



Review Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Living Systematic Surveillance of Safety Signals from Spontaneous Reports to Federated Health Data

Maria Silva^{1*}, Joao Pereira¹, Bruno Costa², Ricardo Alves³

¹Department of Living Systematic Surveillance and Safety Signals, Faculty of Pharmacy, University of Coimbra, Coimbra, Portugal.

²Department of Spontaneous Reports and Federated Data, Faculty of Pharmacy, University of Porto, Porto, Portugal.

³Department of Drug Safety Informatics, Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal.

*Email: maria.silva@uc.pt

ABSTRACT

Pharmacovigilance increasingly operates across data environments that differ fundamentally in how exposures, adverse events, populations, denominators, time, and clinical context are represented. Spontaneous reporting remains indispensable for early detection, but signal interpretation increasingly requires longitudinal healthcare data, registries, and distributed analytical networks. Conventional episodic reviews poorly accommodate the continual arrival of new signals, methodological changes, and cross-source evidence. To develop an evidence-bounded living systematic surveillance approach that preserves source-specific inference while enabling recurrent reassessment of drug-safety signals as evidence accumulates across spontaneous reports and longitudinal health data. A living systematic review framework was constructed for peer-reviewed evidence published from 2017 through 2026. The review distinguishes evidence used for signal detection, contextualization, analytical testing, bias assessment, replication, and escalation. Living-review principles govern updating, versioning, eligibility reassessment, and documentation of methodological changes. Evidence is interpreted according to data provenance, exposure definition, adverse-event phenotype, temporal structure, background risk, confounding, clinical seriousness, and cross-source consistency. The evidence indicates that no surveillance source can independently support the full transition from signal emergence to actionable safety inference. Spontaneous reporting is especially sensitive to unexpected patterns but lacks reliable population denominators and is vulnerable to selective reporting. Longitudinal healthcare data offer denominators and temporality but introduce exposure, outcome, indication, and healthcare-recording biases. The synthesis therefore supports source-specific signal states rather than a single hierarchy of evidence. Cross-source agreement strengthens some interpretations but is not itself proof of causality because shared or correlated biases may generate concordance. Living drug-safety surveillance should be organized as a versioned process of detection, testing, reconciliation, and escalation rather than as a one-time search for statistical signals. The proposed architecture preserves distinctions between statistical disproportionality, supported association, causal interpretation, clinical importance, and regulatory actionability.

Keywords: Pharmacovigilance, Drug safety, Signal detection, Spontaneous reporting, Real-world data, Living systematic review

Received: 06 August 2026; **Revised:** 02 October 2026; **Accepted:** 05 October 2026

INTRODUCTION

Post-marketing drug safety is an intrinsically time-varying evidence problem. New products enter populations that differ from trial cohorts, utilization patterns change, indications expand, treatment switching occurs, and

adverse-event phenotypes may become recognizable only after broad exposure. A conventional systematic review provides an evidence state at a defined time, whereas a living systematic review is designed to incorporate consequential new evidence as it becomes available and to make subsequent versions transparent [1]. For pharmacovigilance, this distinction is more than procedural. The inferential status of a safety concern may change because the underlying evidence changes, because a better analytical design becomes available, or because the clinical context in which earlier evidence was interpreted has changed.

Operational experience with living reviews shows that their sustainability depends on more than repeatedly rerunning a search. Search schedules, screening, extraction, author responsibilities, publication thresholds, and dissemination must be designed as a recurrent system [2]. Methodological guidance has likewise developed unevenly across conduct, reporting, appraisal, and publication, making explicit governance of living mode necessary rather than assuming that a conventional review becomes “living” simply through occasional updating [3]. Practical experience during rapidly changing evidence environments further demonstrates the importance of deciding not only when to initiate and update a living review but also when continued living status ceases to provide sufficient value [4].

These considerations are especially important in drug safety because evidence is generated across sources that answer different questions. Early individual case reports may identify an unusual clinical pattern before comparative epidemiology is possible. Conversely, a longitudinal database may reveal that an apparent excess disappears after appropriate comparison, risk-window specification, or control of indication. The original conception of living systematic reviews therefore provides a useful methodological starting point: currency has value when new evidence is sufficiently likely to alter conclusions or decisions [5]. Sustainability can also be improved by combining human judgement with automation for selected repetitive review tasks, while retaining expert adjudication where contextual interpretation cannot safely be automated [6].

This review applies those principles to a more specific problem: how drug-safety evidence should remain “alive” while moving from spontaneous reports toward increasingly structured longitudinal and federated data. Its central argument is that surveillance should not be represented as a ladder in which later data sources automatically supersede earlier ones. Instead, each source changes what can be observed, estimated, challenged, or excluded. The proposed living-surveillance architecture therefore separates detection strength, inferential strength, clinical seriousness, cross-source concordance, and actionability and treats transitions among these states as revisable rather than permanent.

