



Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Multimodal Pharmacovigilance Needs a Hierarchy for Claims Derived from Reports, Records, and Digital Traces

Olivia Bennett^{1*}, Harry Collins², Jack Foster¹, Ethan Wright³

¹Department of Multimodal Pharmacovigilance and Evidence Hierarchy, Faculty of Pharmacy, University of Nottingham, Nottingham, United Kingdom.

²Department of Reports, Records, and Digital Traces, Faculty of Pharmacy, University of Southampton, Southampton, United Kingdom.

³Department of Drug Safety Informatics, Faculty of Pharmacy, Newcastle University, Newcastle, United Kingdom.

*Email: olivia.bennett@nottingham.ac.uk

ABSTRACT

Modern pharmacovigilance increasingly combines spontaneous reports, electronic health records, administrative claims, registries, regulatory documents, published literature, and patient-generated digital traces. Their coexistence creates an apparent opportunity for stronger safety inference, but multimodality alone does not determine what can legitimately be claimed. Each source observes a different part of the exposure–event process, imposes different selection and measurement mechanisms, and supports different transitions from observation to signal, corroboration, comparative estimation, causal interpretation, and regulatory action. This Evidence-Integration Perspective argues that multimodal pharmacovigilance therefore requires an explicit hierarchy of safety claims rather than an implicit hierarchy of data sources. The proposed approach separates evidence acquisition from claim strength and treats cross-source concordance as informative only after considering phenotype compatibility, exposure definition, risk-window alignment, population comparability, analytical design, source independence, and source-specific bias. Spontaneous reporting remains essential for hypothesis generation; longitudinal healthcare data can contextualize and test selected concerns; registries can provide structured clinical depth; regulatory and literature sources document assessment and action; and digital traces may supply timely patient-generated context while often lacking the conditions necessary for population-risk inference. The resulting hierarchy is intended as a disciplined framework for deciding what additional evidence is needed before a safety claim should be escalated.

Keywords: Pharmacovigilance, Safety signals, Spontaneous reporting, Real-world data, Pharmacoepidemiology, Multimodal evidence

Received: 29 January 2026; **Revised:** 24 March 2026; **Accepted:** 25 March 2026

INTRODUCTION

Post-marketing safety assessment now operates in an environment where a suspected drug–event relationship may be represented simultaneously in individual case safety reports, longitudinal clinical records, insurance claims, disease registries, published studies, regulatory assessments, product labels, online patient narratives, and algorithmically extracted features. This abundance changes the practical problem of pharmacovigilance. The challenge is no longer simply whether another source can be queried, but what inferential role that source can legitimately play. Spontaneous reporting databases and longitudinal healthcare databases are complementary across parts of signal management, yet their strengths, weaknesses, timing, data structures, and inferential

capabilities differ substantially [1]. Safety assessment also remains an iterative reasoning process in which heterogeneous evidence, uncertainty, clinical seriousness, and decision context must be integrated rather than collapsed into one statistic [2].

The need for this distinction is visible when signal generation, corroboration, and regulatory action are compared empirically. Analysis of publicly identified US Food and Drug Administration safety signals showed that regulatory actions and external corroboration did not map onto one another uniformly [3]. Absence of published corroboration does not establish absence of a causal association, and regulatory action does not by itself prove one. Both are better treated as evidence states whose meaning depends on what was examined, when it was examined, how the event and exposure were defined, and what decision had to be made.

This Perspective therefore argues for a hierarchy of claims, not a ranking of sources. The central proposed distinction is between an observation, a detection-level signal, contextual corroboration, comparative risk evidence, a causal interpretation, and an actionability judgment. A single source may support more than one level when its design permits, while several sources may collectively fail to justify escalation when they reproduce the same measurement error or confounding structure. The relevant question is consequently not “How many modalities agree?” but “What claim does each modality support, what uncertainty remains, and is the next inferential step warranted?”

Evidence modalities in modern pharmacovigilance

Source labels conceal substantial methodological variation. Electronic health record studies used for drug-safety signal identification vary in exposure and outcome definitions, data extraction, comparator choice, confounding adjustment, and validation strategy [4]. A systematic review of methods based on routinely collected healthcare data likewise identified multiple methodological families and heterogeneous approaches to performance evaluation; no method could be treated as uniformly optimal across drug–event contexts [5]. Unstructured clinical documentation adds another layer: natural-language processing and machine-learning methods can extract adverse-event information that may be absent from coded fields, but their usefulness depends on annotation, model validation, local documentation practice, and transportability [6].

These findings matter because the evidential status of a result is determined jointly by the source and the transformation applied to it. A clinician’s free-text note, an algorithmically extracted adverse-event concept, and an epidemiological association estimated from a longitudinal cohort are not interchangeable observations simply because all originate in an electronic health record. Across record-based evidence, differences in data representation, analytical method, and validation create distinct inferential boundaries [4–6].

The same principle applies across modalities. Spontaneous reports are unusually sensitive to unusual, serious, or clinically distinctive suspected reactions but generally lack the sampling structure required for incidence estimation. Claims and electronic records may provide denominators and longitudinal information but introduce outcome misclassification, incomplete capture, treatment-channeling effects, and confounding. Registries can collect clinically detailed variables prospectively yet may have selective enrolment and incomplete comparators. Regulatory documents summarize decisions made from broader evidence packages, while digital traces may reveal patient experiences that formal medical systems capture poorly. The proposed hierarchy therefore treats “evidence modality” as a description of where information originates, not as a preassigned level of certainty.

