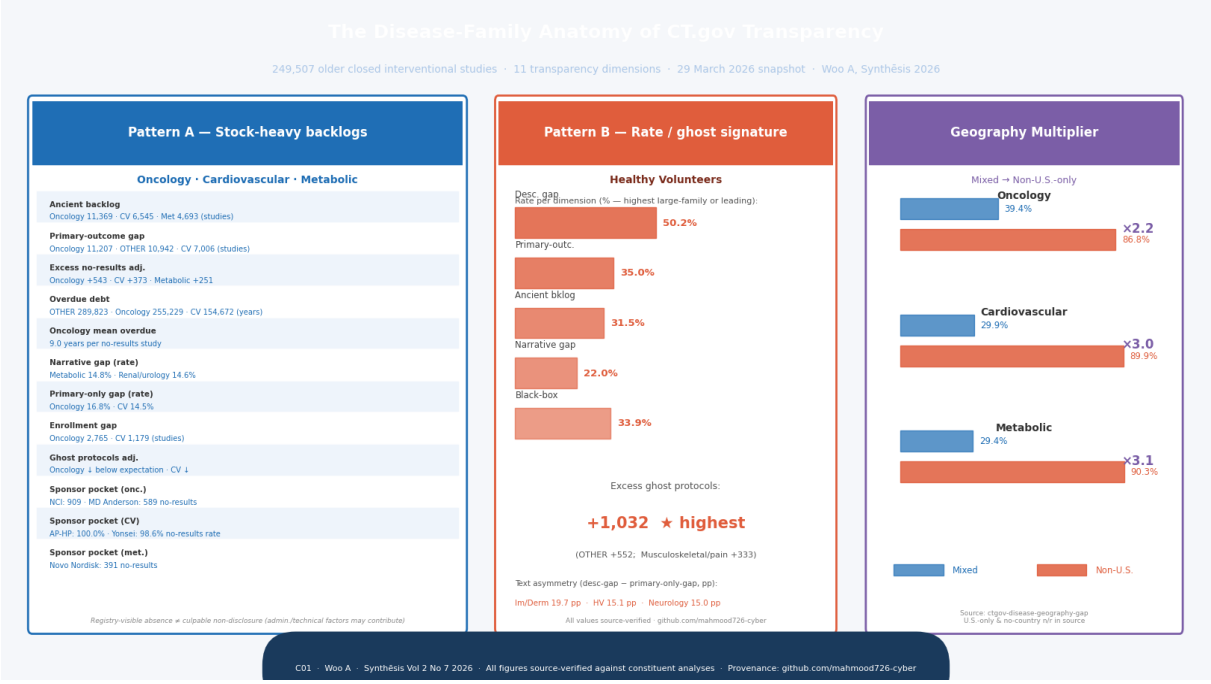

Visual abstract. Two co-existing transparency architectures (Pattern A: stock-heavy oncology/cardiovascular/metabolic; Pattern B: rate/ghost-dominant healthy volunteers) and the geography multiplier. Woo A, Synthēsis Vol 2 No 7 2026.

The Disease-Family Anatomy of CT.gov Transparency: Eleven Condition-Stratified Dimensions across 249,507 Older Closed Interventional Studies

Andrew Woo
St George's, University of London, London, United Kingdom

Synthēsis · Vol 2 No 7 · 2026 | Published 26 June 2026 | doi: 10.66040/synthesis.2026.c01

Abstract

Background. CT.gov transparency has been reviewed at the registry level. It remains unclear, however, how opacity is distributed across therapeutic areas, and whether different types of silence cluster differently by condition family. We decompose transparency using eleven condition-stratified dimensions.

Methods. We analysed 249,507 eligible older closed interventional studies from the 29 March 2026 ClinicalTrials.gov snapshot, assigning each study one condition-family label. We assessed eleven dimensions: (1) ancient backlog, (2) detailed-description gap, (3) enrollment gap, (4) excess no-results stock, (5) excess ghost stock, (6) narrative gap, (7) overdue debt, (8) primary-only gap, (9) primary-outcome gap, (10) sponsor repeaters, and (11) text asymmetry. We also conducted a geography sub-analysis.

