

Who Marks the Trial Record When a Drug Is Withdrawn? Regulatory-Status Metadata on Pivotal Publications for US Safety Withdrawals, 1990 to 2024

Phani Kurada

pkurada1@gmail.com

Independent researcher

Research Article

Keywords: Drug withdrawal; Regulatory-status metadata; Traceability of the scientific record; Research integrity; Pharmacovigilance; Automated evidence synthesis; Pivotal clinical trials; Food and Drug Administration; Rofecoxib; Bibliographic infrastructure

Posted Date: August 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10270651/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Who Marks the Trial Record When a Drug Is Withdrawn? Regulatory-Status Metadata on Pivotal Publications for US Safety Withdrawals, 1990 to 2024

Phani Kurada

Independent researcher, Littleton, Massachusetts, USA

ORCID: <https://orcid.org/0009-0007-4510-024X>

Correspondence: Phani Kurada, pkurada1@gmail.com

Abstract

Background: When the FDA withdraws a drug for safety reasons, the clinical trials that supported its approval remain in the peer-reviewed literature. Clinicians, systematic reviewers, guideline developers, and automated evidence-synthesis tools retrieve those trials through standard bibliographic databases. We asked a question prior to any question of correction or blame: when the regulatory record and the bibliographic record are compared, does the pivotal trial record of a withdrawn drug carry any machine-discoverable signal of the withdrawal, or are the two records discordant? A machine-discoverable signal is a formally linked or structured notice returned with the paper's record, such as a retraction, an expression of concern, or a correction.

Methods: We assembled a cohort of US drugs withdrawn for safety reasons between 1990 and 2024 using the Onakpoya withdrawal registry, FDA Federal Register notices, and confirmatory regulatory sources. For each drug, pivotal trial publications were classified into three identification tiers: FDA-confirmed (Tier 1), retrospective-review-inferred (Tier 2), and no identifiable published pivotal paper (Tier 3). Correction status for each pivotal paper was determined by manual inspection of the live PubMed record, with Crossref, the Retraction Watch database, and Google Scholar as parallel checks. The primary endpoint was a safety-related paper-level action; two secondary endpoints captured any formal action regardless of content, and any machine-discoverable withdrawal-status annotation. Each drug's withdrawal was classified with a pattern typology distinguishing contamination of the pivotal paper itself (Pattern A), failure of a surrogate or clinical-outcome claim on re-examination (Pattern B), and harm in a domain the pivotal trial was not designed to measure (Pattern C).

Results: The primary cohort comprised 29 drugs spanning three decades, seventeen therapeutic classes, and five regulatory pathways, yielding 63 indexed pivotal efficacy papers. One paper, the VIGOR trial of rofecoxib, received an Expression of Concern, the only action meeting the primary endpoint, a rate of one in 63 (approximately 1.6 percent), with zero retractions and no withdrawal-status annotation on any pivotal paper. Most withdrawals, 22 of 29, were Pattern C. In two cases the approval-relevant interpretation of

the pivotal evidence was later reversed or no longer sustained, encainide and drotrecogin alfa, yet no signal attached to the trial record.

Conclusions: The peer-reviewed record carried no structured withdrawal-status annotation on any pivotal paper, and only one safety-related formal action among 63. The regulatory record and the bibliographic record are therefore discordant: the regulatory record documents withdrawal, while the pivotal-paper record carries no corresponding status signal, and the two are not linked. A reader who retrieves the pivotal trial of a withdrawn drug finds the paper unchanged, with no marker of the withdrawal. This is not an editorial failure. We found no routinely implemented mechanism that placed withdrawal-status metadata on the records examined, and responsibility for placing one appears fragmented across journals, regulators, and bibliographic services. The low correction rate is the measurement behind this finding, not a claim that correction was owed. The gap now matters more, because automated systems that rely principally on indexed article-level metadata may not detect a withdrawal that is absent from that metadata.

Keywords: Drug withdrawal; Regulatory-status metadata; Traceability of the scientific record; Research integrity; Pharmacovigilance; Automated evidence synthesis; Pivotal clinical trials; Food and Drug Administration; Rofecoxib; Bibliographic infrastructure

1. Background

Drug regulatory authorities require evidence of favorable benefit-to-harm balance before granting marketing authorization. This evidence typically consists of pivotal randomized clinical trials, published in peer-reviewed journals after or contemporaneously with regulatory approval. These publications constitute the medical literature's record of the trial evidence supporting a drug's approved use.

When a drug is subsequently withdrawn from the market for safety reasons, several institutional responses are possible. Regulatory agencies revoke or modify approval. Clinical practice guidelines update recommendations. Prescribing databases mark the drug as discontinued. Pharmacovigilance systems record the adverse events. Whether the *peer-reviewed pivotal literature itself* receives any formal correction action is less clear. The established correction mechanisms of the peer-reviewed literature, retraction and expression of concern under COPE and ICMJE guidance, are designed to address fraud, error, or misconduct in the claims a paper itself makes; there is no formal mechanism for annotating a paper whose clinical utility has been overtaken by a later regulatory action that its own data did not contain. To our knowledge, no systematic study has measured this rate.

The distinction matters for three reasons. First, the peer-reviewed literature is where clinicians, researchers, and increasingly AI-based systems seek evidence about a drug's efficacy, and the machine-discoverable signal on a pivotal paper is the mechanism by which citation databases surface any post-publication concern about it. Second, retracted or corrected papers are, in principle, marked as such in PubMed, Crossref, and Retraction Watch, and downstream consumers of the literature (including systematic reviews,

guideline developers, and automated evidence-synthesis tools) rely on this metadata rather than on manual regulatory-history retrieval. Third, drug withdrawals for serious harm, including withdrawals following demonstrated harm-signal misrepresentation as in the Vioxx (rofecoxib) case, represent moments at which one might expect the medical literature to record what has been learned.

This paper measures how often that recording occurs. We report a systematic case series of US safety withdrawals 1990-2024, examining the machine-discoverable correction status of the pivotal trial publications for each withdrawn drug. We deliberately scope the endpoint to formal, database-registered correction actions. Editorial commentary, letters to the editor, and discussion of a drug's withdrawal in later review articles all constitute a form of literature response, but they do not carry the paper-level metadata that citation databases and downstream tools surface. Our question is narrower: does the pivotal paper's own record, as retrievable through standard bibliographic infrastructure, carry a signal that the drug it supported was withdrawn?

One objection should be met at the outset. It could be argued that a pivotal trial, by design, does not contain the later safety signal, since a drug whose trial showed that signal would not reach the market. It could be argued further that a paper reporting an accurate result needs no correction. Both points are correct, and we accept them. Our claim is neither that these papers are wrong nor that correction was owed. Our claim is narrower and prior to the question of correction. The bibliographic record of a withdrawn drug's pivotal trial carries no structured signal of the withdrawal itself, whether corrective or informational. Two cases show this holds even when the approval-relevant interpretation of the pivotal evidence was later overturned. The encainide trials measured a real effect on a surrogate endpoint that a later trial showed did not predict survival. The drotrecogin alfa trial reported a mortality benefit that a later trial did not replicate. In both, the basis on which approval rested was no longer sustained, the drug was withdrawn substantially for that reason, and the trial record still shows nothing.

2. Methods

2.1 Study design

Systematic case series with structured case-file extraction and cross-drug quantitative synthesis.

2.2 Study universe: FDA safety withdrawals 1990-2024

The primary cohort consists of drugs withdrawn from the US market for safety reasons between 1990 and 2024. We identified the cohort by filtering the drug-withdrawal registry compiled by Onakpoya, Heneghan, and Aronson (BMC Medicine 2016; 14:10) [1], the most comprehensive published inventory of post-marketing withdrawals, for entries meeting three inclusion criteria: (1) US market withdrawal, (2) year of first withdrawal 1990-2024 inclusive, (3) withdrawal for adverse drug reactions rather than commercial reasons, contamination, or reformulation. Onakpoya's registry covers 1953-2013; entries for US safety withdrawals 2014-2024 were added from FDA Federal Register notices and cross-

referenced to the drug discontinuation analysis by Rodriguez-Monguio et al. (2023), which identified 38 US safety withdrawals in the period 1980-2021.