Spontaneous reports

Spontaneous reporting remains foundational because it can expose suspected drug–event patterns before formal comparative studies are available. Its quantitative use, however, spans numerous signal-detection methods with different assumptions and operating characteristics. A systematic review identified 31 approaches and concluded that the most appropriate method depends on the research purpose, database size, drug characteristics, event characteristics, and the relation being investigated [7]. A disproportionality value should therefore be interpreted as an analytical property of a reporting database under a specified design, not as an effect estimate that directly measures clinical risk.

Interpretability also depends on reporting transparency. READUS-PV was developed specifically because published disproportionality analyses have often incompletely described analytical choices or interpreted results beyond what the design supports [8]. Transparent specification of database, case definitions, reference groups, duplicate handling, time restrictions, statistical approach, sensitivity analyses, and known limitations enables readers to judge what was actually observed. Compliance with reporting guidance does not validate a suspected causal relation, but inadequate reporting weakens the ability to evaluate the result.

A further boundary appears between the statistical output and the language used to describe it. Meta-epidemiological work has shown that causal language associated with disproportionality findings can become stronger in subsequent citations [9]. An earlier evaluation of 100 published disproportionality studies found frequent spin and documented causal wording in abstracts or conclusions despite the hypothesis-generating nature of the underlying analyses [10]. The problem is not that disproportionality is incapable of contributing useful evidence; it is that its contribution changes when an association is silently rewritten as a risk estimate or causal finding.

Within the proposed hierarchy, a spontaneous-report pattern therefore enters primarily as detection-level evidence. Case clinical detail, dechallenge or rechallenge information, biological plausibility, reporting chronology, seriousness, and unusual phenotype may alter prioritization, but escalation of the *claim* requires an explicit reason. A stronger claim may emerge through case-series assessment, independent longitudinal evidence, formal epidemiological testing, or another source capable of addressing the uncertainty that remains. The magnitude of a reporting statistic alone does not supply that transition.

Electronic health records and claims

Longitudinal healthcare data can answer questions that spontaneous reports often cannot: whether an event is documented among exposed patients, when it occurs relative to exposure, what alternative diagnoses or treatments are present, how frequently relevant background events occur, and whether a comparative analytical design is feasible. In a distributed European network study, routinely collected health data contributed useful context to some—but not all—signals selected for validation and prioritization [11]. This conditional usefulness is important. Failure to materially inform an assessment may reflect data availability, phenotype definition, exposure capture, sample size, or analytical feasibility rather than proof against the signal.

Targeted record retrieval can also enlarge the clinical evidence base. A proof-of-concept study linking potential safety signals to hospital electronic health records identified additional relevant cases beyond those initially represented in spontaneous reporting [12]. Additional cases can strengthen the case narrative or reveal clinical context, but they do not become a population-risk estimate merely because they were retrieved from an EHR. The denominator, selection mechanism, ascertainment strategy, and comparator structure still determine which claims are permissible.

Longitudinal designs can support a different inferential step. Adaptation of the self-controlled case series method for recently marketed products illustrates how patient-time can be divided into exposed risk periods and comparison periods, thereby addressing stable between-person confounding while introducing assumptions of its own [13]. Such a design is analytically closer to comparative risk estimation than spontaneous disproportionality, but “self-controlled” should not be read as “confounding-free.” Event-dependent exposure, time-varying disease severity, changing prescribing, observation patterns, and risk-window specification can remain consequential.

Finally, corroboration is weakened when apparently separate analyses are not evidentially independent. Reusing the same data environment for both signal generation and evaluation can reproduce the same coding practices, population structure, and systematic errors [14]. Data reuse is not intrinsically invalid, but agreement generated inside the same information ecosystem should not automatically receive the same corroborative weight as convergence from a genuinely independent design.

Thus, EHRs and claims should not occupy one fixed rung of a hierarchy. A text-mined case, a descriptive prevalence estimate, a self-controlled analysis, and a carefully designed comparative cohort represent different claims even when derived from the same database. In the proposed framework, escalation depends on the analysis performed and the uncertainty addressed rather than on the label “real-world data.”

Registries

Registries occupy a distinctive position because they can deliberately capture clinically meaningful variables over time rather than relying only on information generated for reimbursement or routine documentation. The REGIMS experience in multiple sclerosis illustrates the potential for multicentre prospective pharmacovigilance to capture structured long-term treatment and safety information while also demonstrating the logistical, procedural, and data-quality demands of registry-based surveillance [15]. This depth is valuable, but depth and representativeness are different properties.

Regulatory utility therefore depends on fitness for purpose. Work arising from the European Medicines Agency Patient Registries Initiative identified heterogeneity of registry design, unreliable or variable data quality, data-sharing barriers, governance, and insufficient harmonization as obstacles to wider regulatory use [16]. A survey

focusing on rare-disease registries similarly emphasized common data elements, data quality, and governance as central considerations when registry data are expected to support regulatory decision making [17]. These studies collectively support a conditional interpretation of registry evidence: its strength depends on what population is enrolled, how exposures and outcomes are collected, what follow-up is achieved, what comparators exist, and whether the variables required for the regulatory question are available [15–17].