MATERIALS AND METHODS

The review was designed as a Living Systematic Review and its reporting logic follows contemporary systematic-review and living-review principles. PRISMA 2020 provides the general basis for transparent identification, screening, eligibility, and reporting of systematic reviews [7]. Because much of the substantive evidence considered in this article originates from routinely collected healthcare information, methodological appraisal also incorporates RECORD-PE principles concerning database provenance, coding, study design, exposure and outcome definition, and reproducibility [8].

The evidence window was restricted to peer-reviewed literature published from 2017 through 2026. The scope comprised living-review methodology and pharmacovigilance evidence concerning spontaneous reporting systems, longitudinal electronic health records, administrative claims, registries, distributed or federated data networks, bias control, signal evaluation, and cross-source interpretation. Evidence was coded according to its principal analytical function: signal detection, characterization, comparative testing, bias assessment, methodological validation, replication, or surveillance architecture. No effect estimate was pooled because the included methodological and empirical literature addressed heterogeneous drug-event questions, source structures, and inferential objectives.

A central methodological rule was preservation of the distinction between a reported observation and the inference assigned to it. Disproportionality, incidence, association, causal interpretation, clinical significance, and regulatory actionability were therefore coded as different evidential states. A source could contribute to more than one state, but movement from one state to another required additional evidential justification rather than stronger terminology alone.

Living review protocol

Living mode was defined as a controlled process of evidence reassessment rather than continuous rewriting. Each update cycle is required to retain the prior version's search logic, eligibility criteria, extracted evidence state, and interpretation while separately documenting new evidence and methodological changes. A new version should therefore identify whether its conclusions changed because additional reports accumulated, a new population or product entered the evidence base, an exposure or event definition changed, a more appropriate comparator became available, or a previous methodological limitation was resolved.

The review remains eligible for living status while new safety evidence is likely to alter signal classification, confidence, clinical interpretation, or escalation priority. Living status should be retired when the relevant evidence has stabilized sufficiently that repeated updates are unlikely to change interpretation, when surveillance has been superseded by a more authoritative continuing process, or when the burden of maintaining living mode is disproportionate to the probability of consequential change. Retirement does not imply that the drug-event relationship is definitively resolved; it marks the end of this review's recurrent-update function.

Data sources

Evidence sources were separated by what they make observable. Spontaneous reporting systems preserve individual suspected adverse-event reports and are capable of surfacing rare, unexpected, or clinically distinctive patterns that were not prespecified during development. VigiBase exemplifies the global scale and longitudinal accumulation achievable through international individual-case safety reporting [9]. Such databases, however, primarily record reports rather than exposed populations. Reporting intensity, completeness, duplication, notoriety, stimulated reporting, indication, geography, and time can therefore change the apparent distribution of drug-event combinations.

Longitudinal healthcare data were treated differently. Electronic health records and claims can provide defined populations, observable treatment trajectories, temporal ordering, healthcare utilization, and potential comparators, but the meaning of a coded exposure or outcome depends on the source and operational definition. Registries can provide richer disease- or product-specific information but may select specialized populations. Distributed networks permit common analytical protocols to be executed across multiple databases while data remain locally controlled. The review consequently rejects the idea that surveillance sources can be ranked solely by size or structure. Their value depends on whether the data source is fit for the safety question being asked.

Search and update frequency

Search updating was conceptualized through two mechanisms. The first is scheduled surveillance, in which the literature search is rerun at a prespecified interval to identify methodological or empirical evidence that may alter the review. The second is event-triggered reassessment, initiated when a potentially consequential new safety finding, analytical method, database capability, regulatory action, or cross-source contradiction becomes available.

The distinction prevents search frequency from being treated as an intrinsic measure of review quality. Very frequent searching may produce numerous non-consequential records, whereas an important methodological study can require immediate reassessment even between scheduled searches. The relevant unit is therefore not merely elapsed time but the probability that newly available information changes signal classification or the confidence with which an existing interpretation can be maintained.

Eligibility

Eligible publications were required to contribute directly to at least one component of the surveillance problem: living-review methodology; spontaneous-report signal detection or interpretation; longitudinal healthcare-data safety evaluation; registry-based surveillance; federated or distributed analysis; confounding or bias control; temporal or exposure specification; evidence validation; or cross-source signal management. Studies were not treated as equally informative merely because they contained pharmacovigilance terminology.

Eligibility was also inference-specific. A spontaneous-report study could inform detection methodology without supporting incidence estimation. A claims study could inform comparative risk while being uninformative about clinically nuanced phenotypes absent from coding. Methodological evidence could support a bias-control rule without demonstrating the performance of that rule for every drug-event pair. This role-specific approach was necessary to prevent topical relevance from being mistaken for evidential adequacy.