Exclusion criteria applied to Onakpoya candidate entries were: (a) never US-approved (removed: lumiracoxib, sitaxentan, parecoxib, ximelagatran, rimonabant, sertindole); (b) primary withdrawal reason was commercial failure, product reformulation, or non-safety regulatory action, even if a safety signal was later identified. Inhaled insulin (Exubera) is excluded on this basis, as Pfizer's 2007 withdrawal was driven by commercial factors and the later lung-cancer signal in follow-up did not itself trigger regulatory action; (c) restricted-not-withdrawn drugs that carry boxed warnings and REMS-like restrictions but remain FDA-approved for at least one indication. Tolcapone (Tasmar) is excluded on this basis, as it was restricted rather than withdrawn from the US market.

Drugs withdrawn from the US market for safety reasons and subsequently reintroduced under Risk Evaluation and Mitigation Strategy (REMS) or restricted-access programs (alosetron, natalizumab, gemtuzumab ozogamicin, tegaserod) are included in the primary cohort with a "withdrawn-and-reintroduced" tag. The correction-status question applies to the original pivotal papers regardless of eventual reintroduction; the tag permits stratified analysis to test whether reintroduced-drug pivotal papers show different correction behavior than fully-withdrawn-drug papers.

The full analyzed set is provided as Table 1 and comprises 31 drugs. The strict primary cohort comprises 29 drugs after holding out LAAM/ORLAAM and nefazodone as a border-case sensitivity set.

2.3 Pivotal paper identification: three-tier classification

For each drug in the cohort, "pivotal paper" was defined as a peer-reviewed publication reporting a clinical trial that supported the drug's FDA approval for its primary indication. Because FDA archive completeness varies substantially across the study period, we classified pivotal papers into three tiers to allow tier-stratified analysis:

- **Tier 1 (FDA-confirmed pivotal papers).** Papers explicitly cited in the FDA medical review, Advisory Committee briefing document, Summary Basis of Approval, or FDA-approved drug label at time of original approval. For post-2000 approvals, Tier 1 identification is generally straightforward because FDA's online archive (accessdata.fda.gov) preserves these primary documents. Where dose-supplement approvals were separately consequential to the withdrawal, as with cerivastatin's 0.8 mg dose supplement, pivotal papers supporting the specific supplement were also classified as Tier 1.
- **Tier 2 (retrospective-inferred pivotal papers).** Papers not directly recoverable from FDA primary documents but identified as pivotal by authoritative post-approval reviews, for example Karha and Topol's *Cleveland Clinic Journal of Medicine* review of rofecoxib [2], or Sorkin and Heel's *Drugs* review of terfenadine [3]. Tier 2 identification is necessary for pre-1995 approvals where FDA online archives are incomplete, and represents a defined fallback rather than a discretionary choice.

- **Tier 3 (no identifiable pivotal paper).** Drugs whose pre-approval trials were never substantially published in the peer-reviewed literature. Bromfenac (Duract) is the pilot example: FDA reviewer records document the pre-approval hepatotoxicity signal, but almost none of the pre-approval efficacy or safety data appeared as pivotal peer-reviewed publications. Goldkind and Laine's 2006 systematic review of NSAIDs withdrawn for hepatotoxicity [4] explicitly notes "There were very limited published data on... bromfenac." Tier 3 is not missing data; it is an emergent finding, because if pre-approval evidence is never published, no correction mechanism can operate on it.

The primary analysis includes all indexed Tier 1 and Tier 2 pivotal papers. A Tier 1-only analysis is conducted as a sensitivity analysis, to assess the effect of retrospective pivotal-paper identification. Tier 3 drugs are reported as a distinct category, described qualitatively.

Pivotal-paper identification for post-approval trials cited in subsequent label expansions (for example, VIGOR / rofecoxib for the 2002 rheumatoid arthritis expansion) is included when the trial appears in the drug's approved-use evidence base. Trials investigating potential *new* indications not granted (for example, APPROVe / rofecoxib for colorectal polyp chemoprevention) are noted but not classified as pivotal. By the same test, a post-approval trial with no documented role in the approved-use evidence base is not classified as pivotal even where it studies the approved indication; on this criterion two post-approval trials, Levy et al. (*Circulation* 1995) [5] for aprotinin and Olanow et al. (*Mov Disord* 1994) [6] for pergolide, are held from the strict pivotal denominator and enter only the held-trials sensitivity check reported in Appendix B, while each drug's pre-approval registration trials count as pivotal in the usual way. This is the same documented-role criterion that admits VIGOR, whose rheumatoid-arthritis labeling role is documented, and excludes APPROVe, applied identically regardless of a paper's publication date; both held trials are correction-clean, so the hold does not change the primary-endpoint count.

2.4 Pre/post-2000 sensitivity analysis

Because both pivotal-paper identification (via FDA archive completeness) and the correction infrastructure itself (PubMed linked-articles, Crossref DOI metadata, Retraction Watch) have strengthened substantially over the study period, we conduct a pre-specified sensitivity analysis stratified by decade of first US approval: pre-2000 approvals versus 2000-onward approvals. This analysis addresses the possibility that any observed low correction rate is an artifact of infrastructure incompleteness in earlier decades rather than a stable structural finding.

2.5 Correction-status endpoint and assessment

The endpoint is a *machine-discoverable paper-level action on the pivotal publication*, detected through standard bibliographic infrastructure and then graded by content. Detection used four parallel database checks; a pivotal paper was coded as carrying an action if any of the following returned a documented correction, retraction, expression of concern, or erratum linked to the paper by identifier:

- PubMed linked-articles field for the paper's PubMed ID (retraction notices, corrections, comments, and expressions of concern all appear here when formally linked by the journal via NLM data-supply channels)
- Crossref API record for the paper's DOI (formal correction notices are captured in Crossref metadata via publisher-registered relationships)
- Retraction Watch database search
- Google Scholar search for "erratum" or "correction" or "expression of concern" citing the paper's title, DOI, or citation

Papers with only editorial commentary, letters to the editor, or critical review-article discussion, but with no formally linked correction record retrievable through the four checks above, were coded as *no formal action*. This scoping is deliberate. Editorial discussion, review-article critique, and post-approval regulatory notices all constitute a form of literature response to a drug withdrawal, but they do not attach to the pivotal paper as machine-discoverable metadata. Our endpoint measures the signal a downstream consumer of the literature (clinician, systematic reviewer, guideline developer, or automated evidence-synthesis tool) would encounter when retrieving the pivotal paper through standard bibliographic infrastructure.

Any action detected was graded against a prespecified three-level endpoint hierarchy, ordered from the strictest signal to the weakest:

- **Primary endpoint: safety-related paper-level action.** A retraction, expression of concern, or correction/erratum whose content concerns the safety, efficacy, or risk-benefit evidence that led to the withdrawal. The study's headline rate reports this endpoint.
- **Secondary endpoint A: any formal paper-level action, regardless of content.** Any retraction, expression of concern, or correction/erratum on the pivotal paper, including routine numerical corrections unrelated to the withdrawal. A paper can satisfy this endpoint without satisfying the primary endpoint.
- **Secondary endpoint B: withdrawal-status annotation.** Any machine-discoverable notice on the paper's bibliographic record indicating the drug was withdrawn, suspended, or reintroduced, independent of any correction to the paper's scientific content.

This separation makes borderline records unambiguous: a routine erratum with no bearing on the withdrawal satisfies secondary endpoint A but not the primary endpoint, and is reported as such. The safety-related action was retained as the prespecified primary endpoint from the original study design. The withdrawal-status annotation, secondary endpoint B, is the outcome that most directly addresses the record-linkage question, and its result, zero across all pivotal papers, is reported alongside the primary rate throughout.