Harmonization can improve the ability to compare and synthesize registry-derived evidence without eliminating these limitations. The IMI ConcePTION initiative developed 98 core data elements for primary-source pregnancy pharmacovigilance to reduce heterogeneity in data collection and improve evidence synthesis [18]. Standard definitions can make disagreements easier to interpret because exposure, pregnancy timing, outcomes, and covariates are less likely to differ merely through inconsistent data architecture. They cannot, however, make populations exchangeable or remove confounding simply by standardizing fields.

For the proposed hierarchy, registry evidence is therefore most useful when its clinical depth resolves a specific uncertainty left by earlier signal evidence. The same registry may provide strong evidence about phenotype, timing, treatment persistence, pregnancy exposure, or disease course while remaining weak for comparative risk if an appropriate comparator is absent. Cross-source agreement should consequently be evaluated against the registry’s capture mechanism rather than awarded a generic “registry confirmation” label.

The registry problem is difficult to express as a single continuum because the same longitudinal dataset can strengthen one dimension of a claim while leaving another unresolved. **Table 1** therefore separates the conditions that determine whether registry evidence should contextualize, corroborate, test, or merely preserve uncertainty around a safety concern.

Table 1. Registry evidence as a conditional bridge between a safety concern and a stronger multimodal claim

Registry inference scenario	Dominant bias or information constraint	Evidential strength Proposed escalation rule	Clinical seriousness and reversibility	Cross-source concordance interpretation	Proposed hierarchy endpoint
Scientific anchor	Dominant bias or information constraint	Evidential strength Proposed escalation rule	Clinical seriousness and reversibility	Cross-source concordance interpretation	Proposed hierarchy endpoint
Detailed longitudinal cases reproduce the suspected phenotype and compatible exposure timing	Selective enrolment or incomplete comparator data	Evidential strength available from the registry Stronger clinical characterization; limited population-risk inference Proposed escalation rule Seek independent comparative evidence while preserving the clinical signal	Serious or irreversible events increase decision urgency	Concordance strengthens phenotype credibility if case definitions are aligned	Contextually corroborated safety concern
Registry captures exposure and outcome systematically but background risk remains poorly characterized	Disease-related event frequency or incomplete unexposed comparison	Evidential strength available from the registry Structured event occurrence with uncertain excess risk Proposed escalation rule Add external comparator or fit-for-purpose longitudinal design	Urgency depends on event severity and available alternatives	Agreement with reports supports persistence of concern, not excess-risk magnitude	Corroborated signal awaiting comparative estimation

Registry includes suitable treatment comparators and relevant confounders	Residual confounding, treatment switching, missing data	Evidential strength available from the registry Comparative association may be estimable Proposed escalation rule Test robustness across definitions, windows, and independent data	Seriousness informs action threshold but not causal certainty	Independent agreement with other designs increases confidence	Comparative safety evidence
Registry and spontaneous or record-based evidence disagree	Population, phenotype, exposure, follow-up, or ascertainment mismatch	Evidential strength available from the registry Discordance itself is diagnostically informative Proposed escalation rule Identify the source of discordance before changing claim level	High seriousness may justify continued assessment despite disagreement	Do not average conflicting findings	Unresolved multimodal safety claim
Registry is clinically rich but too small or too selective for the required inference	Sparse events and limited transferability	Evidential strength available from the registry Case-level mechanistic or clinical context only Proposed escalation rule Escalate investigation rather than certainty	Severe rare outcomes may still justify regulatory attention	Concordance may support prioritization without statistical confirmation	High-priority signal with limited inferential support

A registry-centred diagnostic map is also needed because bias does not operate only downstream of analysis; it can enter before enrolment, during phenotype construction, through exposure follow-up, and again when registry evidence is transferred to another population. **Figure 1** makes these diagnostic pathways explicit rather than presenting registry evidence as a one-directional bridge to stronger claims.