Results. Two distinct silence patterns emerged. Oncology and cardiovascular dominate absolute counts: oncology contributed 11,369 ancient-backlog studies, 255,229 cumulative overdue years (9.0 years per study), and 11,207 primary-outcome gaps; cardiovascular contributed 6,545 ancient-backlog studies, 154,672 overdue years, and 7,006 primary-outcome gaps. Healthy volunteers showed the opposite pattern — higher rates: 50.2% description gap, 35.0% primary-outcome gap, 31.5% ancient-backlog rate, 22.0% narrative gap, and 1,032 excess ghost protocols above expectation. Geography amplified these differences: non-U.S. studies showed 86.8–90.3% no-results rates, compared with 29.4–39.4% for mixed-geography studies in the same disease families.

Conclusions. CT.gov silence is structurally patterned by condition family; accountability requires condition-stratified, geography-disaggregated, and sponsor-resolved metrics.

Keywords: ClinicalTrials.gov · non-disclosure · results reporting · condition family · transparency · evidence synthesis

1. Introduction

When a trial's results go unreported, the silence is not uniform. Accountability campaigns have described non-disclosure at the registry level — a single rate, a sponsor league table, a rule-era comparison. They do not, however, ask whether silence is structured by therapeutic area, or whether the *type* of silence (text-stripped page, enrolled-but-absent sample size, fully ghosted protocol, deep antiquity) clusters differently by disease family.

This distinction matters for policy: a study that never posts results but retains its description and endpoint is different from one stripped of both, or one whose enrollment is blank, or one whose sponsor shows 100% non-disclosure. Tools that target missing results as a uniform failure miss these type-specific pockets.

This paper synthesises a thirteen-analysis stratified audit that applied eleven dimensions to the same 249,507-study corpus — the first systematic disease-family anatomy of ClinicalTrials.gov transparency.

2. Methods

Data. We used the 29 March 2026 full-registry snapshot of ClinicalTrials.gov. Eligible studies were interventional and closed, with a primary completion date at least two years in the past (the FDAAA 801 reporting window). Studies with unresolvable completion dates were excluded.

Condition-family assignment. Each study received one condition-family label, derived from MeSH terms and free-text keywords. Families: oncology, cardiovascular, metabolic, infectious disease, healthy volunteers, musculoskeletal and pain, neurology, immunology and dermatology, renal and urology, gastrointestinal and hepatic, and OTHER.

Dimensions. Eleven transparency dimensions were assessed: (1) ancient backlog; (2) detailed-description gap; (3) enrollment gap; (4) excess no-results stock; (5) excess ghost stock; (6) narrative gap; (7) overdue debt; (8) primary-only gap; (9) primary-outcome gap; (10) sponsor repeaters; (11) text asymmetry.

Geography sub-analysis. Studies in the three largest families were stratified into four geography buckets: U.S.-only, U.S.+non-U.S. (mixed), non-U.S.-only, and no-country. All analyses were cross-sectional and observational; no patient-level or non-public data were used.

3. Results

3.1 Corpus

The eligible corpus comprised 249,507 older closed interventional studies from the 29 March 2026 ClinicalTrials.gov snapshot. Reproduction on the AACT 12 April 2026 snapshot yielded 249,580–249,599 eligible studies (Δ +73–+92; +0.03–0.04%), within snapshot-drift tolerance.

Table 1. Corpus and condition-family structure.

Parameter	Value
Total eligible studies	249,507
Snapshot date	29 March 2026
Study type	Older closed interventional

Eligibility criterion	Primary completion ≥2 years before snapshot
Condition families (n)	11 (keyword-derived approximations)
AACT reproduction	249,580–249,599 (Δ +73–92; +0.03–0.04%) — REPRODUCED

3.2 Eleven transparency dimensions

Two transparency architectures emerged (Table 2). Pattern A — oncology, cardiovascular, and metabolic — is defined by absolute stock dominance and positive study-mix-adjusted excess no-results. Pattern B — healthy volunteers — is defined by rate dominance and excess ghost concentration. See Figure 1 for the ancient-backlog stock distribution.