Signal classification

Reporting standards for pharmacovigilance disproportionality analyses emphasize that database construction, case definitions, analytical choices, thresholds, sensitivity analyses, and interpretation must be sufficiently transparent for readers to understand how a reported signal was generated [10]. A systematic review of spontaneous-report signal-detection approaches identified substantial methodological diversity and found that different approaches have different advantages, disadvantages, and appropriate application contexts rather than one universally superior method [11]. Accordingly, this review does not define a signal solely by whether one algorithm crosses a numerical threshold.

Four inferential states are distinguished. A detected statistical pattern is an unusual drug-event reporting configuration requiring assessment. A supported safety association has additional temporal, clinical, comparative, or replicated evidence but remains distinguishable from causal proof. An unresolved safety concern is a pattern for which plausible competing explanations remain material. An escalated safety concern is one whose seriousness and evidential maturity justify more intensive evaluation or action. These states constitute a proposed classification rather than an externally validated scoring system.

Causal interpretation requires particular restraint. Recent work applying causal graphs to disproportionality analysis demonstrates that pharmacovigilance associations can be affected by structures involving indication, reporting behavior, co-medication, event-related drug use, selection, and other dependencies that simple disproportionality measures do not resolve [12]. Meta-epidemiological evidence also shows that causal wording can become stronger when disproportionality findings are subsequently cited, creating interpretive inflation beyond what the original analysis establishes [13]. Signal classification must therefore encode what remains unresolved rather than reward linguistic certainty.

Risk-of-bias and evidence certainty

Risk of bias was assessed at the level of the specific inference rather than assigned as a single quality label to an entire data source. A spontaneous-report database can be highly informative for identifying an unusual event pattern while simultaneously being unsuitable for estimating population incidence. Conversely, longitudinal data can provide denominator information yet remain vulnerable to confounding by indication, outcome misclassification, incomplete capture of treatment, differential healthcare contact, and incorrectly specified risk windows.

Evidence certainty was therefore treated as multidimensional. The first dimension is measurement credibility: whether drug exposure, event phenotype, and relevant timing are represented with sufficient validity. The second is comparative credibility: whether the design addresses the counterfactual question implied by the interpretation. The third is bias resistance: whether plausible alternative explanations remain capable of producing the observed pattern. The fourth is cross-source stability: whether evidence generated under meaningfully different source structures is compatible after accounting for differences in population, estimand, exposure, and outcome definition. The fifth is clinical consequence, which concerns seriousness, reversibility, preventability, and the consequences of delayed action.

These dimensions are deliberately not collapsed into a numerical certainty score. A clinically catastrophic but uncertain signal may justify rapid investigation, whereas a statistically stable association involving a minor reversible event may have a different escalation pathway. **Table 1** therefore organizes bias assessment as a claim-to-test problem rather than a hierarchy.

The distinctions required for signal adjudication are summarized below because the same apparent association can demand different tests depending on the mechanism capable of generating it.

Table 1. Claim-to-Test Adjudication of Bias and Evidential Certainty in Living Drug-Safety Surveillance

Evidence configuration	Interpretive alternatives	Adjudication requirement	
Inference scenario / Background-risk problem / Cross-source interpretation	bias mechanism	Evidence needed to discriminate explanations / Clinical consequence	Permitted surveillance endpoint
Inference scenario: Unexpected reporting excess without a reliable denominator	Selective or stimulated reporting	Evidence needed to discriminate explanations: Clinical case review plus denominator-bearing longitudinal analysis	Statistical signal requiring targeted evaluation

Background-risk problem: Baseline occurrence cannot be estimated from reports alone Cross-source interpretation: Concordance is informative only if the second source measures a comparable phenotype		Clinical consequence: Seriousness may justify accelerated evaluation	
Inference scenario: Association concentrated within the treated disease population Background-risk problem: Disease itself predicts the outcome Cross-source interpretation: Persistence after indication-sensitive analysis increases confidence		Evidence needed to discriminate explanations: Clinically comparable treatment groups and explicit indication adjustment Clinical consequence: Depends on event severity and available alternatives	Association with residual causal uncertainty
Inference scenario: Event appears immediately after treatment initiation Background-risk problem: Background events may cluster around treatment changes Cross-source interpretation: Temporal replication across sources is supportive, not decisive		Evidence needed to discriminate explanations: Clinically justified risk windows and pre-exposure trajectory assessment Clinical consequence: Short latency may increase urgency but not causal certainty	Time-structured safety concern
Inference scenario: Signal changes after coding or phenotype refinement Background-risk problem: Apparent risk depends on measurement definition Cross-source interpretation: Discordance may reflect measurement rather than biology		Evidence needed to discriminate explanations: Validated phenotype or manual/clinical adjudication Clinical consequence: Consequence depends on false-positive versus false-negative costs	Measurement-sensitive signal
Inference scenario: Similar association appears across heterogeneous databases Background-risk problem: Baseline risks and healthcare processes differ Cross-source interpretation: Agreement strengthens robustness only after shared-bias assessment		Evidence needed to discriminate explanations: Source-specific diagnostics and design-sensitive replication Clinical consequence: Greater stability can raise escalation priority	Cross-source supported association

Version control and updating rules

Each living update should be treated as a new evidential state with an auditable relationship to the previous version. Version records should distinguish four kinds of change: evidence additions, in which new studies or surveillance analyses are incorporated; evidence corrections, in which duplicate, erroneous, retracted, or materially revised information is addressed; method changes, including altered search, phenotype, analytical, or appraisal procedures; and interpretation changes, in which the same evidence is reclassified because a competing explanation has strengthened or weakened.