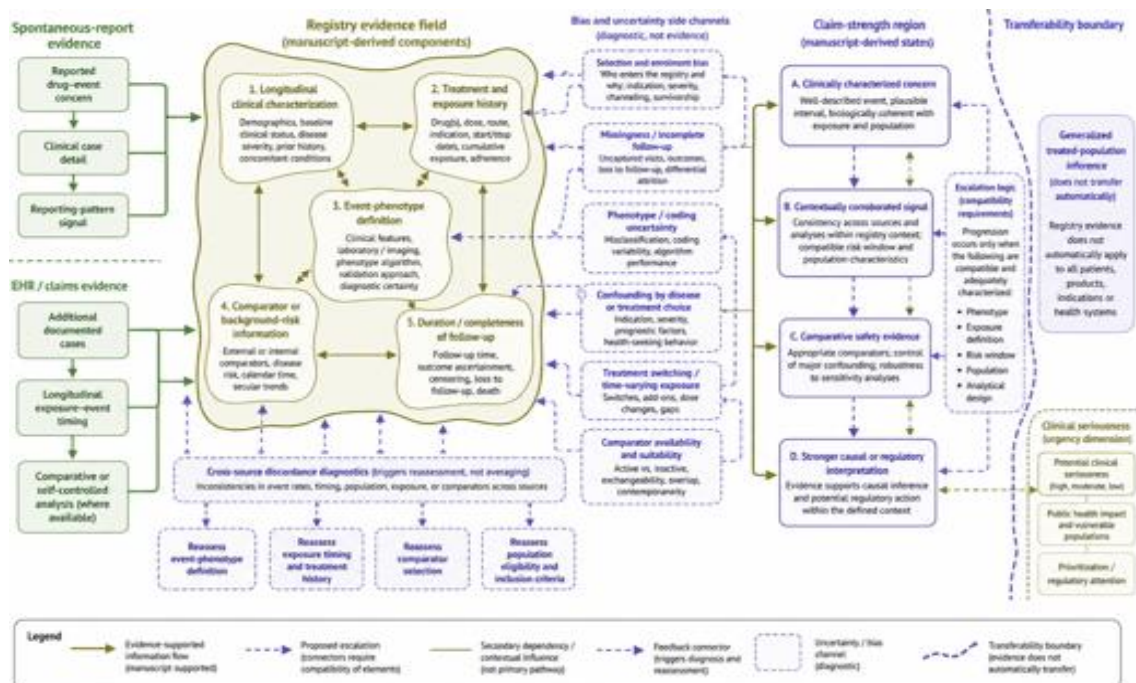


Figure 1. Registry concordance can strengthen a safety claim only after source-specific bias and transferability are diagnosed

Literature and regulatory sources

Published literature and regulatory documentation occupy another distinct evidential layer because they may summarize analyses, evaluations, decisions, or product-information changes rather than constitute a new

population observation. This distinction becomes especially important when regulatory action is interpreted retrospectively. Evaluation of European Union risk-minimisation interventions showed that impact studies varied in their ability to meet objectives and in the evidence they provided about effectiveness [19]. A regulatory intervention is therefore an action taken within a benefit–risk and public-health context; its existence should not be reverse-engineered into proof that every element of the underlying causal hypothesis was established.

PRAC activity illustrates how safety concerns may be processed through different regulatory procedures. A quantitative review of antidiabetic-drug evaluations found that drug–event concerns could move through periodic safety update reports, signal assessment and prioritisation, product-information changes, risk-management measures, and other procedures, with many evaluations resulting in no additional regulatory action [20]. Regulatory process consequently provides information about signal maturity and decision context, but the meaning of a given action depends on the procedure, evidence package, urgency and question being addressed. The increasing use of computational tools creates an additional boundary between regulatory information and its extraction. AskFDALabel demonstrated that large-language-model and retrieval-augmented methods can structure adverse-event information from FDA labeling documents [21]. The authority belongs to the underlying label; the model is an information-retrieval and transformation layer. Similarly, an AI-supported narrative search engine improved retrieval of relevant Vigibase case narratives for case-series assessment [22]. Improved retrieval can reduce information loss and assessor burden, but ranking or finding a case does not determine whether that case changes the causal interpretation.

These distinctions are sequential rather than merely categorical. Literature may trigger concern, a regulatory assessment may request additional evidence, a label may change, and later studies may support, qualify or weaken the original interpretation. **Table 2** therefore separates the transition between evidence states from the final claim that the article permits.

Table 2. Sequence conditions that determine what literature and regulatory evidence add to a multimodal safety claim

A reported or database-derived concern enters formal assessment Additional published observational evidence becomes available		
Governing condition	A reported or database-derived concern enters formal assessment	Additional published observational evidence becomes available
Risk-window requirement	Exposure and event timing must remain clinically compatible	Study risk window must correspond to the safety hypothesis
Precursor evidence state	Detection-level signal	Unresolved or contextually corroborated concern
Transition trigger	Regulatory review or focused literature evaluation	Comparative or longitudinal analysis
Observable regulatory or literature readout	Formal assessment, request for information, or documented evaluation	Association, null finding, or context-specific result
Condition that interrupts or reverses the sequence	Phenotype mismatch, implausible timing, or major alternative explanation	Incompatible population, exposure definition, comparator or outcome
Proposed article-level test	Does the new source address uncertainty or merely restate the original signal?	Is the evidence sufficiently independent and fit for the claim being escalated?
Regulatory action is recommended Regulatory text is computationally extracted or searched		
Governing condition	Regulatory action is recommended	Regulatory text is computationally extracted or searched
Risk-window requirement	Urgency may be shorter than the time required for causal resolution	Source-document version and temporal validity must be preserved
Precursor evidence state	Evidence sufficient for a decision under uncertainty	Authoritative text already exists

Transition trigger	Seriousness, preventability, accumulated evidence or benefit–risk concern	NLP, retrieval or LLM processing
Observable regulatory or literature readout	Label update, risk-management action, further study or other regulatory measure	Structured adverse-event information or retrieved case narratives
Condition that interrupts or reverses the sequence	Subsequent evidence changes benefit–risk interpretation	Extraction error, outdated source, retrieval failure or loss of context
Proposed article-level test	Is actionability being kept separate from causal certainty?	Is the algorithm being treated as an access layer rather than a new clinical source?
Later evidence conflicts with the earlier regulatory or literature interpretation		
Governing condition	Later evidence conflicts with the earlier regulatory or literature interpretation	
Risk-window requirement	Comparable exposure periods and event definitions are required	
Precursor evidence state	Mature but revisable safety interpretation	
Transition trigger	Independent contradictory evidence	
Observable regulatory or literature readout	Reassessment, qualification or unchanged decision	
Condition that interrupts or reverses the sequence	Non-comparable designs may create apparent contradiction	
Proposed article-level test	Has discordance been diagnosed before the claim is strengthened or weakened?	