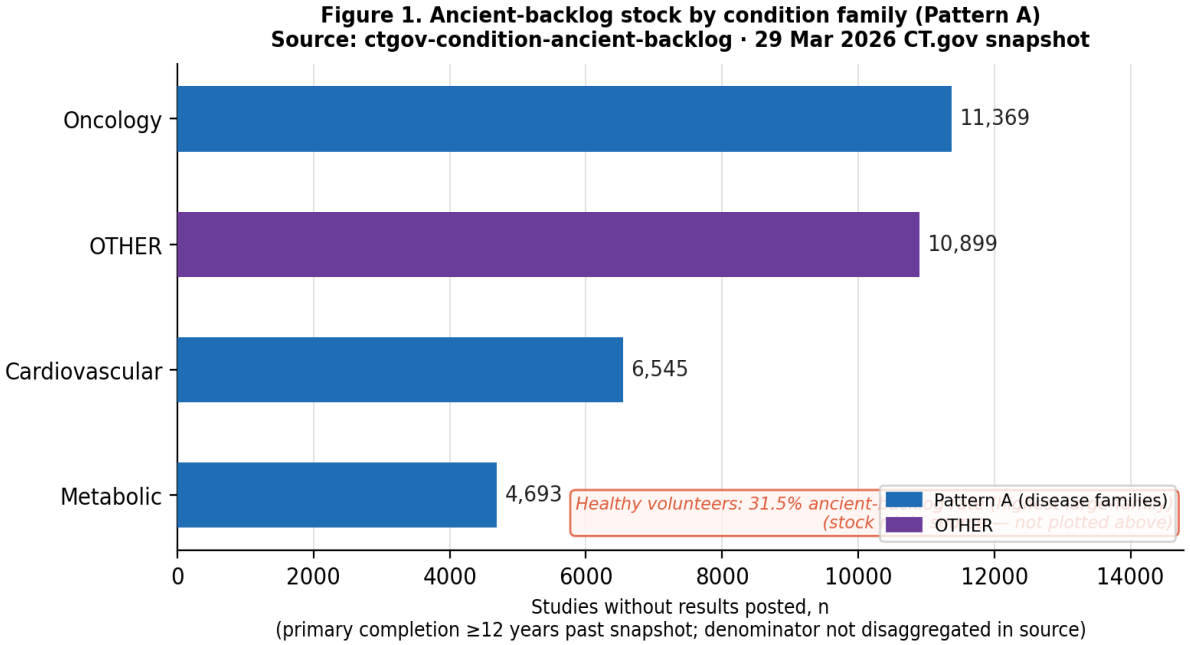

Figure 1. Ancient-backlog stock by condition family. Studies without posted results, primary completion ≥12 years before snapshot. Healthy volunteers rate = 31.5% (highest large-family; stock not in reported top-N).

Table 2. Eleven transparency dimensions — leading condition families.

Dimension	Metric	Leading families and values	Source
1. Ancient backlog	Count (n)	OTHER 10,899; Oncology 11,369; CV 6,545; Metabolic 4,693	ancient-backlog
	Rate (%)	Healthy volunteers 31.5% (highest large-family)	ancient-backlog
2. Desc.-description gap	Count (n)	OTHER 18,641; Oncology 12,321; CV 8,808; HV 7,082	desc-description-gap
	Rate (%)	HV 50.2%; Im/derm 41.7%; Renal/urology 38.3%	desc-description-gap
3. Enrollment gap	Count (n)	Oncology 2,765; OTHER 1,815; CV 1,179; Infect. dis. 747	enrollment-gap
	Rate (%)	Oncology 6.5%; CV 4.5%; GI/hepatic 4.5%	enrollment-gap
4. Excess no-results	Adj. excess	Oncology +543; CV +373; Metabolic +251; HV ~0	excess-watchlist
5. Excess ghost	Adj. excess	HV +1,032; OTHER +552; Musculoskeletal/pain +333	ghost-watchlist
	Direction	Oncology ↓; CV ↓ (below expectation)	ghost-watchlist
6. Narrative gap	Count (n)	OTHER 5,124; Oncology 4,105; CV 3,240; HV 3,100	narrative-gap
	Rate (%)	HV 22.0%; Metabolic 14.8%; Renal/urology 14.6%	narrative-gap
7. Overdue debt	Total yrs	OTHER 289,823; Oncology 255,229; CV 154,672	overdue-debt
	Mean/study	Oncology 9.0 yrs (highest named-family mean)	overdue-debt
8. Primary-only gap	Count (n)	Oncology 7,102; OTHER 5,818; CV 3,766; Infect. dis. 2,584	primary-only-gap
	Rate (%)	Oncology 16.8%; CV 14.5%; Infect. dis. 14.3%	primary-only-gap