A change in signal magnitude does not automatically constitute a change in signal state. Additional reports may increase precision without resolving reporting bias, while a smaller study with better temporality or comparative design may materially change interpretation. Likewise, a database update can alter exposure prevalence, coding coverage, or follow-up without changing the biological hypothesis. Version control must therefore record *why* an interpretation moved, not merely whether an estimate or report count changed.

Table 2 formalizes this distinction as a boundary-condition register. Its purpose is to prevent routine updating from being mistaken for evidential progression.

Table 2. Boundary Conditions Governing Interpretation Changes Across Living Surveillance Versions

Source state and scale	Qualification endpoint	
Update case / Clinical seriousness	Reversibility or preventability / Cross-source status / Version-level interpretation	Escalation consequence
Update case: More reports accrue but source structure and biases are unchanged Clinical seriousness: May remain high	Reversibility or preventability: Context dependent Cross-source status: No independent corroboration Version-level interpretation: Evidence volume changed; inferential state may remain unchanged	Intensify evaluation if seriousness warrants
Update case: Independent longitudinal evidence supports the same event pattern Clinical seriousness: High or moderate	Reversibility or preventability: Context dependent Cross-source status: Concordant under a different data structure Version-level interpretation: Cross-source support strengthens association without proving causality	Priority for focused causal evaluation rises
Update case: Apparent signal attenuates after comparator or indication refinement Clinical seriousness: Any	Reversibility or preventability: Any Cross-source status: Earlier evidence becomes method-sensitive Version-level interpretation: Previous interpretation weakened by improved confounding control	Escalation may be reduced or redirected
Update case: Discordance is explained by different event or exposure definitions Clinical seriousness: Any	Reversibility or preventability: Any Cross-source status: Superficial disagreement Version-level interpretation: Evidence cannot yet be interpreted as biological inconsistency	Harmonize definitions before escalation decision
Update case: New evidence reveals a serious, preventable phenotype missed by earlier coding Clinical seriousness: High	Reversibility or preventability: Potentially preventable Cross-source status: Source-specific detection followed by targeted confirmation Version-level interpretation: Signal state changes because clinically relevant measurement improved	Rapid escalation may be warranted

RESULTS AND DISCUSSION

The reviewed literature supports a surveillance model in which different evidence sources contribute different inferential capabilities rather than successive degrees of inherent validity. Living-review methodology establishes the need for repeated, explicitly versioned reassessment. Pharmacovigilance methodology establishes that statistical signal detection is strongly dependent on source construction, analytical choice, event definition, and interpretive discipline. Together, these literatures indicate that a safety signal should be treated as a changing evidence state rather than a binary property of a drug-event pair.

Across source types, three recurring distinctions emerged. First, detection sensitivity and causal interpretability are separable: data that are effective for noticing unusual patterns may be weak for estimating comparative risk. Second, data richness does not eliminate design dependence: longitudinal structure introduces denominators and temporality but still requires valid exposure, outcome, comparator, and risk-window definitions. Third, **cross-source** concordance is conditional: agreement becomes more informative when sources differ in their principal biases, whereas concordance produced by common coding, indication, selection, or healthcare-process structures may add less independent evidence than its visual consistency implies.

Recent pharmacovigilance guidance explicitly emphasizes the complementary roles of spontaneous reporting and longitudinal healthcare databases in signal management rather than presenting either as a complete replacement for the other [14]. This provides the empirical-methodological basis for the living surveillance model developed in later sections.

Spontaneous reporting signals

Spontaneous reporting occupies a distinct position in the proposed surveillance architecture because it can reveal events before a denominator-bearing epidemiological study is feasible. Individual reports can preserve clinical

detail about temporal sequences, dechallenge or rechallenge, unusual syndromes, co-medications, product characteristics, and reporter suspicions. Aggregated across national and international systems, these observations can generate hypotheses across large and diverse exposure settings. Their value is therefore greatest when the surveillance task is to identify that an unexpected pattern may exist.

The same properties constrain the inference that can follow. Reporting databases generally do not provide complete counts of exposed individuals or systematically observed non-events. Reporting probability can change after publicity, regulatory communication, litigation, changes in product use, professional awareness, or the recognition of a suspected adverse reaction. Drug-event pairs may also reflect confounding by indication, concomitant treatment, disease progression, protopathic prescribing, differential diagnostic intensity, or differences in reporting quality. Consequently, disproportionality should be interpreted as a property of the reporting data under specified analytical conditions rather than as an estimate of causal risk.