Patient-generated and digital data

Patient-generated digital information expands the observable safety landscape because symptoms, treatment experiences, adherence problems, perceived formulation changes, and concerns may be discussed outside conventional clinical encounters. Social-media data have consequently been proposed as real-world data that could support selected pharmaceutical-development and regulatory questions, but causal inference, representativeness, data provenance, and alignment with regulatory standards remain substantial constraints [23]. The fact that a digital trace is temporally immediate or patient-authored does not determine the level of safety claim it can support.

Comparative evidence reinforces this distinction. A scoping review of social-media analysis for adverse-event detection found heterogeneous methods and mixed comparative results, with social platforms potentially providing patient perspectives or information not readily visible in traditional sources while not establishing a consistent superiority for pharmacovigilance detection [24]. Digital data are particularly vulnerable to unknown denominators, repeated or copied content, platform-specific participation, differential visibility, changing algorithms, publicity effects, and uncertainty over whether medication use and clinical events have been accurately interpreted.

Temporal patterns can nevertheless serve as useful investigative triggers. An AI-based analysis of patient behavior around the French levothyroxine formulation controversy evaluated changes in linguistic and behavioral indicators over time and proposed their use for identifying abnormal periods requiring attention [25]. Such a pattern is appropriately interpreted as an attention signal: it can identify when investigation should intensify, but it does not itself distinguish a pharmacological adverse reaction from media amplification, nocebo effects, changes in healthcare access, product-switching experiences, or other contemporaneous influences.

The strongest boundary comes from comparative signal-detection evidence. The WEB-RADR evaluation found poor performance for broad-ranging statistical signal detection using the examined Twitter and Facebook data

relative to established spontaneous-report data [26]. Subsequent project recommendations therefore emphasized purpose-specific rather than indiscriminate incorporation of social media into pharmacovigilance [27]. Importantly, this negative result does not render patient-generated evidence useless. It demonstrates why a multimodal framework must separate novelty of source from fitness for a particular claim.

Across the available evidence, digital traces may contribute information about patient experience, unexpected terminology, emerging attention, misuse, adherence, or contextual changes while performing poorly for other tasks such as broad statistical signal detection [23, 24, 26]. The proposed hierarchy therefore places unvalidated digital observations at a hypothesis or context-generating level. Escalation requires progressively stronger checks: credible medication and event identification, duplication control, temporal anchoring, evidence that the digital pattern is not explained primarily by publicity or platform dynamics, and comparison with clinically grounded sources. Agreement with spontaneous reports or records becomes meaningful only if the digital observations are genuinely independent rather than repetitions of the same public event.

Digital pharmacovigilance thus illustrates the broader argument of this Perspective. A new modality does not acquire evidential strength merely by adding another channel to a multimodal system. Its value depends on what it observes better than existing sources, which biases it introduces, and whether a proposed inference exceeds those observational capabilities. The next analytical problem is therefore not simply concordance, but how concordance and discordance should be interpreted when the contributing modalities have different and sometimes correlated failure modes.

Signal concordance and discordance

Cross-source concordance is often treated as the strongest argument for multimodal pharmacovigilance, but agreement is informative only when the contributing evidence is sufficiently independent. Reanalysis within the same underlying population or data environment may reproduce common coding, exposure, ascertainment, or analytical errors [14]. Concordance should therefore be interpreted after examining provenance, phenotype definition, exposure definition, risk window, population overlap, and methodological dependence.

Discordance can be equally informative. Regulatory signals may receive uneven external corroboration [3], while longitudinal healthcare analyses can identify alternative explanations or fail to add useful information for particular signals [11]. The proposed framework therefore treats disagreement as a diagnostic state rather than a simple failure of replication. Before one source is privileged over another, investigators should ask whether different populations, event definitions, exposure windows, treatment patterns, or ascertainment mechanisms explain the divergence.

Digital evidence makes this requirement especially visible. Social-media findings may contribute patient perspective while differing substantially from conventional signal-detection outputs [24]. Apparent agreement between digital and formal sources should consequently be checked for common publicity events, duplicated content, and transfer of information from one channel into another. Multimodal confirmation requires more than repeated visibility of the same concern.

Source-specific bias

Source-specific bias can enter before statistical analysis begins. Ambiguous adverse-event narratives may be translated differently into standardized MedDRA terms, producing variation at the phenotype-construction stage [28]. A coded event should therefore not be treated as a transparent copy of the underlying clinical narrative when ambiguity materially affects classification.

Analytical choices can introduce another layer. In FAERS disproportionality analyses, alternative comparator restrictions changed the behavior of signal estimates, and no single comparator produced a uniformly preferable bias pattern across the evaluated drug–event examples [29]. Signal magnitude is consequently conditional on the reference structure used to generate it.

These mechanisms differ across modalities. Electronic records are affected by documentation and confounding structures [4]; spontaneous-report analyses depend on reporting behavior and detection method [7]; registries may be selective despite deep longitudinal characterization [15]; and digital platforms combine participation, duplication, publicity, and denominator problems [24]. The proposed hierarchy therefore does not assign a fixed rank to a source. It asks which bias mechanisms remain capable of explaining the observed pattern and which additional evidence would discriminate among them.

Because source bias can make the same apparent concordance either persuasive or misleading, **Table 3** organizes the boundary conditions that must be resolved before a multimodal claim is escalated.