Correction actions, when present, were subclassified by type:

- **Retraction.** The paper is formally withdrawn from the literature record
- **Expression of concern.** The journal formally records substantive concerns about the paper's integrity without full retraction
- **Correction / erratum.** A specific error is formally corrected while the paper remains valid
- **No formal action.** None of the above appears in the paper's record

2.6 Pattern typology

Each drug was classified by withdrawal reason using the pattern typology below, developed and refined against the cases examined. Pattern B is divided into two subtypes, B1 and B2, distinguishing failure of a surrogate from non-replication of a directly-measured clinical outcome:

- **Pattern A.** Pivotal paper directly contaminated. The paper misrepresented, omitted, or selectively reported data material to the safety signal that later triggered withdrawal. Example: VIGOR (Bombardier et al. NEJM 2000), which omitted three myocardial infarctions in the rofecoxib arm known to authors before publication.
- **Pattern B1.** Surrogate-to-clinical-outcome failure. The pivotal paper measured a validated surrogate endpoint (LDL cholesterol, tumor response, ventricular-ectopy suppression) correctly, and the drug's effect on that surrogate was real, but the surrogate later failed to translate into clinical benefit. Example: encainide, whose pivotal trials correctly demonstrated ventricular-ectopy suppression that CAST-I later showed did not translate into a survival benefit.
- **Pattern B2.** Pivotal clinical-outcome non-replication. The pivotal paper measured a clinical endpoint directly and reported benefit, but that clinical result was not replicated in a later confirmatory trial and the drug was withdrawn substantially because of the non-replication. Example: drotrecogin alfa, whose PROWESS mortality reduction was not confirmed in PROWESS-SHOCK.
- **Pattern C.** Harm in unmeasured domain. The pivotal paper correctly demonstrated efficacy on its measured endpoint, and the drug's efficacy claim was not disputed. Withdrawal occurred for a harm in a physiological domain the pivotal trials were not designed or powered to detect. Example: troglitazone (idiosyncratic hepatotoxicity in a paper measuring HbA1c reduction).

One drug, natalizumab, falls outside this A/B/C assignment and is treated as a defined exceptional sub-case: its pivotal papers reported the withdrawal-triggering harm, progressive multifocal leukoencephalopathy, directly (Section 3.5).

Drugs frequently exhibit hybrid or intermediate patterns. Classification records the *predominant* pattern with hybrid tags noted.

2.7 LLM-assisted workflow and audit protocol

Following current transparency practice for AI-assisted research, we disclose that a large language model (Claude, Anthropic) was used to assist with pivotal-paper identification from FDA documents and retrospective reviews, classification of correction status from PubMed and Crossref records, and drafting of case-file text. The author retained decision authority for every case: (1) confirmation of pivotal-paper inclusion and tier classification, (2) adjudication of borderline correction-status cases, (3) pattern typology classification, and (4) all interpretive claims.

The workflow discipline was formalized as “LLM extracts, human adjudicates”: the language model performed literature retrieval and initial candidate classification, and the author independently verified each determination against primary sources before it was recorded in the case file.

Audit. A prespecified audit of the primary cohort was conducted. From the enumerated pivotal-paper pairs, a sample of 36 (38 percent, above the prespecified 25 percent minimum) was drawn as a census of all pairs carrying any recorded formal action, a census of all pairs from drugs whose pivotal papers were classified Tier 2 (retrospective-inferred, the identifications least supported by FDA archives), and a seed-fixed 25 percent random draw of the remainder. For each sampled pair the author (i) where the pair resolved to a specific publication, confirmed against the FDA record that the paper was a pivotal efficacy publication supporting approval and noted any pivotal paper cited by the FDA source but absent from the case file; (ii) independently re-checked correction status against PubMed linked-notice fields, Crossref, the Retraction Watch Database, and Google Scholar, without reference to the case file; and (iii) compared the independent finding with the recorded classification, resolving any discrepancy by reading the correction notice itself. A separate large language model performed first-pass record retrieval constrained to return only source links for verification; every correction status was confirmed by the author against the live record, and none was recorded on the model’s assertion alone. The prespecified re-audit trigger was correction-status error above 5 percent on the primary endpoint. Audit results are reported in Section 3.3. Aggregate case-file entries that the initial pass could not resolve to a specific citation were subsequently resolved against the FDA medical review where one was available online, and otherwise against the published literature; each resolved publication was correction-checked by the same procedure, and none carried a safety-related action.

3. Results

3.1 Cohort derivation

Filtering the Onakpoya registry to US safety withdrawals in the 1990 to 2024 window, then cross-referencing FDA Federal Register notices and Rodriguez-Monguio et al. (2023) [7], produced 39 candidate entries. Applying the exclusion criteria of Section 2.2 removed eight: six drugs never approved in the United States (lumiracoxib, sitaxentan, parecoxib, ximelagatran, rimonabant, sertindole); one drug whose primary withdrawal was commercial rather than safety-driven (inhaled insulin, Exubera); and one drug that was

restricted rather than withdrawn (tolcapone, Tasmar). This left 31 drugs, all of which were analyzed. Of these, two do not cleanly satisfy the strict criterion of withdrawal from the US market by the FDA for safety reasons: FDA’s 2011 Federal Register determination categorized levacetylmethadol (LAAM/ORLAAM) as not withdrawn for safety or effectiveness, and nefazodone was never formally withdrawn by the FDA (the branded product was discontinued commercially while generic nefazodone remains marketed). These two are held out as a border-case sensitivity set (Appendix C), leaving a strict primary cohort of 29 drugs. Four primary-cohort drugs (alosetron, natalizumab, gemtuzumab ozogamicin, tegaserod) were later reintroduced under REMS or restricted-access programs and carry a withdrawn-and-reintroduced tag. The full analyzed set, with per-drug withdrawal year, tier assignment, and reintroduction status, is given as Table 1 (Appendix A). The strict 29-drug primary cohort spans seventeen therapeutic classes; the full 31-drug analyzed set spans eighteen, the additional class, opioid-maintenance agents, contributed solely by the LAAM/ORLAAM border case.

Derivation step	Count
Onakpoya candidate US withdrawals, 1990 to 2024	39
Removed: never US-approved	6
Removed: commercial or non-safety withdrawal	1
Removed: restricted, not withdrawn	1
Analyzed set (all analyzed)	31
Held out as border-case sensitivity set (LAAM/ORLAAM; nefazodone)	2
Strict primary cohort	29

3.2 Pivotal-paper identification

Across the 29-drug primary cohort we identified 63 indexed pivotal efficacy papers, reported by tier in Appendix B. The full analyzed set of 31 drugs, including the two border cases, comprises 66 indexed pivotal efficacy papers (Appendix C). Individual citations for every pivotal paper and correction notice across all 31 drugs are listed in the bibliography of Additional file 1. Papers were assigned to the three identification tiers of Section 2.3. Tier 1 (FDA-confirmed) predominates for post-2000 approvals, where FDA primary documents are recoverable through the online archive. Tier 2 (retrospective-inferred) is required for several pre-1995 approvals, where FDA archives are incomplete and pivotal status was established from authoritative post-approval reviews. Per-tier drug and paper counts are reported in Appendix B. Registration-era records that fall outside PubMed indexing (gray-literature and pre-MEDLINE) contribute no machine-discoverable paper and are not counted in the indexed denominator; three drugs (bromfenac, gatifloxacin, pemoline) contribute zero indexed pivotal papers on this basis. Identification tier and bibliographic indexing status are independent dimensions: the tier records how the pivotal-trial evidence for a drug was located (FDA primary documents for Tier 1, authoritative retrospective review for Tier 2), whereas indexing status records whether the resulting publication has a live PubMed or Crossref record able to carry a correction. A drug can be Tier 1 yet contribute no indexed paper, as gatifloxacin illustrates: its pivotal trials are FDA-confirmed

but were published only in gray literature and never entered the indexed record. Abstract-only and supplement-only publications of this kind lack a Crossref DOI deposit and a MEDLINE record, so they cannot be surfaced, or annotated, by the automated evidence-synthesis systems that read the indexed literature.