A signal becomes more informative when its internal structure is clinically coherent and survives relevant sensitivity analyses, but these features do not erase the need for independent evaluation. Temporal plausibility, phenotype specificity, consistency across reporting strata, dose information where available, dechallenge or rechallenge patterns, and absence of obvious competing explanations can refine prioritization. Equally important are negative or destabilizing observations: loss of the signal after deduplication, restriction to one reporting environment, sensitivity to event coding, concentration around publicity, or disappearance after indication-sensitive analysis. Living surveillance should retain these weakening features rather than reporting only the strongest disproportionality estimate.

The role of spontaneous reporting is therefore neither to prove nor merely to “generate” signals in an undifferentiated sense. It identifies and characterizes candidate safety patterns whose next evidential requirement depends on what remains uncertain. A rare, distinctive and serious phenotype may require intensive clinical case assessment before comparative epidemiology. A common event with substantial background incidence may require denominator-bearing longitudinal analysis. A pattern strongly entangled with treatment indication may require an active comparator or other design strategy capable of addressing that confounding structure. In this sense, the output of spontaneous reporting is best understood as a structured set of hypotheses and uncertainty states that determine what evidence should be sought next.

Electronic health records and claims

EHR-based pharmacovigilance extends signal evaluation beyond reported suspicions by adding longitudinal exposure, clinical context, and denominator-bearing populations, although methods and validation remain heterogeneous [15]. Tree-based scanning can expand outcome discovery across structured and clinically enriched EHR phenotypes [16], while large-language-model approaches may extract safety information from narrative records but require explicit validation before surveillance use [17]. Sentinel experience illustrates active surveillance in pregnancy [18], and fit-for-purpose taxonomies show that some drug–outcome pairs are unsuitable for routine-data signal detection [19]. Active-comparator self-controlled designs can reduce selected confounding structures [20], but validity still depends on coherent time zero [21], exposure measurement [22], indication-sensitive comparison [23], recognition of database and design biases [24], quantitative sensitivity analysis [25], and appropriately chosen negative controls [26]. Sequential calibration [27], self-controlled methods [28], and explicit risk-window modeling [29] can strengthen inference without eliminating residual uncertainty.

Table 3 is needed because a longitudinal-data signal changes meaning as measurement, comparison, temporality, and bias control are strengthened.

Table 3. Translation Boundaries in EHR- and Claims-Based Safety Evaluation

Source state and scale	Transfer conditions	Qualification endpoint
State	Transfer boundary	Measurement requirement / Dominant bias test / Permitted endpoint
Coded event excess	Code to clinical event	Measurement requirement: Reproducible phenotype Dominant bias test: Coding and healthcare-contact bias Permitted endpoint: Candidate signal

Temporally localized excess	Timing to etiologic relevance	Measurement requirement: Exposure and risk-window alignment Dominant bias test: Reverse causation Permitted endpoint: Time-supported association
Comparator-persistent excess	Comparison to counterfactual	Measurement requirement: Clinically credible comparator Dominant bias test: Indication and severity Permitted endpoint: Comparative association
Sensitivity-stable excess	Model stability to causal interpretation	Measurement requirement: Alternative definitions and bias analyses Dominant bias test: Residual systematic error Permitted endpoint: Strengthened, not proven, safety concern

Registries

Registries may add disease severity, product detail, or clinically adjudicated outcomes unavailable in general claims. Validation frameworks emphasize that observational estimates require explicit treatment-assignment and sampling assumptions [30]. OMOP-based tools improve reproducible population estimation [31], but registry harmonization demonstrates that mapping itself can alter meaning and completeness [32].

These characteristics make registries a useful anchor for showing that richer observation does not remove inference boundaries, as shown in **figure 1**.