Table 3. Boundary conditions that determine whether cross-source agreement strengthens or weakens a pharmacovigilance claim

Source-specific case	Boundary	Context	Interpretation and Validation and Observation and Decision
Scientific anchor	Boundary variable beside the interpretation	Cross-source concordance	Proposed escalation rule Validation need Signal source involved Manuscript-specific consequence
Similar phenotype appears across sources but coding rules differ	Event-definition compatibility	Potentially superficial	Proposed escalation rule Reconcile phenotype construction before escalation Validation need Clinical or narrative adjudication Signal source involved Reports, records, registries Manuscript-specific consequence Agreement remains provisional
Exposure–event timing aligns but populations differ materially	Population transferability	Context-dependent	Proposed escalation rule Preserve population-specific interpretation Validation need Replication in relevant treated population Signal source involved Records, registries Manuscript-specific consequence Claim cannot automatically generalize
Similar signal emerges from non-independent databases	Evidence independence	Potentially correlated	Proposed escalation rule Trace data provenance and overlap Validation need Independent source or design Signal source involved Reports, claims, records Manuscript-specific consequence Apparent replication receives reduced corroborative weight
Disproportionality changes with comparator restriction	Analytical reference structure	Unstable	Proposed escalation rule Test alternative defensible comparators Validation need Sensitivity analysis and external evidence Signal source involved Spontaneous reports Manuscript-specific consequence Signal magnitude remains design-conditional
Digital attention rises with formal reporting after publicity	Common external stimulus	Potentially inflated	Proposed escalation rule Separate media-driven amplification from clinical evidence Validation need Clinically grounded independent assessment Signal source involved Digital traces, reports Manuscript-specific consequence Visibility is not equivalent to confirmation

Sources disagree despite aligned definitions and timing	Residual biological or methodological heterogeneity	Informative discordance	Proposed escalation rule
			Investigate effect modification, confounding, and source limitations
			Validation need
			Fit-for-purpose comparative evidence
			Signal source involved
			Any multimodal combination
			Manuscript-specific consequence
			Claim remains unresolved rather than averaged

The same limitations operate through time, exposure and background risk rather than as isolated methodological defects. **Figure 2** therefore depicts bias as a set of parallel evidence trajectories that can converge only after temporal and population boundaries have been examined.

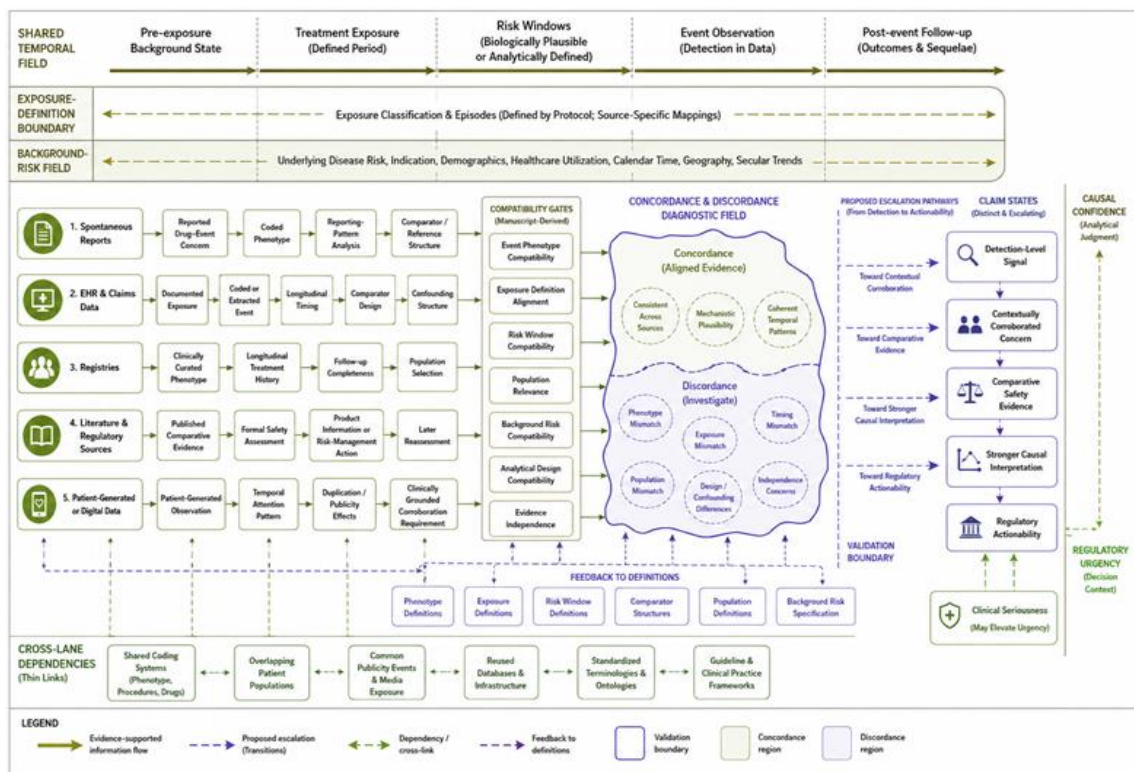


Figure 2. Time, exposure and background risk determine whether multimodal concordance is evidential or merely apparent

A hierarchy for multimodal safety claims

The proposed hierarchy begins from the premise that claim strength should follow inferential burden, not source prestige. Spontaneous reports may justify detection; independent clinical records or registries may provide contextual corroboration; suitable longitudinal comparative designs may support relative-risk inference; and stronger causal claims require progressively better control of alternative explanations. The complementarity of spontaneous and longitudinal evidence supports this separation without establishing the hierarchy itself [1].