Two categories emerged that the initial A/B/C typology did not anticipate, and audit sharpened the line between them. First, one drug had pivotal trials that were never substantially published as peer-reviewed papers. Bromfenac (Duract) is that instance: Goldkind and Laine's 2006 systematic review confirms that very limited pre-approval data were published, and we code bromfenac as having zero identifiable published pivotal papers. This is itself a structural finding: where the pivotal evidence never entered the peer-reviewed record, no paper-level correction mechanism can operate on it, because there is no pivotal paper to correct. Second, a distinct intermediate category, identified when the aggregate case-file entries were resolved at audit, comprises drugs whose pivotal trials were published but outside the PubMed-indexed record, so that standard bibliographic correction infrastructure does not reach them: gatifloxacin, whose US community-acquired-pneumonia registration trials appeared in a journal supplement and in conference abstracts rather than indexed primary reports, and pemoline, whose registration trials appeared in journals that predate MEDLINE indexing. Technetium fanolesomab (NeutroSpec) belongs to neither category: its two pivotal diagnostic-accuracy trials are indexed in PubMed under the drug's earlier trade name LeuTech, and it is reclassified accordingly.

3.3 Correction-status endpoints

Correction status was assessed against the three-level endpoint hierarchy of Section 2.5.

On the primary endpoint, a safety-related paper-level action, one pivotal paper in the cohort qualified: the Expression of Concern issued for VIGOR (Bombardier et al., NEJM 2000) [8] in December 2005 [9] and reaffirmed in March 2006 [10]. No pivotal paper was retracted. The cohort-wide primary-endpoint rate is one in 63, approximately 1.6 percent. The audit (Section 2.7) reproduced this result without discordance on the primary endpoint, and the aggregate case-file entries have since been resolved to specific publications and correction-checked (see below).

On secondary endpoint A, any formal paper-level action regardless of content, one further action attaches to a pivotal efficacy paper: a 1992 erratum to the Weintraub weight-control trial [11], verified at audit on the live record as a routine non-safety correction predating the withdrawal signal. It meets secondary endpoint A only. Two further errata sit on non-pivotal cohort papers, a 1997 correction to the Connolly fenfluramine-phentermine report [13] and an erratum to the BART aprotinin trial [14], and neither is a pivotal efficacy paper. Across the 63 pivotal papers, therefore, two carry a formal action (VIGOR and Weintraub), of which one is safety-related, and 61 carry none. The four-source check was run on every pivotal paper, so the secondary-endpoint-A count is a complete census of formal actions on the 63 pivotal papers, not a lower bound.

On secondary endpoint B, a machine-discoverable withdrawal-status annotation, no pivotal paper in the cohort carried any such notice: a reader retrieving these papers through

standard bibliographic infrastructure finds no indication on the paper record that the drug was withdrawn, suspended, or reintroduced.

Endpoint level	Count in cohort
Primary: safety-related paper-level action	1 (VIGOR expression of concern), audit-confirmed
of which retractions	0
Secondary A: any formal action on a pivotal paper (census)	2 (VIGOR expression of concern; Weintraub 1992 erratum, both non-safety-related except VIGOR)
Secondary B: withdrawal-status annotation on the paper record	0
No formal action of any level	61 of the 63 pivotal papers

Audit result. The validation proceeded in a fixed order. Initial case extraction produced two kinds of record: specific paper-level entries, and aggregate trial entries not yet resolved to a citation (for example, “multiple Phase III trials”). At audit, 26 of the 36 selected records were already specific publications, and 10 were unresolved aggregates. The 10 aggregates were then resolved to specific publications, against the corresponding FDA medical review where one was available online (gatifloxacin, technetium fanolesomab) and otherwise against the published registration-era literature (terfenadine, trovafloxacin, aprotinin, pergolide, pemoline, cisapride, temafloxacin, astemizole, encainide). The paper-level denominator was then frozen at 63 indexed pivotal papers, and the prespecified four-source search was applied to every one of them. Validation was assessed on two distinct questions. First, whether each included publication was genuinely a pivotal efficacy paper: confirmed against the FDA record for the audited set. Second, whether its correction status was coded correctly: independently re-checked against the four sources without reference to the case file. On the second question, primary-endpoint classification was concordant in every case, primary correction-status error was zero, and the re-audit trigger was not met. No resolved aggregate publication carried a safety-related action, so resolving the aggregates sharpened the denominator without changing the numerator. The one non-safety action on a pivotal paper, the Weintraub 1992 erratum, was verified at audit on the live record and affects secondary endpoint A only, not the primary endpoint.

3.4 Pattern distribution

Classifying each primary-cohort drug by its predominant withdrawal reason under the Section 2.6 typology gives the distribution below; the two border-case drugs are classified separately in Appendix C. Alatrofloxacin shares its pivotal literature with trovafloxacin and is counted once, within Pattern C.

Withdrawal pattern	Drugs
A: pivotal paper directly contaminated	1 (rofecoxib / VIGOR; A with a C element)
B1: surrogate-to-clinical-outcome	3 (encainide / CAST; cerivastatin, B/C; gemtuzumab

Withdrawal pattern	Drugs
failure	ozogamicin / Mylotarg, accelerated-approval surrogate)
B2: pivotal clinical-outcome non-replication	1 (drotrecogin alfa / PROWESS)
C: harm in an unmeasured domain	22
Pivotal literature never published in the peer-reviewed record	1 (bromfenac)
Distinctive sub-case: withdrawal signal already in the pivotal paper	1 (natalizumab)
Total (primary cohort)	29

As an exploratory descriptive layer, not part of the prespecified analysis, ten recurring withdrawal-mechanism sub-patterns were noted across the cohort: safer-alternative withdrawals (terfenadine, astemizole, mibefradil); drug-interaction withdrawals (mibefradil, cisapride, terfenadine); idiosyncratic organ toxicity (trovafloxacin, alatrofloxacin, troglitazone, bromfenac); post-approval mandated safety-trial withdrawals (sibutramine, tegaserod); long-latency risk (efalizumab); known-at-approval harm (bepridil, pemoline, alosetron); non-replicated pivotal finding (drotrecogin alfa / PROWESS); correct effect with an incorrect clinical rationale (encainide / CAST); class-effect anticipated from a prior withdrawal (pergolide, following the fen-phen precedent); and accelerated-approval confirmatory failure (gemtuzumab ozogamicin / Mylotarg).

3.5 Key illustrative cases

Five cases illustrate the range of ways a drug's evidence base can change without any change to its pivotal literature. Full per-drug case files for the complete cohort are provided in Supplementary Material S1.

VIGOR / rofecoxib: the sole safety-related formal action. The VIGOR trial (Bombardier et al., NEJM 2000) supported the rheumatoid-arthritis expansion of rofecoxib. Its Expression of Concern (NEJM 2005; reaffirmed 2006) was issued more than five years after publication and fourteen months after the drug was withdrawn, prompted by litigation evidence that three myocardial infarctions in the rofecoxib arm had been deleted from the manuscript before submission. The authors declined to submit a correction. This is the only case in the cohort with a documented Pattern A element, and the only one that produced a formal safety-related action.

PROWESS / drotrecogin alfa: efficacy non-replication, uncorrected. The pivotal PROWESS trial (Bernard et al., NEJM 2001) [15] reported a 6.1 percent absolute mortality reduction that supported approval of drotrecogin alfa for severe sepsis. A decade later PROWESS-SHOCK [17] found no mortality benefit, and the drug was withdrawn. The scientific interpretation of the discrepancy remains contested, and our study does not adjudicate it. What is robust across interpretations is narrower: the central efficacy claim was not replicated, the drug was withdrawn substantially because of that non-replication,

and PROWESS itself carries no correction, expression of concern, or annotation linking its record to the withdrawal.

Encainide / CAST: surrogate-endpoint invalidation, uncorrected. The pre-approval pivotal trials of encainide correctly showed suppression of ventricular ectopy, and the drug was approved in 1986 for life-threatening ventricular arrhythmias. In 1989 the Cardiac Arrhythmia Suppression Trial (CAST-I; Echt et al., NEJM 1991) [16] showed that suppressing premature ventricular contractions after myocardial infarction with encainide, flecainide, or moricizine increased mortality, and encainide was withdrawn in 1991. The pivotal papers were correct about the drug's biological effect; the surrogate-to-benefit inference on which approval depended was overturned. That lesson lives in the CAST papers and in review articles, not in any annotation of the original encainide pivotal papers, which stand uncorrected today.