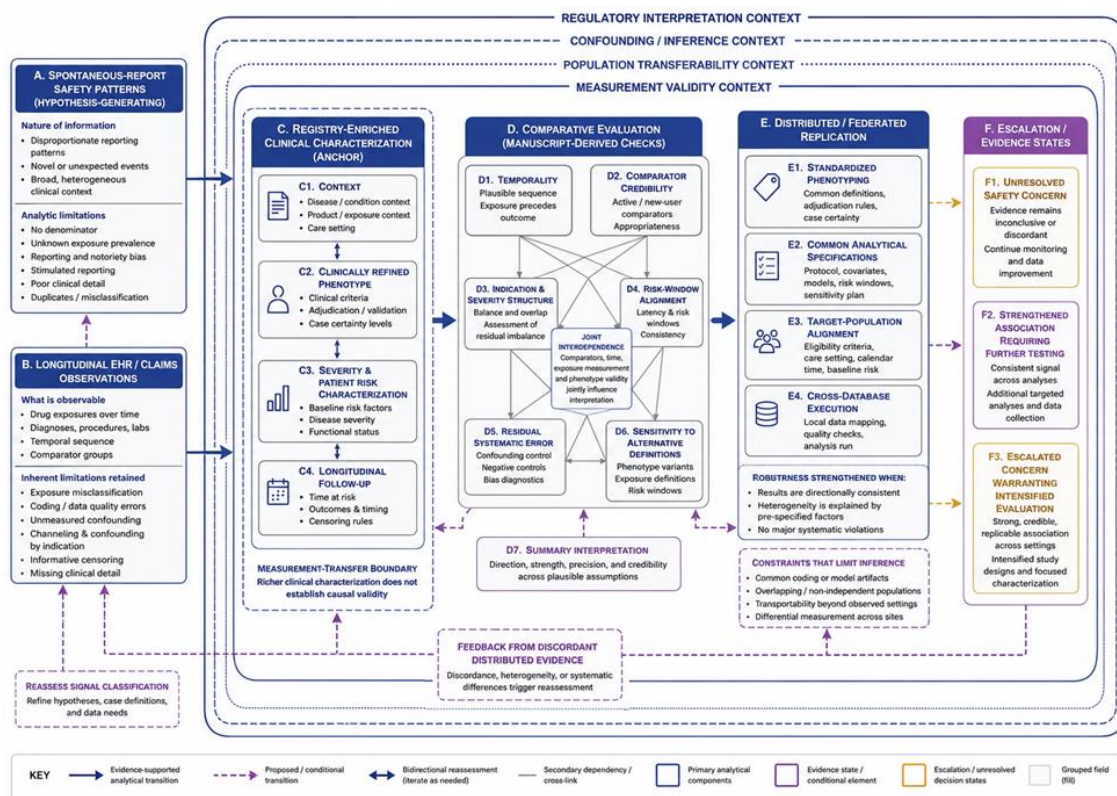


Figure 1. Registry-Anchored Signal Escalation Across Measurement, Comparison, and Network Boundaries

Distributed and federated networks

Distributed networks extend evaluation without requiring centralized patient-level data. DARWIN EU demonstrates standardized phenotyping across distributed analytics [33], while Sentinel work emphasizes executable transparency and reproducibility [34]. Multisite inference also depends on the target population because standardization can change precision and interpretation [35]. Contemporary network methodology distinguishes common-data-model and common-protocol architectures and their respective governance demands [36]. Linked EHR and claims demonstrations further show that federated infrastructure can make previously difficult safety questions analyzable while leaving missingness and residual confounding unresolved [37].

Cross-source signal concordance

Concordance is most informative when sources differ in their dominant biases yet address sufficiently comparable exposures, phenotypes, populations, and risk windows. Agreement produced by shared coding or indication structures may add less independent evidence; discordance may instead reveal different estimands or measurement regimes. **Table 4** therefore treats agreement as an inference problem rather than a vote count.

Table 4. Reconciliation States for Cross-Source Safety Evidence

Evidence configuration	Interpretive alternatives	Adjudication requirement
Pattern / Source relationship / Justified inference	uncertainty	Needed discriminator
Pattern: SRS positive, longitudinal null Source relationship: Different denominators and designs Justified inference: Unresolved signal	Reporting or outcome dilution	Phenotype and risk-window alignment
Pattern: SRS and longitudinal positive Source relationship: Partly independent Justified inference: Supported association	Shared indication bias possible	Bias-sensitive replication
Pattern: Registry positive, claims null Source relationship: Different clinical granularity Justified inference: Measurement-dependent discordance	Outcome misclassification	Validated common phenotype
Pattern: Multiple networks concordant Source relationship: Distributed replication Justified inference: Robust cross-source association	Shared model or coding bias	Site-specific diagnostics and external evidence

The synthesis rejects both source hierarchy and simple evidence accumulation. Signal maturity increases when later evidence resolves a specific uncertainty left by earlier evidence. Seriousness, preventability, and reversibility may accelerate investigation, but they do not convert weak inference into causal certainty.

A living design also changes how discordance should be handled. A negative result should not automatically close a signal when the second source fails to capture the suspected phenotype, exposure episode, or vulnerable population. Conversely, repeated positive findings should not be counted as independent confirmation when databases share coding systems, analytic specifications, or overlapping populations. The useful question is therefore whether new evidence changes the set of plausible explanations.

Criteria for signal escalation

The proposed escalation rule combines clinical consequence, phenotype credibility, temporal coherence, comparative credibility, bias resistance, and cross-source stability. Escalation is justified when the expected cost of delayed evaluation is high and the evidence is sufficiently specific to define the next test; no universal numerical threshold is proposed.

Living-surveillance architecture

Sentinel demonstrates scalable active surveillance infrastructure [38], while LEGEND illustrates standardized large-scale evidence generation across databases [39]. Building on these foundations, the proposed architecture records each signal as a versioned evidence state: detection, targeted characterization, comparative testing, cross-source reconciliation, and decision escalation. Downstream contradiction returns the signal to the earliest unresolved assumption rather than simply lowering a score.