9. Primary-outc. gap	Count (n)	Oncology 11,207; OTHER 10,942; CV 7,006; Metabolic 5,006	primary-outcome-gap
	Rate (%)	HV 35.0%; Metabolic 28.9%; Renal/urology 28.8%	primary-outcome-gap
10. Sponsor repeaters		See Table 4	sponsor-repeaters
11. Text asymmetry	Net (n)	OTHER 7,699; MSK/pain 2,521; HV 2,134; CV 1,802	text-asymmetry
	Rate (pp)	Im/derm 19.7 pp; HV 15.1 pp; Neurology 15.0 pp	text-asymmetry
(dim. 4/5)	Black-box %	HV 33.9%	excess-watchlist

n/r = not reported. adj. = adjusted. pp = percentage points. ↓ = below expectation. Source identifiers reference constituent analyses (Provenance Appendix, Section 7). Registry-visible absence ≠ culpable non-disclosure.

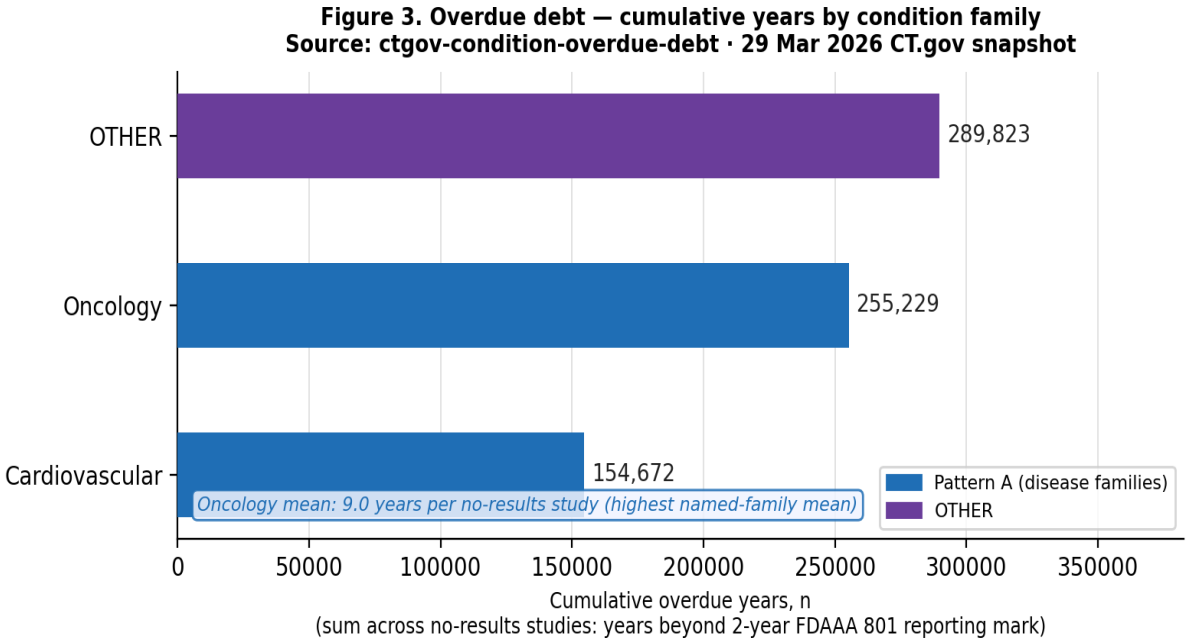

Figure 3. Overdue debt — cumulative years by condition family. Cumulative overdue years (sum across no-results studies: years beyond the two-year FDAAA 801 reporting mark). Oncology mean = 9.0 years/study (highest named-family mean). Source: ctgov-condition-overdue-debt.

3.3 Geography multiplier

Within the three largest families, shifting from mixed U.S.+non-U.S. studies to non-U.S.-only studies approximately doubled to tripled the no-results rate (Table 3, Figure 2). Any completeness estimate that does not stratify by geography will systematically understate the opacity of the non-U.S. evidence base.

Table 3. Geography multiplier — no-results rate by geography bucket.

Condition family	U.S.+non-U.S. (mixed)	Non-U.S.-only	Multiplier	Source
Oncology	39.4%	86.8%	×2.2	disease-geography-gap
Cardiovascular	29.9%	89.9%	×3.0	disease-geography-gap
Metabolic	29.4%	90.3%	×3.1	disease-geography-gap

U.S.-only and no-country buckets not reported in source. Multiplier computed: 86.8/39.4=2.2×; 89.9/29.9=3.0×; 90.3/29.4=3.1×.