Bromfenac / Duract: pivotal evidence never published. Bromfenac was approved in 1997 for short-term acute pain and withdrawn in 1998 for hepatotoxicity. Almost none of its pre-approval evidence was published as peer-reviewed pivotal trial papers; the hepatotoxicity signal was documented in the FDA medical review but never entered the peer-reviewed literature. With no pivotal paper in the record, no correction mechanism can operate. Bromfenac anchors the Tier 3 category described in Section 3.2.

Natalizumab / Tysabri: the withdrawal signal is in the paper, without annotation. Natalizumab received accelerated approval in 2004, was suspended in February 2005 for progressive multifocal leukoencephalopathy, and was reintroduced under REMS in 2006. Both pivotal Phase III papers [18,19] were published in NEJM in 2006, after the suspension, and acknowledged the PML cases directly. No formal correction, retraction, or expression of concern was attached to either paper. Here the absence of correction is arguably appropriate, since the papers were accurate and remain the evidence base for the currently approved restricted product; the case nonetheless shows that even a paper whose subject was suspended and reintroduced carries no paper-level annotation of that history.

4. Discussion

4.1 Principal finding

Across the primary cohort of twenty-nine FDA safety withdrawals from 1990 to 2024, the pivotal trial record carried no machine-discoverable signal of the withdrawal. One paper of 63 carried a safety-related action, an Expression of Concern on VIGOR; none was retracted; and none carried a withdrawal-status annotation. This is a finding about the record, not about the papers, which are in almost every case accurate accounts of what their trials measured. It is that the record of an accurate trial, once its drug is withdrawn, was not updated in any case in this cohort to state the subsequent withdrawal. Drawn from drugs spanning three decades, seventeen therapeutic classes, and five regulatory pathways, it shows that the peer-reviewed medical literature does not routinely carry a durable, machine-discoverable, paper-level signal of post-approval drug withdrawal.

The consistency of the finding across widely different regulatory pathways is particularly notable. Whether a drug was approved via standard NDA (Vioxx, cerivastatin, encainide),

fast-track (Vioxx), accelerated approval (Mylotarg, natalizumab), BLA (efalizumab, natalizumab, Mylotarg, Xigris, NeutroSpec), or pre-1995 processes (aprotinin, terfenadine, bepridil, pemoline), the near-absence of safety-related formal action and the complete absence of withdrawal-status annotation were observed across regulatory pathways. This suggests the observation is a property of the general peer-review-plus-correction infrastructure, not of any particular regulatory pathway.

Three cases in the cohort deserve emphasis because they demonstrate that even where the pivotal claim or the clinical rationale connecting it to benefit did not survive re-examination, formal correction does not attach.

The Xigris/PROWESS case. The pivotal PROWESS trial (Bernard et al., NEJM 2001) reported a 6.1% absolute mortality reduction that led to approval of drotrecogin alfa for severe sepsis. A decade later, PROWESS-SHOCK found no mortality benefit, and the drug was withdrawn. The scientific interpretation of the discrepancy is contested: some investigators argue the original effect was real in a differently-treated patient population under earlier standards of sepsis care that PROWESS-SHOCK's updated protocol did not preserve; others attribute the original result to interim-analysis stopping, protocol changes during recruitment, or subgroup selection. Our study does not adjudicate this dispute. What we observe is narrower and more robust than either interpretation: whichever account of the discrepancy is correct, PROWESS's central efficacy claim has not been replicated, the drug was withdrawn substantially because of that non-replication, and PROWESS itself carries no formal correction, expression of concern, or annotation of any kind linking its record to the subsequent regulatory action.

The encainide/CAST case. The pre-approval Phase III pivotal trials of encainide correctly demonstrated ventricular ectopy suppression, and the drug was approved in 1986 for the treatment of life-threatening ventricular arrhythmias. In 1989, the post-approval Cardiac Arrhythmia Suppression Trial (CAST-I; Echt et al., NEJM 1991) [16] demonstrated that suppressing PVCs with encainide, flecainide, or moricizine after myocardial infarction increased mortality. Encainide was withdrawn in December 1991. The pre-approval pivotal literature was correct about the drug's biological effect; the surrogate-to-clinical-benefit inference the approval depended on was overturned by CAST-I. This is a widely taught example of surrogate-endpoint failure in cardiology. That curricular status exists in the CAST papers themselves and in review articles about them, not in any annotation of the original encainide pivotal papers. Those pivotal papers stand uncorrected today, thirty-five years after CAST-I. If PROWESS is the strongest efficacy-non-replication case, encainide is the strongest surrogate-endpoint-invalidating case, and both produced zero formal correction actions on their pivotal literatures.

The ten exploratory withdrawal-mechanism sub-patterns noted across the cohort ("safer alternative," drug-interaction, idiosyncratic organ toxicity, post-approval mandated safety trial, long-latency risk, "known-at-approval" harm, non-replicated pivotal finding, correct-effect-incorrect-clinical-rationale, class-effect anticipated from prior withdrawal, and accelerated-approval confirmatory failure) each represent a distinct mechanistic route to withdrawal, ranging from harm in a domain the pivotal trial was not designed to measure to failure of the pivotal efficacy claim or its clinical rationale. This mechanistic diversity

across cases that share the same primary-endpoint outcome (no safety-related paper-level action) supports the structural interpretation: the correction infrastructure is scoped to demonstrable paper-level error, while many withdrawals arise from later evidence that changes the clinical interpretation of an otherwise accurate pivotal publication.

4.2 The correction mechanism is structurally narrower than the withdrawal mechanism

We do not interpret this finding as evidence that journals or peer reviewers have failed. Journals are not designed to function as ongoing post-marketing surveillance infrastructure, and peer reviewers cannot be held accountable for information that emerged after publication. The finding is more precisely stated as a structural observation about the mismatch between two institutional mechanisms operating on different evidence tiers.

Onakpoya et al. (2016) found that 71% of drug withdrawal decisions in their 462-product registry were based on Level 4 evidence (case reports and case series), and only 1.3% on Level 1 evidence (systematic reviews). This is the withdrawal evidence base, not the pivotal-paper evidence base, but it establishes a relevant asymmetry. Withdrawal is triggered by low-tier evidence (case reports); formal literature correction requires paper-level demonstrated wrongness in the specific claims a pivotal paper made. The two mechanisms do not share evidence types. Withdrawals do not by their nature produce the material that would trigger corrections.

4.3 The pattern typology explains the rate

Our pattern typology helps explain why the correction rate is low. In Pattern A cases, where a pivotal paper directly misrepresented data relevant to the safety signal, formal correction is warranted and, in principle, achievable. In Pattern B1 cases (surrogate-to-clinical-outcome failure), the pivotal paper was not wrong about the surrogate it measured, and there is no clean basis for correcting it. In Pattern B2 cases (clinical-outcome non-replication, as in PROWESS), the pivotal paper's clinical result was not replicated in a later trial, but non-replication of a result is not, under prevailing norms, treated as a paper-level error warranting correction of the original publication. In Pattern C cases (harm in unmeasured domain), the paper is entirely correct about what it claimed; there is nothing to correct.

Across the primary cohort, only one of the 29 withdrawals exhibited a Pattern A element, and that case produced the sole safety-related formal action. Because most withdrawals fall in Patterns B and C, the low correction rate is not a failure of the correction infrastructure but a structural consequence of what the withdrawal reasons are.

4.4 The gap this leaves

A clinician, patient advocate, or researcher accessing the peer-reviewed literature today for evidence about, for example, cerivastatin 0.8 mg efficacy finds Insull et al. (2000) [20] presenting the pivotal evidence for the 0.8 mg dose; within roughly a year of the drug's launch the FDA had received reports of 31 US deaths from rhabdomyolysis linked to cerivastatin, and the drug was withdrawn [21]. Nothing on the paper's PubMed page

indicates the drug was withdrawn. Its bibliographic record contains no structured indication distinguishing its withdrawn regulatory status from that of papers supporting currently marketed products. Similar retrievals for the pivotal papers of all 29 drugs in the primary cohort, including VIGOR, return no machine-discoverable annotation of the drug's withdrawal or subsequent regulatory history.