Cross-source agreement increases confidence only when the sources are sufficiently compatible and their errors are not simply repeated [14]. The hierarchy therefore asks whether phenotype, exposure, timing, population and analytical question align before concordance is treated as escalation evidence. It also distinguishes causal confidence from decision urgency. Serious or preventable harm may justify rapid regulatory attention even while causal uncertainty remains, because pharmacovigilance decisions integrate evidence with clinical and public-health consequences [2].

The framework is explicitly proposed rather than empirically validated. Its purpose is to prevent an evidential shortcut: moving from “seen in several places” to “causally established” without identifying which uncertainty each source has actually resolved.

Escalation and validation rules

Escalation should begin by defining the unresolved inferential deficit. If case identity is uncertain, additional case characterization may be appropriate; if background risk or temporality is unresolved, longitudinal healthcare data may add value [11]; if apparent agreement may arise from coding ambiguity, the phenotype requires reassessment [28]; and if disproportionality depends strongly on the reference group, alternative comparator structures and independent evidence are required [29].

The proposed rule is therefore question-matched validation: choose the next evidence source or design because it can address a specified uncertainty, not merely because it is available. Multimodal evidence should be escalated only after its independence, compatibility and bias structure are examined. Regulatory action may still occur before complete causal resolution when seriousness or preventability justifies precaution, but subsequent evaluation should remain capable of strengthening, qualifying or reversing the interpretation [19].

CONCLUSION

Multimodal pharmacovigilance is most useful when it preserves differences among evidence sources rather than erasing them. Spontaneous reports, longitudinal records, registries, regulatory and literature sources, and digital traces observe different parts of the safety problem and carry different failure modes. Their agreement can strengthen inference, but only after compatibility, independence, timing, population, background risk and source-specific bias are examined. The proposed hierarchy therefore separates detection, contextual corroboration, comparative safety evidence, causal interpretation and actionability while keeping regulatory urgency distinct from causal certainty. Its central principle is simple: stronger safety claims should require evidence that resolves stronger alternative explanations, not merely a larger number of data streams.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Crisafulli S, Bate A, Brown JS, Candore G, Chandler RE, Hammad TA, et al. Interplay of spontaneous reporting and longitudinal healthcare databases for signal management: Position statement from the Real-World Evidence and Big Data Special Interest Group of the International Society of Pharmacovigilance. *Drug Saf.* 2025;48(9):959-76. doi:10.1007/s40264-025-01548-3
2. Hammad TA, Davies S. Navigating diverging perspectives: Reasoning, evidence, and decision-making in drug safety. *Drug Saf.* 2025;48(6):587-93. doi:10.1007/s40264-025-01537-6
3. Dhodapkar MM, Shi X, Ramachandran R, Chen EM, Wallach JD, Ross JS. Characterization and corroboration of safety signals identified from the US Food and Drug Administration Adverse Event Reporting System, 2008-19: Cross sectional study. *BMJ.* 2022;379. doi:10.1136/bmj-2022-071752
4. Davis SE, Zobotka L, Desai RJ, Wang SV, Maro JC, Coughlin K, et al. Use of electronic health record data for drug safety signal identification: A scoping review. *Drug Saf.* 2023;46(8):725-42. doi:10.1007/s40264-023-01325-0
5. Coste A, Wong A, Bokern M, Bate A, Douglas IJ. Methods for drug safety signal detection using routinely collected observational electronic health care data: A systematic review. *Pharmacoepidemiol Drug Saf.* 2023;32(1):28-43. doi:10.1002/pds.5548
6. Golder S, Xu D, O'Connor K, Wang Y, Batra M, Gonzalez Hernandez G. Leveraging natural language processing and machine learning methods for adverse drug event detection in electronic health/medical records: A scoping review. *Drug Saf.* 2025;48(4):321-37. doi:10.1007/s40264-024-01505-6
7. Jiao XF, Pu L, Lan S, Li H, Zeng L, Wang H, et al. Adverse drug reaction signal detection methods in spontaneous reporting system: A systematic review. *Pharmacoepidemiol Drug Saf.* 2024;33(3). doi:10.1002/pds.5768

