NOTES FROM THE BOT LAB · EDITION OF NOVEMBER 25, 2026

We Pointed Our Search Math at 3,000 FDA Device Filings After Hours

By Will Tygart · Tygart Media · Canonical edition at https://tygartmedia.com/we-pointed-our-search-math-at-3000-fda-device-filings/

Scope & Regulatory Note: This essay describes an independent after-hours data-engineering experiment over public, de-identified government records (OpenFDA, ClinicalTrials.gov, and California CHHS Open Data) containing zero patient health information. Nothing here constitutes medical, clinical-trial, or FDA regulatory legal advice.

When a shop builds a diagnostic tool that actually works on its day job, sooner or later curiosity gets the better of you after hours. You pour another cup of coffee, clear off the workbench, and point the math at something completely outside your industry just to see whether you built a real instrument — or just memorized your own backyard.

All week we had been using our paired rival-autopsy engine to figure out why AI systems pick one technical web page over another. Last night we asked a nerdier question: what happens if we point that exact two-stage comparison pipeline at public government medical paperwork — zero patient data, zero marketing — and forbid ourselves from fudging the score when it breaks?

Step One: Watching Our Web Math Faceplant

First we pulled 3,000 Traditional medical-device clearances from the OpenFDA 510(k) database spanning 2022 through 2025. Despite everyone talking about 90-day FDA reviews, only 26.4 percent cleared inside 90 days; the median took 143 days, and 37.87 percent dragged past 180 days. To test like against like, we downloaded 228 full-text 21 CFR 807.92 summary PDFs — 114 fast-versus-slow pairs locked 1-to-1 inside the exact same 41 three-letter FDA product codes.

Right away we caught our first trap: FDA staples its signed approval cover letter onto pages one through three of every PDF. Leaving those cover pages attached let a model look smart (0.5943 AUC) just by sniffing review-office boilerplate. Once we quarantined the cover letters and ran our web-citation feature extractor blind on the applicant summaries alone, it scored a LightGBM ROC-AUC of 0.4857. Worse than flipping a quarter.

Just to make sure that was not a fluke, we pulled 600 Phase 2 and Phase 3 interventional protocols from ClinicalTrials.gov — 300 completed versus 300 terminated across 64 disease clusters. Blind web-SEO math scored 0.5374 there too, and the reason was fascinating: the core rule inverted. On the web, packing crisp numeric cutoffs near the top of a guide wins citations. In a clinical trial, stuffing strict lab and organ washout gates into the eligibility list (mean 5.43 gates on terminated trials versus 2.47 on completed, p = 2.51e-8) chokes off patient enrollment — which accounted for 38.0 percent of dead trials. Once we split inclusion from exclusion bloat, accuracy jumped +0.1249 to 0.6624 (and 0.6865 on post-2018 trials).

Catching the 99.5 Percent Mirage on Insurance Appeals

Before coming back to the FDA PDFs, we ran one more honesty check across 800 Independent Medical Review appeals from the California Health and Human Services Open Data Portal, where patients appealed health-plan denials and won 71.25 percent of the time. Fed the raw case narrative, a naive model immediately bragged about a 0.9950 ROC-AUC.

Anytime a model hits 99.5 percent on messy human decisions, check its pockets. Even after deleting the final verdict sentence, the reviewer's hindsight phrasing still leaked 0.8466. Only when we stripped every post-decision adjective down to the honest intake floor did the real pre-decision signal settle at 0.6216 (0.6361 on Random Forest) — isolating one plain kitchen-table receipt that more than doubled a patient's odds of overturning a denial: documented prior therapy failure (odds ratio 2.35, p = 0.00663).

The Family-Tree Ghost Hiding in 27.6 Percent of Cleared Filings

That brought us back to the 228 FDA 510(k) device PDFs. A 510(k) clearance works like a family tree: a manufacturer argues that a new device is substantially equivalent to an older, already-cleared predicate device identified by a six-digit K-number. When we swapped out web-SEO features for actual regulatory genealogy — applicant geography (U.S. applicant odds ratio 2.96), citing your own prior device family (odds ratio 2.11), and predicate age (older predicates slowed reviews down, p = 0.024) — within-code discrimination rose +0.1177 to 0.6062.

Then we traced all 544 unique K-numbers cited inside those 228 cleared summaries against the OpenFDA Medical Device Recall Registry all the way back to 1976. Sixty-three of the 228 cleared filings — 27.6 percent — cited at least one predicate device with an FDA recall on its record, spanning 97 distinct recalled K-numbers tied to component selection, nonconforming materials, process design, and labeling mix-ups. Fewer than 4.4 percent included an explicit clause attesting to no design-related recall history.

To keep both feet on the honest floor: citing a recalled predicate did not mean those 63 devices were unsafe or rejected — every one of them cleared (33 fast versus 30 slow, p = 0.631), and rejected NSE filings are not public in OpenFDA at all. What it showed is human nature. Busy professionals in every technical trade pull the last approved binder off the shelf, copy the precedent table forward, and rarely audit forty years of genealogy unless a script checks the tree in two seconds.

Why Stay Late Pointing Tools at Strange Problems?

We are not opening a medical-device law firm tomorrow morning. Pointing our search tools at 4,400 government records was just a way to pressure-test our own shop discipline: strip the cover letter before you brag about a score, expect domain rules to flip upside down until you read the data, and check the foundation before you copy yesterday's template. Everything else is just keeping your eyes on the mountain.

Public .gov Dataset Tested	Blind Web-SEO Port (Wave 1)	Domain-Adapted Signal (Wave 2)	Biggest Surprise Under the Hood
OpenFDA 510(k) Clearances (3,000 macro; 228 paired 21 CFR 807.92 PDFs)	0.4857 ROC-AUC (worse than a coin flip)	0.6062 ROC-AUC within exact same product code (+0.1177, p = 0.027)	63 of 228 (27.6%) cleared devices cited >= 1 recalled predicate back to 1976
ClinicalTrials.gov Phase 2/3 (600 completed vs. terminated protocols)	0.5374 ROC-AUC (noise)	0.6624 ROC-AUC (+0.1249, p < 0.00001; 0.6865 post-2018)	Rule inverted: strict numeric lab cutoffs strangled patient recruitment
California CHHS IMR Appeals (800 health-plan denials)	0.9950 naive / 0.8466 verdict-stripped (hindsight leak)	0.6216 LGBM / 0.6361 RF honest pre-decision floor	Documented prior therapy failure more than doubled overturn odds (OR = 2.35)

Key Questions Answered: Why did web-search features score 0.4857 on FDA 510(k) summaries? Generic web-SEO metrics (such as HTML word count and generic ISO regexes) ignore product-code-controlled regulatory drivers like predicate device age, self-predicate continuity, and applicant review logistics. **Does citing a predicate device with a past recall mean an FDA 510(k) submission fails?** No. All 63 of the 228 submissions in our sample that cited a recalled predicate still earned FDA substantial-equivalence clearance, though flagging recalled predicates pre-filing helps teams address design differences explicitly. **Why did a naive model score 0.9950 on California IMR insurance appeals before dropping to 0.6216?** Public case summaries are written after the decision is made; until post-hoc reviewer phrasing is stripped down to pre-decision intake facts, models cheat by reading hindsight language.