This is the paper's core descriptive finding. Regulatory withdrawal notices, guideline updates, prescribing databases, and pharmacovigilance systems each carry a signal of current drug status. The peer-reviewed literature does not, and under the current division of institutional labor is not expected to. Journals are not designed to function as ongoing post-marketing surveillance infrastructure, and the correction mechanism they operate is scoped to paper-level demonstrable error. Most drug withdrawals do not produce demonstrable paper-level error in the pivotal literature, and the correction mechanism therefore does not attach.

A structural observation follows. The institutional division of labor leaves no single actor responsible for annotating the peer-reviewed record when the regulatory status of a drug changes: journals are not surveillance systems, regulators treat the withdrawal notice itself as the primary status signal, guideline committees address prescribing rather than the underlying literature, and pharmacovigilance databases record adverse events rather than paper-level annotations. Each scope is defensible on its own terms. We found no routinely implemented mechanism that placed withdrawal-status metadata on the records examined, and responsibility for placing one appears fragmented across these actors rather than assigned to any of them. Whether that gap ought to be closed, and by whom, is a normative question beyond the scope of this descriptive study.

The gap is nonetheless technically closable. A machine-discoverable link between the regulatory record and the bibliographic record, for example a structured regulatory-status field populated from the FDA withdrawn-drug registry and surfaced alongside the PubMed or Crossref metadata of a pivotal paper, would give downstream systems the signal they currently lack. Whether such a mechanism should be built, and which body should own it, is an infrastructural and normative question this study does not attempt to answer.

One objection to this framing should be stated fairly, because it is a legitimate normative position rather than an error. On this view, a drug's regulatory status belongs in regulatory databases, prescribing systems, and clinical guidelines, and not in the metadata of a journal article; the silence of the pivotal record is then correct by design, a sensible division of labor rather than a defect. Our study does not disprove this position, and does not attempt to. It is worth separating two responsibilities that the objection can run together. The first is responsibility for the scientific content at the time of publication. Where the later safety information was genuinely unknowable, that responsibility was discharged: the paper accurately reported the evidence then available, and neither authors nor editors owed a correction. The second is responsibility for maintaining a link between an evolving regulatory fact and the evidence record that a reader retrieves. This is a post-publication infrastructure question, it need not fall on the original authors or journal, and it need not involve altering the paper at all. The distinction is familiar from record-keeping: later context need not alter the original work, but it may still be added to the surrounding

catalogue record. Whether such a status link should be built for the trial literature, and whether it should be owned by a regulator, a bibliographic service, a publisher, or no institution at all, is the governance question we leave open. What our result establishes, independent of how that question is answered, is narrower: today the two records are not linked, so a reader who retrieves the pivotal paper through standard bibliographic infrastructure receives no machine-discoverable signal of the drug's withdrawal.

Framed at the level of records rather than papers, the finding is a failure of alignment between two authoritative systems of record. The regulatory record states that the drug was withdrawn for safety reasons. The bibliographic record of the pivotal evidence states nothing of the kind, and the two are not linked. This discordance is not unique to withdrawal status. Structured links between the trial literature and its own registrations are frequently missing: in one analysis a majority of completed trials had no structured trial-to-article link, and identifiers present in an article's full text were often absent from its indexed metadata [22]. The absence we report is therefore one instance of a broader and documented weakness, that the bibliographic record does not reliably carry structured links to the regulatory and registration facts that determine how its contents should be read.

The consequence of a missing status signal is not merely hypothetical, and an adjacent literature bounds its likely size. Studies of retracted papers, which do carry a formal notice, find that they continue to be cited, and that most such citations are positive, treating the work as valid despite the notice [23,24]. In one biomedical case a fraudulent clinical trial report was still cited positively to support the same intervention eleven years after its retraction, with the retraction rarely mentioned in the citing text [25]. If the strongest signal the record can carry, a formal retraction, propagates this poorly through the citation network, it is reasonable to expect that a class of papers carrying no status signal at all may propagate at least as poorly, because there is nothing for even a diligent citation-checker to detect. The same failure appears in the clinical layer: among medicines withdrawn from one national market, most remained recommended by at least one clinical guideline at the time of withdrawal, and more did so afterward [26]. We did not measure citation or guideline persistence for the present cohort, and we draw no quantitative estimate from these adjacent findings. They establish only that the channel through which a missing status signal would cause harm, continued use of superseded evidence as if current, is real and measurable, and they motivate the citation-persistence study we set out in Section 4.6.

4.5 Limitations

- **Cohort size.** The strict primary cohort is the complete set of 29 US safety withdrawals meeting the inclusion criteria for 1990 to 2024 after two border cases (LAAM/ORLAAM and nefazodone) are held out for sensitivity analysis; the full analyzed set contains 31 drugs. This is a full enumeration of the eligible population rather than a sample, but it is modest in absolute terms, and because the primary endpoint occurred once, the cohort-wide rate rests on a single event. The primary indexed denominator is exact (63 pivotal papers in the strict cohort, enumerated in Appendix B); broader paper universes that would additionally count non-indexed, gray-literature, or held post-approval records remain analytically distinct but do not bear on the primary-endpoint rate. Because the cohort is a complete enumeration of

the drugs identified through the prespecified source framework and eligibility criteria, the single event is best understood as an accurate census of a rare phenomenon rather than an underpowered estimate of a common one: the quantity of interest is the near-total absence of correction across a fully enumerated population, and a larger sample could not make that enumeration more complete.

- **Tier 2 pivotal-paper identification.** Pivotal-paper identification for pre-2000 drugs relies more heavily on Tier 2 (retrospective-review-inferred) papers than post-2000 drugs, because FDA online archives are incomplete for older approvals. The pre-specified sensitivity analysis stratified by approval era shows the finding is robust: one safety-related action among 44 papers supporting pre-2000 approvals, and none among 19 papers supporting approvals from 2000 onward (Appendix B). The pattern is therefore not an artifact of the greater Tier 2 reliance among older approvals.
- **Endpoint scope.** We deliberately scope the endpoint to formal paper-level correction actions discoverable through the prespecified PubMed, Crossref, Retraction Watch, and Google Scholar searches. Correction actions confined to journal websites without cross-database registration would not be captured; this could result in modest undercounting. However, from a downstream-user perspective (clinician, systematic reviewer, guideline developer, automated evidence synthesis tool), an unregistered journal-website correction functions identically to no correction for practical purposes. Editorial commentary, letters to the editor, and post-approval retrospective discussion are all forms of literature response, but are not the machine-discoverable paper-level annotation our endpoint measures. The prespecified four-source search was run on every one of the 63 pivotal papers, so secondary endpoint A (any formal paper-level action, including non-safety errata) is a complete census within that search protocol. A formal correction confined to a publisher website and not represented in any of the four prespecified sources could still be missed. Because such errata bear no relation to the regulatory status or safety profile of a drug, an exhaustive census of them would not alter the primary endpoint or the structural conclusions.
- **LLM-assisted workflow.** LLM assistance introduces a systematic risk of subtle misclassification. The workflow discipline (“LLM extracts, human adjudicates”), the completed 36-pair audit, and the pre-specified re-audit trigger at >5% primary-endpoint error were the mitigation strategy. The audit found zero primary-endpoint discordance, so the re-audit trigger was not met.
- **Withdrawn-and-reintroduced tag.** Four drugs (alosetron, natalizumab, gemtuzumab ozogamicin, tegaserod) were subsequently reintroduced under REMS or restricted access. Their inclusion in the primary cohort with a distinguishing tag permits stratified analysis but complicates the “post-withdrawal” endpoint definition, since these drugs are technically re-approved. We measure correction status as of a fixed reference date (analysis date) rather than “after first withdrawal event” to standardize the endpoint across reintroduced and fully-withdrawn drugs.