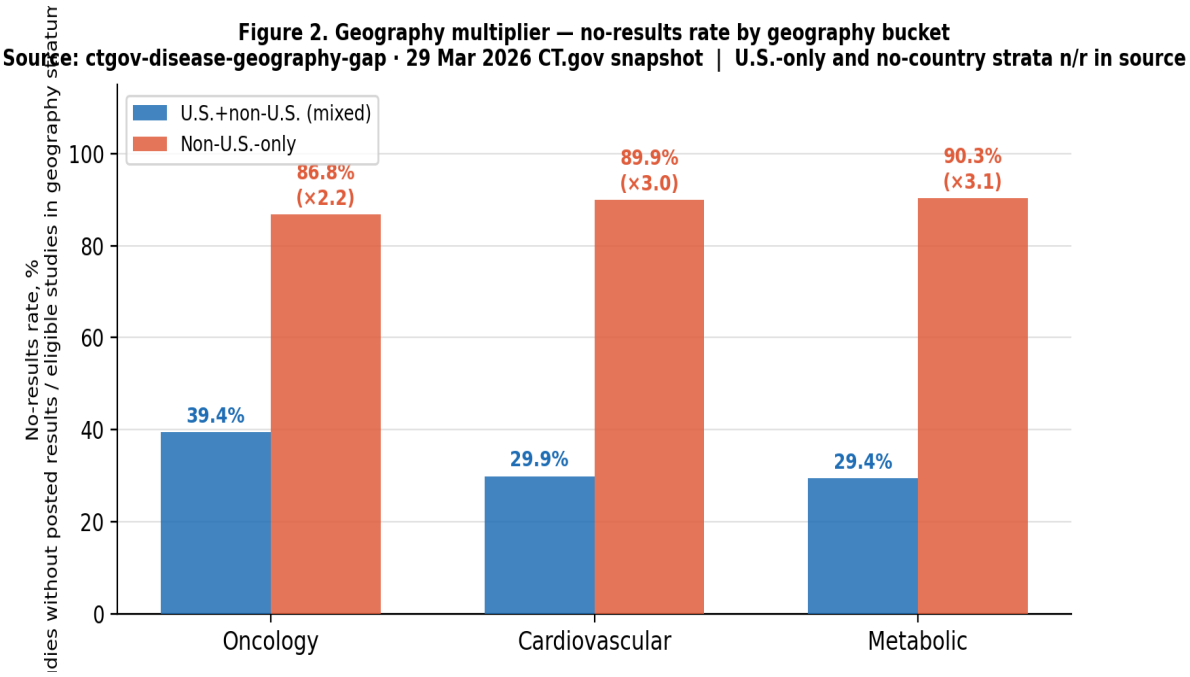

Figure 2. Geography multiplier — no-results rate by geography bucket. No-results rate (%) within three largest condition families by geography stratum. Non-U.S.-only rates approximately 2–3× the mixed U.S.+non-U.S. rates (×2.2 Oncology; ×3.0 Cardiovascular; ×3.1 Metabolic). Source: ctgov-disease-geography-gap.

3.4 Sponsor pockets

Sponsor-level concentration was invisible in registry-wide tables and shifted by disease area (Table 4).

Table 4. Sponsor pockets — leading sponsors within three condition families.

Condition family	Sponsor	No-results studies (n)	No-results rate (%)	Source
Oncology	National Cancer Institute	909	n/r	sponsor-repeaters
Oncology	M.D. Anderson Cancer Center	589	n/r	sponsor-repeaters
Cardiovascular	AP–Hôpitaux de Paris	n/r	100.0%	sponsor-repeaters
Cardiovascular	Yonsei University	n/r	98.6%	sponsor-repeaters
Metabolic	Novo Nordisk	391	n/r	sponsor-repeaters

Values are registry-visible. They do not establish legal breach or intentional non-disclosure. n/r = not reported.

4. Discussion

Silence on ClinicalTrials.gov has structure — an anatomy of its own. Stock-heavy portfolios (oncology, cardiovascular, metabolic) accumulate opacity through portfolio mass combined with moderate rates and concentrated primary-only and enrollment gaps. The 255,229 oncology overdue years means that the typical oncology no-results study is nine years overdue: decades of accumulated obligation, not a recent lapse.