8. Fusaroli M, Salvo F, Begaud B, AlShammari TM, Bate A, Battini V, et al. The reporting of a disproportionality analysis for drug safety signal detection using individual case safety reports in pharmacovigilance (READUS-PV): Development and statement. *Drug Saf.* 2024;47(6):575-84. doi:10.1007/s40264-024-01421-9
9. Bernardeau C, Revol B, Salvo F, Fusaroli M, Raschi E, Cracowski JL, et al. Are causal statements reported in pharmacovigilance disproportionality analyses using individual case safety reports exaggerated in related citations? A meta-epidemiological study. *Drug Saf.* 2025;48(6):679-88. doi:10.1007/s40264-025-01524-x
10. Mouffak A, Lepelley M, Revol B, Bernardeau C, Salvo F, Pariente A, et al. High prevalence of spin was found in pharmacovigilance studies using disproportionality analyses to detect safety signals: A meta-epidemiological study. *J Clin Epidemiol.* 2021;138:73-9. doi:10.1016/j.jclinepi.2021.06.022
11. Gauffin O, Brand JS, Hedfors Vidlin S, Sartori D, Asikainen S, Català M, et al. Supporting pharmacovigilance signal validation and prioritization with analyses of routinely collected health data: Lessons learned from an EH DEN network study. *Drug Saf.* 2023;46(12):1335-52. doi:10.1007/s40264-023-01353-w
12. Abedian Kalkhoran H, Zwaveling J, van Hunsel F, Kant A. An innovative method to strengthen evidence for potential drug safety signals using electronic health records. *J Med Syst.* 2024;48(1):51. doi:10.1007/s10916-024-02070-2
13. Zhou X, Douglas IJ, Shen R, Bate A. Signal detection for recently approved products: Adapting and evaluating self-controlled case series method using a US claims and UK electronic medical records database. *Drug Saf.* 2018;41(5):523-36. doi:10.1007/s40264-017-0626-y
14. Wang SV, Kulldorff M, Glynn RJ, Gagne JJ, Pottegård A, Rothman KJ, et al. Reuse of data sources to evaluate drug safety signals: When is it appropriate? *Pharmacoepidemiol Drug Saf.* 2018;27(6):567-9. doi:10.1002/pds.4442
15. Simbrich A, Thibaut J, Khil L, Maximov S, Wiendl H, Berger K, et al. Chances and challenges of registry-based pharmacovigilance in multiple sclerosis: Lessons learnt from the implementation of the multicenter REGIMS registry. *Drug Saf.* 2021;44(1):7-15. doi:10.1007/s40264-020-01007-1
16. McGettigan P, Alonso Olmo C, Plueschke K, Castillon M, Nogueras Zondag D, Bahri P, et al. Patient registries: An underused resource for medicines evaluation: Operational proposals for increasing the use of patient registries in regulatory assessments. *Drug Saf.* 2019;42(11):1343-51. doi:10.1007/s40264-019-00848-9
17. Jonker CJ, de Vries ST, van den Berg HM, McGettigan P, Hoes AW, Mol PGM. Capturing data in rare disease registries to support regulatory decision making: A survey study among industry and other stakeholders. *Drug Saf.* 2021;44(8):853-61. doi:10.1007/s40264-021-01081-z
18. Richardson JL, Moore A, Bromley RL, Stellfeld M, Geissbühler Y, Bluett-Duncan M, et al. Core data elements for pregnancy pharmacovigilance studies using primary source data collection methods: Recommendations from the IMI ConcePTION project. *Drug Saf.* 2023;46(5):479-91. doi:10.1007/s40264-023-01291-7
19. Goedecke T, Martirosyan L, Gault N, Seifert K, Morales DR, Bahri P, et al. Studying the impact of European Union regulatory interventions for minimising risks from medicines: Lessons learnt and recommendations. *Pharmacoepidemiol Drug Saf.* 2024;33(8). doi:10.1002/pds.5874
20. Le H, Sindahl P, Andersen M, Hallgreen CE. Understanding the work of the Pharmacovigilance Risk Assessment Committee (PRAC): A quantitative review of the post-authorisation safety evaluation of antidiabetic drugs from 2012 to 2022. *Drug Saf.* 2025;48(7):781-94. doi:10.1007/s40264-025-01536-7
21. Wu L, Fang H, Qu Y, Xu J, Tong W. Leveraging FDA labeling documents and large language model to enhance annotation, profiling, and classification of drug adverse events with AskFDALabel. *Drug Saf.* 2025;48(6):655-65. doi:10.1007/s40264-025-01520-1
22. Zekarias A, Meldau EL, Bista S, China JF, Sandberg L. Narrative search engine for case series assessment supported by artificial intelligence query suggestions. *Drug Saf.* 2025;48(7):767-79. doi:10.1007/s40264-025-01529-6
23. Wessel D, Pogrebnyakov N. Using social media as a source of real-world data for pharmaceutical drug development and regulatory decision making. *Drug Saf.* 2024;47(5):495-511. doi:10.1007/s40264-024-01409-5
24. Golder S, O'Connor K, Wang Y, Klein A, Gonzalez Hernandez G. The value of social media analysis for adverse events detection and pharmacovigilance: Scoping review. *JMIR Public Health Surveill.* 2024;10. doi:10.2196/59167

25. Roche V, Robert JP, Salam H. A holistic AI-based approach for pharmacovigilance optimization from patients behavior on social media. *Artif Intell Med.* 2023;144:102638. doi:10.1016/j.artmed.2023.102638
26. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN, et al. Assessment of the utility of social media for broad-ranging statistical signal detection in pharmacovigilance: Results from the WEB-RADR project. *Drug Saf.* 2018;41(12):1355-69. doi:10.1007/s40264-018-0699-2
27. van Stekelenborg J, Ellenius J, Maskell S, Bergvall T, Caster O, Dasgupta N, et al. Recommendations for the use of social media in pharmacovigilance: Lessons from IMI WEB-RADR. *Drug Saf.* 2019;42(12):1393-407. doi:10.1007/s40264-019-00858-7
28. Garmann T, Samdal H, Sartori D, Jahanlu D, Andersen F, Rocca E. Strategies and challenges in coding ambiguous information using MedDRA: An exploration among Norwegian pharmacovigilance officers. *Drug Saf.* 2025;48(11):1253-69. doi:10.1007/s40264-025-01573-2
29. Gravel CA, Bai W, Douros A. Comparators in pharmacovigilance: A quasi-quantification bias analysis. *Drug Saf.* 2024;47(8):809-19. doi:10.1007/s40264-024-01433-5