- **Time-varying correction infrastructure.** PubMed’s linked-articles system, Crossref DOI infrastructure, and Retraction Watch have all matured over the study period. Older withdrawals may have received corrections that were not machine-indexed at the time and are no longer discoverable now. The sensitivity analysis by decade addresses this partially; a citation-network analysis of the cohort (proposed in Section 4.6) would address it more fully.
- **Scope of the claim.** This study measures the machine-discoverable status of the pivotal record. It does not measure downstream effect, and four boundaries follow. It does not establish that any reader, clinician, reviewer, or automated tool was in fact misled; the retrieval examples show what a reader would encounter, not a documented harm. It does not measure whether the pivotal papers continue to be cited or recommended as valid, which is the separate study proposed in Section 4.6. It does not establish that annotation is the correct or feasible remedy, or that its absence is any actor’s dereliction; the fragmented-responsibility reading is an inference from the observed silence, not a governance finding. And the enumeration is complete only with respect to the prespecified source framework and the four bibliographic services searched, not with respect to an unknowable absolute set of withdrawals or channels.

4.6 Directions for future work

Building on this finding, several follow-up questions become natural:

- **Citation persistence.** Do the uncorrected pivotal papers of withdrawn drugs continue to be cited as valid evidence in subsequent reviews and guidelines? A citation-network analysis of the same cohort could quantify how the medical literature *does* respond to withdrawal, absent formal correction.
- **Cross-mechanism comparison.** How do medical guideline updates and prescribing-database annotations propagate the withdrawal signal in the period before, during, and after any formal literature correction? This would clarify the *effective* correction channel even where formal action is absent.
- **International comparison.** Do medical literatures in journals with different editorial policies (national medical journals, open-review journals, journals in different regulatory environments) show different formal-correction rates for the same drug’s pivotal papers?

5. Conclusions

Formal correction of pivotal trial publications following FDA safety drug withdrawal is uncommon: across the 29-drug primary cohort, one of 63 indexed pivotal papers carried a safety-related paper-level action, a rate of approximately 1.6 percent, with zero retractions and no pivotal paper carrying a machine-discoverable withdrawal-status annotation. The peer-reviewed record carried no structured withdrawal-status annotation on any pivotal paper, and only one safety-related formal action among 63; it does not routinely carry a durable, machine-discoverable signal of subsequent withdrawal. This is not an editorial

failure. The mechanism is a scope mismatch: formal correction is built for demonstrated paper-level error, and most withdrawals follow harms in domains the pivotal trials were not designed to measure. But the deeper finding is structural: we found no routinely implemented mechanism that marks the trial record when a drug's status changes, so the record is not marked, and that gap is now read by automated systems that cannot see what the record does not state. Even where the approval-relevant interpretation of the pivotal evidence was later overturned, formal correction did not, in practice, follow.

Appendices

Appendix A. Full analyzed cohort (Table 1)

Table 1. The full analyzed set of 31 drugs (strict 29-drug primary cohort plus the two border-case drugs, LAAM/ORLAAM and nefazodone). Withdrawal pattern and identification tier are as defined in Methods.

Drug (generic)	Trade name	First US approval	Withdrawal year	Withdrawal pattern	Tier	Reintroduced
Rofecoxib	Vioxx	1999	2004	A (with C element)	Tier 1	No
Cerivastatin	Baycol	1997	2001	B1 (B/C hybrid)	Tier 1	No
Troglitazone	Rezulin	1997	2000	C	Tier 1	No
Sibutramine	Meridia	1997	2010	C	Tier 1	No
Terfenadine	Seldane	1985	1998	C	Tier 2	No
Valdecoxib	Bextra	2001	2005	C	Tier 1	No
Bromfenac	Duract	1997	1998	C	Tier 3	No
Trovafloracin	Trovan	1997	2001 (restricted)	C	Tier 1	No
Grepafloxacin	Raxar	1997	1999	C	Tier 1	No
Temafloxacin	Omniflox	1992	1992	C	Tier 2	No
Mibefradil	Posicor	1997	1998	C	Tier 1	No
Cisapride	Propulsid	1993	2000	C	Tier 1	No
Tegaserod	Zelnorm	2002	2007	C	Tier 1	Yes (2019)
Astemizole	Hismanal	1988	1999	C	Tier 2	No
Sparfloxacin	Zagam	1996	2001	C	Tier 1	No
Efalizumab	Raptiva	2003	2009	C	Tier 1	No
Natalizumab	Tysabri	2004	2005 (suspended)	Distinctive (signal in pivotal)	Tier 1	Yes (2006)

Drug (generic)	Trade name	First US approval	Withdrawal year	Withdrawal pattern	Tier	Reintroduced
Gemtuzumab ozogamicin	Mylotarg	2000	2010	B1 (accelerated-approval paper)	Tier 1	Yes (2017)
Drotrecogin alfa	Xigris	2001	2011	B2 (non-replication)	Tier 1	No
Aprotinin	Trasylol	1993	2008	C	Tier 2	No
Bepridil	Vascor	1990	2004	C	Tier 1	No
Encainide	Enkaid	1986	1991	B1	Tier 2	No
Pergolide	Permax	1988	2007	C	Tier 1	No
Rapacuronium	Raplon	1999	2001	C	Tier 1	No
Fenfluramine	Pondimin	1973	1997	C	Tier 1	No
Gatifloxacin	Tequin	1999	2006	C	Tier 1 (pivotal pubs gray-literature)	No
Pemoline	Cylert	1975	2005	C	Tier 2 (pivotal pubs pre-MEDLINE)	No
Technetium fanolesomab	NeutroSpec	2004	2005 (suspended)	C	Tier 1	No
Alosetron	Lotronex	2000	2000	C	Tier 1	Yes (2002)
Levomethadyl acetate (LAAM)	ORLAAM	1993	2003	C (border case)	Tier 1	No
Nefazodone	Serzone	1994	2004	C (border case)	Tier 1	No

Appendix B. Sensitivity analyses

The audit described in Section 2.7 is closed and the numerator is fixed: across the cohort there are zero retractions and a single safety-related formal action (the VIGOR Expression of Concern). The indexed pivotal-paper denominator is 63 for the strict 29-drug primary

cohort (Tier 1 plus Tier 2 pooled) and 66 for the full 31-drug set, and the rate is approximately 1.6 percent. Drug-level and paper-level counts in the tables below are exact, enumerated from the reference list. Three drugs contribute no PubMed-indexed pivotal paper to the paper-level denominator: bromfenac (pivotal trials never published) and gatifloxacin and pemoline (pivotal trials published, but in a journal supplement, conference abstracts, or pre-MEDLINE journals not reached by bibliographic indexing); they appear in the drug counts but not the indexed-paper denominator.

Table B1. Safety-related formal-action rate by pivotal-paper identification tier.

Analysis set	Drugs	Pivotal papers	Safety-related formal actions	Retractions	Rate
Tier 1 papers only	22	51	1 (VIGOR)	0	1/51 = 2.0%
Tier 1 and Tier 2 pooled	28	63	1 (VIGOR)	0	1/63 = 1.6%

Table B2. Safety-related formal-action rate by decade of first US approval.

Analysis set	Drugs	Pivotal papers	Safety-related formal actions	Retractions	Rate
Pre-2000 approvals	21	44	1 (VIGOR)	0	1/44 = 2.3%
2000-onward approvals	8	19	0	0	0%

The sole safety-related formal action attaches to rofecoxib, approved in May 1999, which falls in the pre-2000 stratum under stratification by decade of first US approval. Twenty-one of the twenty-nine drugs were approved before 2000 and eight from 2000 onward, so the single correction sits in the larger, older, more Tier-2-reliant stratum, and the newer, better-indexed stratum of eight drugs carries none. This cuts in favor of the finding rather than against it: an artifactually low rate driven by weak identification of older drugs would concentrate any corrections in the newer, better-indexed stratum, not the older one. Stratification is by approval decade, not withdrawal decade; several drugs withdrawn after 2000, including rofecoxib, cisapride, and gatifloxacin, were approved before it.

Table B3. Safety-related formal-action rate by withdrawal and reintroduction status.

