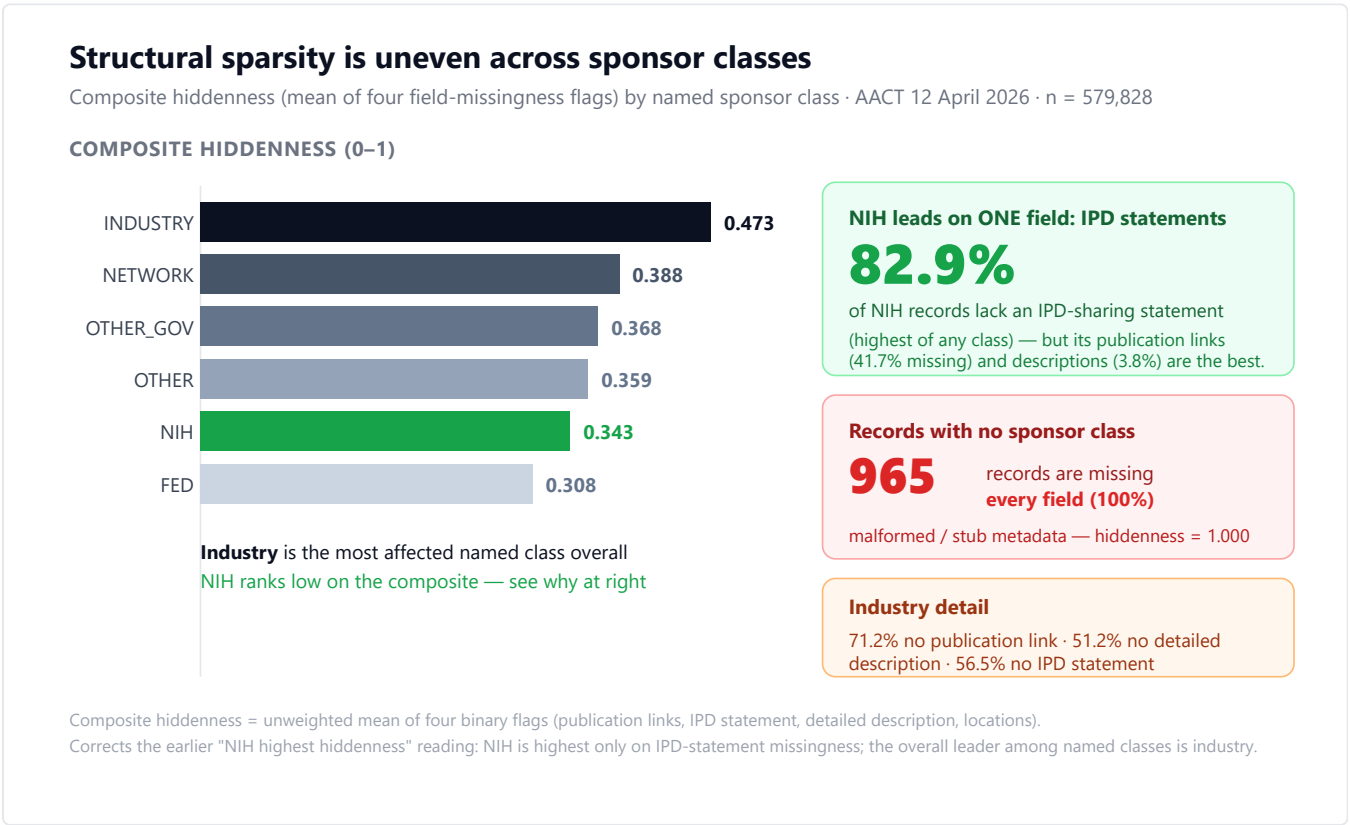


Figure 2. Composite hiddenness (unweighted mean of four field-missingness flags) by named sponsor class. Industry is the most affected named class overall (0.473); NIH ranks low on the composite (0.343) yet carries the highest IPD-sharing-statement missingness of any class (82.9%). A residue of 965 records with no sponsor class are missing every field (malformed metadata).

Boundary. These metrics capture registry-visible information loss, not proven intent to conceal. Rates were recomputed on the AACT 12 April 2026 full-registry snapshot (n = 579,828); all four headline field-omission rates were independently confirmed two ways.

REFERENCES

1. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet*. 2020;395(10221):361–369.
2. Zarin DA, Tse T, Williams RJ, Rajakannan T. Update on Trial Registration 11 Years after the ICMJE Policy Was Established. *N Engl J Med*. 2017;376(4):383–391.

ARTICLE METADATA

Funding. None.

Competing interests. None declared.

Reporting checklist. STROBE (registry-audit variant).

Data availability. Derived entirely from the public AACT mirror of ClinicalTrials.gov (snapshot 2026-04-12); no patient-level data. Recomputation scripts archived with the project.

Ethics. Not required; secondary analysis of publicly available aggregate registry metadata, no human participants and no patient-identifiable data.

E156 micro-paper format (7 sentences, ≤156 words). Target journal: *Synth sis* (Brief Update). Galley generated 2026-06-23. Byline variant Zaki Khan.