The healthy-volunteer pattern is qualitatively different. Healthy volunteers show the highest rates across nearly every dimension and are the most-ghosted condition family (+1,032 excess ghost protocols), a finding plausibly driven by phase-I PK/safety units operating under weaker disclosure cultures than large disease-portfolio sponsors.

The text-structure analysis adds what results-only audits miss: registry pages fail in two separable ways — narrative lost versus endpoint sentence lost — that are not co-extensive across condition families. Targeting either failure mode alone misses the other entirely.

Geography is the most consequential finding for evidence synthesis: a shift from mixed US-plus-non-US studies to non-US-only studies moves the no-results rate from 29–39% to 87–90% within the same therapeutic family. Any completeness assessment that does not stratify by geography will badly misjudge the non-US evidence base.

The concentrated sponsor pockets shift by disease area, making accountability therapeutically specific. Un-stratified pressure on sponsors lands on the wrong targets; condition-stratified audit changes who is named.

5. Conclusions

Across 249,507 studies and eleven dimensions, condition-family stratification reveals two co-existing silence architectures — stock-heavy disease backlogs and high-rate healthy-volunteer ghosting and text-stripping — together with an approximately two- to three-fold geography multiplier and therapeutically specific sponsor pockets that are invisible in aggregate analyses. Effective accountability requires condition-stratified, geography-disaggregated, and sponsor-resolved metrics; registry-wide aggregates alone are insufficient.

6. Limitations

Condition families are keyword-derived, non-exclusive approximations; boundary cases introduce misclassification. A single cross-sectional snapshot is itself a limitation. The two-year reporting clock is the baseline statutory window, but enforcement lag and discretionary extensions are not modelled. Geography uses recorded site locations. The excess and ghost watchlists depend on the prior-wave study-mix model and are not independently validated here. Sponsor identifiers are taken from registry text, with no corporate-parent or CRO consolidation applied. Registry-visible absence does not establish culpable non-disclosure.

References

1. Food and Drug Administration Amendments Act (FDAAA) of 2007. Pub. L. No. 110-85.

2. Clinical Trials Registration and Results Information Submission. 42 CFR Part 11. 2016.

3. Zarin DA et al. N Engl J Med 2011;364:852–860. PMID 21366476.

4. Anderson ML et al. N Engl J Med 2015;372:1031–1039. PMID 25760355.

5. DeVito NJ et al. BMJ 2020;368:m128. PMID 31958402.

6. Tasneem A et al. PLOS ONE 2012;7:e31014. PMID 22438982.

7. Chan A-W et al. JAMA 2004;291:2457–2465. PMID 15161896.

8. Ross JS et al. BMJ 2012;344:d7292. PMID 22214755.

9. Zarin DA et al. N Engl J Med 2019;381:264–273. PMID 31722160.

10. Hart B et al. BMJ 2012;344:d7323. PMID 22214754.

11. Macleod MR et al. Lancet 2014;383:101–104. PMID 24411643.

7. Provenance Appendix

Constituent analyses underpinning this synthesis. All repositories at github.com/mahmood726-cyber.

Short identifier	Analysis name	Repository
ancient-backlog	CT.gov Condition Ancient Backlog	ctgov-condition-ancient-backlog
detailed-description-gap	CT.gov Condition Detailed-Description Gap	ctgov-condition-detailed-description-gap
enrollment-gap	CT.gov Condition Enrollment Gap	ctgov-condition-enrollment-gap

excess-watchlist	CT.gov Condition Excess Watchlist	ctgov-condition-excess-watchlist
ghost-watchlist	CT.gov Condition Ghost Watchlist	ctgov-condition-ghost-watchlist
narrative-gap	CT.gov Condition Narrative Gap	ctgov-condition-narrative-gap
overdue-debt	CT.gov Condition Overdue Debt	ctgov-condition-overdue-debt
primary-only-gap	CT.gov Condition Primary-Only Gap	ctgov-condition-primary-only-gap
primary-outcome-gap	CT.gov Condition Primary-Outcome Gap	ctgov-condition-primary-outcome-gap
sponsor-repeaters	CT.gov Condition Sponsor Repeaters	ctgov-condition-sponsor-repeaters
text-asymmetry	CT.gov Condition Text Asymmetry	ctgov-condition-text-asymmetry
disease-geography-gap	CT.gov Disease Geography Gap	ctgov-disease-geography-gap