Limitations

The framework inherits limitations of the underlying sources: incomplete reporting, coding and exposure error, unmeasured confounding, variable phenotype transportability, heterogeneous healthcare systems, and changing treatment practice. It has not been prospectively validated as an algorithm, and its proposed classifications should therefore be evaluated against real signal-management decisions.

CONCLUSION

Living safety surveillance is best treated as a revisable chain of source-specific inferences rather than a progression from weak to strong databases. Spontaneous reports remain valuable for detection; longitudinal data, registries,

and federated networks add distinct tests of temporality, comparison, measurement, and reproducibility. A defensible escalation decision depends on which uncertainty has actually been resolved, not on data volume or concordance alone.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Akl EA, Khabsa J, Iannizzi C, Piechotta V, Kahale LA, Barker JM, et al. Extension of the PRISMA 2020 statement for living systematic reviews (PRISMA-LSR): Checklist and explanation. *BMJ*. 2024;387:e079183. doi:10.1136/bmj-2024-079183
2. Butler AR, Hartmann-Boyce J, Livingstone-Banks J, Turner T, Lindson N. Optimizing process and methods for a living systematic review: 30 search updates and three review updates later. *J Clin Epidemiol*. 2024;166:111231. doi:10.1016/j.jclinepi.2023.111231
3. Iannizzi C, Akl EA, Anslinger E, Weibel S, Kahale LA, Aminat AM, et al. Methods and guidance on conducting, reporting, publishing, and appraising living systematic reviews: A scoping review. *Syst Rev*. 2023;12(1):238. doi:10.1186/s13643-023-02396-x
4. Heron L, Buitrago-Garcia D, Ipekci AM, Baumann R, Imeri H, Salanti G, et al. How to update a living systematic review and keep it alive during a pandemic: A practical guide. *Syst Rev*. 2023;12(1):156. doi:10.1186/s13643-023-02325-y
5. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. *J Clin Epidemiol*. 2017;91:23-30. doi:10.1016/j.jclinepi.2017.08.010
6. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. *J Clin Epidemiol*. 2017;91:31-37. doi:10.1016/j.jclinepi.2017.08.011
7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
8. Langan SM, Schmidt SAJ, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532. doi:10.1136/bmj.k3532
9. Brand JS, Gauffin O, Sartori D, Fusaroli M, Sköld H, Bergvall T, et al. VigiBase: Resource profile update with a summary of global patterns and trends in adverse event reports for medicines and vaccines. *Drug Saf*. 2026;49(6):613-29. doi:10.1007/s40264-025-01642-6
10. 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
11. 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):e5768. doi:10.1002/pds.5768
12. Fusaroli M, Mitchell J, Rudolph A, Rocca E, Fusaroli R. Causal inference tools for pharmacovigilance: Using causal graphs to identify and address biases in disproportionality analysis. *Drug Saf*. 2026;49(4):449-69. doi:10.1007/s40264-025-01628-4
13. 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
14. 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
15. 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
 16. Russo M, Sreedhara SK, Smith J, Davis SE, Maro JC, Deramus T, et al. Electronic health record-enhanced signal detection using tree-based scan statistic methods. *Am J Epidemiol.* 2026;195(1):178-87. doi:10.1093/aje/kwaf199
 17. Matheny ME, Yang J, Smith JC, Walsh CG, Al-Garadi MA, Davis SE, et al. Enhancing postmarketing surveillance of medical products with large language models. *JAMA Netw Open.* 2024;7(8):e2428276. doi:10.1001/jamanetworkopen.2024.28276
 18. Lyons JG, Shinde MU, Maro JC, Petrone A, Cosgrove A, Kempner ME, et al. Use of the Sentinel System to examine medical product use and outcomes during pregnancy. *Drug Saf.* 2024;47(10):931-40. doi:10.1007/s40264-024-01447-z
 19. Coste A, Wong AYS, Warren-Gash C, Matthewman J, Bate A, Douglas IJ. Implementation of a taxonomy-based framework for the selection of appropriate drugs and outcomes for real-world data signal detection studies. *Drug Saf.* 2024;47(2):183-92. doi:10.1007/s40264-023-01382-5
 20. Coste A, Wong AYS, Haguinet F, Bate A, Douglas IJ. Performance of the self-controlled case series with active comparators for drug safety signal detection using the French administrative healthcare database (SNDS). *Pharmacoepidemiol Drug Saf.* 2025;34(11):e70224. doi:10.1002/pds.70224
 21. Her QL, Rouette J, Young JC, Webster-Clark M, Tazare J. Core concepts in pharmacoepidemiology: New-user designs. *Pharmacoepidemiol Drug Saf.* 2024;33(12):e70048. doi:10.1002/pds.70048
 22. Thai TN, Winterstein AG. Core concepts in pharmacoepidemiology: Measurement of medication exposure in routinely collected healthcare data for causal inference studies in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2024;33(3):e5683. doi:10.1002/pds.5683
 23. Sendor R, Stürmer T. Core concepts in pharmacoepidemiology: Confounding by indication and the role of active comparators. *Pharmacoepidemiol Drug Saf.* 2022;31(3):261-9. doi:10.1002/pds.5407
 24. Acton EK, Willis AW, Hennessy S. Core concepts in pharmacoepidemiology: Key biases arising in pharmacoepidemiologic studies. *Pharmacoepidemiol Drug Saf.* 2023;32(1):9-18. doi:10.1002/pds.5547
 25. Brown JP, Hunnicutt JN, Ali MS, Bhaskaran K, Cole A, Langan SM, et al. Core concepts in pharmacoepidemiology: Quantitative bias analysis. *Pharmacoepidemiol Drug Saf.* 2024;33(10):e70026. doi:10.1002/pds.70026
 26. Zafari Z, Park JE, Shah CH, dosReis S, Gorman EF, Hua W, et al. The state of use and utility of negative controls in pharmacoepidemiologic studies. *Am J Epidemiol.* 2024;193(3):426-53. doi:10.1093/aje/kwad201
 27. Schuemie MJ, Bu F, Nishimura A, Suchard MA. Adjusting for both sequential testing and systematic error in safety surveillance using observational data: Empirical calibration and MaxSPRT. *Stat Med.* 2023;42(5):619-31. doi:10.1002/sim.9631
 28. Bots SH, Brown J, Wong AYS, Martin I, Douglas I, Klungel OH, et al. Core concepts: Self-controlled designs in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2025;34(1):e70071. doi:10.1002/pds.70071
 29. Rippin G, Salmasi S, Sanz H, Largent J. Core concepts in pharmacoepidemiology: Time-to-event analysis approaches in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2024;33(11):e5886. doi:10.1002/pds.5886
 30. Shen L, Visser E, van Erning F, Geleijnse G, Kaptein M. A two-step framework for validating causal effect estimates. *Pharmacoepidemiol Drug Saf.* 2024;33(9):e5873. doi:10.1002/pds.5873
 31. Raventós B, Català M, Du M, Guo Y, Black A, Inberg G, et al. IncidencePrevalence: An R package to calculate population-level incidence rates and prevalence using the OMOP common data model. *Pharmacoepidemiol Drug Saf.* 2024;33(1):e5717. doi:10.1002/pds.5717
 32. Trinh NTH, Houghtaling J, Bernal FLM, Hayati S, Maglanoc LA, Lupattelli A, et al. Harmonizing Norwegian registries onto OMOP common data model: Mapping challenges and opportunities for pregnancy and COVID-19 research. *Int J Med Inform.* 2024;191:105602. doi:10.1016/j.ijmedinf.2024.105602
 33. Dernie F, Corby G, Robinson A, Bezer J, Mercade-Besora N, Griffier R, et al. Standardised and reproducible phenotyping using distributed analytics and tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). *Pharmacoepidemiol Drug Saf.* 2024;33(11):e70042. doi:10.1002/pds.70042