Analysis set	Drugs	Pivotal papers	Safety-related formal actions	Retractions	Rate
Fully withdrawn	25	55	1 (VIGOR)	0	1/55 = 1.8%
Withdrawn and reintroduced	4	8	0	0	0%

Appendix C. Border-case sensitivity analysis

The strict primary cohort of 29 drugs comprises drugs withdrawn from the US market by the FDA for safety reasons, 1990 to 2024. Two drugs in the analyzed set of 31 satisfy the Onakpoya inclusion criterion and were fully analyzed but do not cleanly meet that strict definition. They are held out here as a sensitivity set.

LAAM / levacetylmethadol (ORLAAM). FDA's 2011 Federal Register determination categorized ORLAAM as not withdrawn from sale for reasons of safety or effectiveness. The withdrawal-relevant safety signal (QT prolongation and torsades de pointes) drove regulatory action in the European Union and the drug's commercial discontinuation, but the formal US determination was non-safety. Withdrawal pattern: C (drug-interaction / QT prolongation). No pivotal paper carries any formal action at any endpoint level.

Nefazodone (Serzone). The FDA never formally withdrew nefazodone. Bristol-Myers Squibb withdrew the branded product in 2004 citing declining sales after other jurisdictions had withdrawn it for hepatotoxicity, and generic nefazodone remains marketed in the US. The case is closer to a commercial withdrawal accompanied by a substantial safety signal than to a pure FDA safety withdrawal. Withdrawal pattern: C (idiosyncratic hepatotoxicity). No pivotal paper carries any formal action at any endpoint level.

Effect on the result. Both border-case drugs are Pattern C, and neither carries a correction, expression of concern, retraction, or withdrawal-status annotation on any pivotal paper. Including them expands the cohort from 29 to 31 drugs and the indexed pivotal-paper count from 63 to 66, while the count of primary-endpoint actions remains one (the VIGOR expression of concern). The primary-endpoint rate is therefore approximately 1.5 to 1.6 percent whether the two border cases are included (1/66) or excluded (1/63). The principal finding is not sensitive to their classification.

Declarations

Ethics approval and consent to participate: Not applicable. This study used only publicly available published literature and regulatory documents and involved no human participants, human data, or human tissue.

Consent for publication: Not applicable.

Availability of data and materials: All data generated or analysed during this study are included in this published article and its supplementary information files. The full per-drug case files and audit trail are provided as Additional file 1 (Supplementary Material S1), and the cohort, pivotal-paper identifications, tier assignments, and author-verified correction-status determinations with their sources are enumerated in Additional file 2 (denominator and tier-resolution workbook).

Competing interests: The author declares that they have no competing interests.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions: PK conceived and designed the study, assembled the cohort, performed the analysis, verified all findings and citations against primary sources, and prepared the manuscript. The author read and approved the final manuscript.

Acknowledgements: Not applicable.

Use of large language models: A large language model was used as an analysis-support and drafting aid, under the audit protocol described in Methods (Section 2.7). All correction-status determinations and all citations were verified by the author against primary records (live PubMed, Crossref, Retraction Watch, and FDA documents); no correction status or citation was recorded on model assertion. The large language model does not meet authorship criteria and is not listed as an author.

References

The list below contains the works cited in the main text. The complete verified bibliography of pivotal-paper and correction-notice citations for all 31 analyzed drugs is provided in Additional file 1.

1. Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med.* 2016;14:10.
2. Karha J, Topol EJ. The sad story of Vioxx, and what we should learn from it. *Cleve Clin J Med.* 2004;71(12):933-939.
3. Sorkin EM, Heel RC. Terfenadine. A review of its pharmacodynamic properties and therapeutic efficacy. *Drugs.* 1985;29(1):34-56.
4. Goldkind L, Laine L. A systematic review of NSAIDs withdrawn from the market because of hepatotoxicity: lessons learned from the bromfenac experience. *Pharmacoepidemiol Drug Saf.* 2006;15(4):213-220.
5. Levy JH, Pifarre R, Schaff HV, et al. A multicenter, double-blind, placebo-controlled trial of aprotinin for reducing blood loss and the requirement for donor-blood transfusion in patients undergoing repeat coronary artery bypass grafting. *Circulation.* 1995;92(8):2236-2244.
6. Olanow CW, Fahn S, Muenter M, et al. A multicenter double-blind placebo-controlled trial of pergolide as an adjunct to Sinemet in Parkinson's disease. *Mov Disord.* 1994;9(1):40-47.
7. Rodriguez-Monguio R, Seoane-Vazquez E, Powers JH 3rd. A comparative assessment of approvals and discontinuations of systemic antibiotics and other therapeutic areas. *Healthcare (Basel).* 2023;11(12):1759.
8. Bombardier C, Laine L, Reicin A, et al. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med.* 2000;343(21):1520-1528.
9. Curfman GD, Morrissey S, Drazen JM. Expression of concern: Bombardier et al., Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med.* 2005;353(26):2813-2814.

10. Curfman GD, Morrissey S, Drazen JM. Expression of concern reaffirmed. *N Engl J Med*. 2006;354(11):1193.
11. Weintraub M, Sundaresan PR, Madan M, et al. Long-term weight control study. I (weeks 0 to 34). *Clin Pharmacol Ther*. 1992;51(5):586-594.
12. Kumar S, Boulton AJ, Beck-Nielsen H, et al. Troglitazone, an insulin action enhancer, improves metabolic control in NIDDM patients. *Diabetologia*. 1996;39(6):701-709.
13. Connolly HM, Crary JL, McGoon MD, et al. Valvular heart disease associated with fenfluramine-phentermine. *N Engl J Med*. 1997;337(9):581-588.
14. Fergusson DA, Hebert PC, Mazer CD, et al. A comparison of aprotinin and lysine analogues in high-risk cardiac surgery. *N Engl J Med*. 2008;358(22):2319-2331.
15. Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med*. 2001;344(10):699-709.
16. Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med*. 1991;324(12):781-788.
17. Ranieri VM, Thompson BT, Barie PS, et al. Drotrecogin alfa (activated) in adults with septic shock. *N Engl J Med*. 2012;366(22):2055-2064.
18. Polman CH, O'Connor PW, Havrdova E, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med*. 2006;354(9):899-910.
19. Rudick RA, Stuart WH, Calabresi PA, et al. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *N Engl J Med*. 2006;354(9):911-923.
20. Insull W Jr, Isaacsohn J, Kwiterovich P, et al. Efficacy and safety of cerivastatin 0.8 mg in patients with hypercholesterolaemia: the pivotal placebo-controlled clinical trial. *J Int Med Res*. 2000;28(2):47-68.
21. Charatan F. Bayer decides to withdraw cholesterol lowering drug. *BMJ*. 2001;323(7309):359.
22. Bashir R, Bourgeois FT, Dunn AG. A systematic review of methods to identify unreported links between trial registrations and their published results, and an analysis of ClinicalTrials.gov. *BMC Med Res Methodol*. 2022;22:44.
23. Bar-Ilan J, Halevi G. Post retraction citations in context: a case study. *Scientometrics*. 2017;113(1):547-565.
24. Hsiao TK, Schneider J. Continued use of retracted papers: temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. *Quant Sci Stud*. 2021;2(4):1144-1169.
25. Schneider J, Ye D, Hill AM, Whitehorn AS. Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. *Scientometrics*. 2020;125(3):2877-2913.
26. Staab TR, Walter M, Mariotti Nesurini S, Dintsios CM, Graf von der Schulenburg JM, Amelung VE, Ruof J. Market withdrawals of medicines in Germany after AMNOG: a comparison of HTA ratings and clinical guideline recommendations. *Health Econ Rev*. 2018;8(1):23.

Additional files

Additional file 1: Supplementary Material S1. Per-drug case files and audit trail for all 31 analyzed drugs, with the full pivotal-paper and correction-notice bibliography.

Additional file 2: Denominator and tier-resolution workbook (xlsx). Cohort by approval decade and tier; aggregate resolutions with PMIDs; exclusion list.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [03additionalfile1casefiles1.docx](#)
- [04additionalfile2denominatorworkbook1.xlsx](#)