34. Rai A, Maro JC, Dutcher S, Bright P, Toh S. Transparency, reproducibility, and replicability of pharmacoepidemiology studies in a distributed network environment. *Pharmacoepidemiol Drug Saf.* 2024;33(6):e5820. doi:10.1002/pds.5820
35. Webster-Clark M, Fillion KB, Platt RW. Standardizing to specific target populations in distributed networks and multisite pharmacoepidemiologic studies. *Am J Epidemiol.* 2024;193(7):1031-9. doi:10.1093/aje/kwae015
36. Haber R, Webster-Clark M, Pratt N, Barclay N, Li X, Maro JC, et al. Core concepts in pharmacoepidemiology: Multi-database distributed data networks. *Pharmacoepidemiol Drug Saf.* 2025;34(7):e70177. doi:10.1002/pds.70177
37. Desai RJ, Weberpals J, Pillai H, Ajao A, Desibhatla M, Hawrusik R, et al. Strengthening inferential studies in the FDA Sentinel initiative: Results from a methodological demonstration project. *NPJ Digit Med.* 2026;9(1):52. doi:10.1038/s41746-025-02234-5
38. Cocoros NM, Fuller CC, Adimadhyam S, Ball R, Brown JS, Dal Pan GJ, et al. A COVID-19-ready public health surveillance system: The Food and Drug Administration's Sentinel System. *Pharmacoepidemiol Drug Saf.* 2021;30(7):827-37. doi:10.1002/pds.5240
39. Schuemie MJ, Ryan PB, Pratt N, Chen R, You SC, Krumholz HM, et al. Principles of large-scale evidence generation and evaluation across a network of databases (LEGEND). *J Am Med Inform Assoc.* 2020;27(8):1331-7. doi:10.1093/jamia/ocaa103