Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2025, 14(4):66-77
https://doi.org/10.51847/ZurxT6wisX

Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Modern Pharmacovigilance across Spontaneous Reports, Real-World Data, Machine Learning, and Active Surveillance: An Umbrella Review of Contemporary Evidence

Olivia Bennett[1]*, Harry Collins[2], Jack Foster[1], Ethan Wright[3]

[1]Department of Modern Pharmacovigilance and Umbrella Review, Faculty of Pharmacy, University of Nottingham, Nottingham, United Kingdom.
[2]Department of Spontaneous Reports and Real-World Data, Faculty of Pharmacy, University of Southampton, Southampton, United Kingdom.
[3]Department of Machine Learning and Active Surveillance, Faculty of Pharmacy, Newcastle University, Newcastle, United Kingdom.

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

ABSTRACT

Modern pharmacovigilance increasingly combines spontaneous reports with electronic health records, administrative claims, registries, distributed data networks, active monitoring, natural-language processing, and machine-learning methods. Yet the evidentiary roles of these approaches are often conflated: signal generation may be treated as causal assessment, computational performance as operational utility, and implementation activity as evidence of improved patient safety. This umbrella review evaluates contemporary review-level evidence across major surveillance approaches and examines methodological quality, primary-study overlap, convergence, contradiction, transferability, and implementation maturity. Systematic reviews, meta-analyses, scoping reviews, and mapping reviews were treated as the principal evidence units, while methodological and implementation papers were used only to define appraisal, overlap, reporting, or deployment boundaries. Review-level quality was assessed through critical domains appropriate to each review type rather than through unsupported numerical scores. Primary-study overlap was managed through planned citation-matrix and domain-specific interpretation, without assuming that agreement among overlapping reviews represented independent confirmation. The evidence indicates that spontaneous reports remain central to early signal generation and narrative characterization, whereas routinely collected healthcare data can provide denominators, longitudinal context, and comparative analyses when exposure, outcome, confounding, and data-provenance limitations are addressed. Machine learning and natural-language processing may improve extraction, prioritization, and evidence integration, but external validation, benchmark consistency, model maintenance, and demonstrated workflow benefit remain uneven. Active surveillance, registries, and distributed networks offer more structured ascertainment, although implementation maturity varies by setting, population, and governance infrastructure. Across domains, evidence converges on complementarity rather than replacement. Conflicts commonly reflect different surveillance tasks, data structures, validation standards, and decision contexts. Integrated pharmacovigilance should therefore preserve source provenance and maintain explicit boundaries among detection, validation, causal assessment, implementation, and decision use.

Key words: *Pharmacovigilance, Drug safety, Safety signal, Causal inference, Adverse event, Real-world data*
Received: 16 August 2025; **Revised:** 14 November 2025; **Accepted:** 18 November 2025

INTRODUCTION

Pharmacovigilance encompasses the detection, assessment, understanding, communication, and prevention of adverse effects and other medicine-related problems. Its contemporary evidence base extends beyond spontaneous

adverse-event reports to electronic health records, administrative claims, registries, distributed networks, social-media data, active monitoring, and computational systems for extracting or prioritizing potential safety information. This expansion has increased the volume and diversity of available evidence but has also complicated its interpretation. Surveillance outputs may represent reports, statistical associations, model predictions, corroborated signals, or causal judgments, and these categories should not be treated as interchangeable. Transparent systematic-review reporting is therefore essential when evidence from heterogeneous surveillance platforms is combined [1].

The prevailing framing of modern pharmacovigilance is often technology-centred. New data sources or algorithms may be described as improvements because they increase processing speed, identify more candidate events, or produce higher internal classification performance. Such gains can be useful, but they do not establish that a method detects clinically meaningful harms earlier, transports across institutions, improves causal assessment, supports regulatory decisions, or benefits patients. Review-level appraisal must consequently examine whether eligibility, search, extraction, risk-of-bias assessment, synthesis, and interpretation were methodologically adequate rather than relying on publication venue or technical novelty as proxies for credibility. Critical-domain appraisal provides a structured basis for making these distinctions [2].

Another limitation is the fragmentation of the literature into parallel methodological communities. Reviews of spontaneous-reporting databases emphasize disproportionality and reporting behaviour; reviews of electronic healthcare data focus on phenotyping, confounding, and longitudinal design; computational reviews emphasize extraction, classification, knowledge representation, or signal ranking; and active-surveillance reviews examine structured ascertainment within specific populations or medicines. Without reproducible searching across these domains, an evidence synthesis may privilege one surveillance paradigm, miss relevant terminology, or obscure important differences between analytical performance and pharmaceutical decision use. Complete reporting of search sources, strategies, limits, and supplementary retrieval methods is therefore necessary for evaluating the breadth and reproducibility of an umbrella review [3].

This article uses an umbrella-review design to synthesize systematic reviews, meta-analyses, scoping reviews, and mapping reviews concerning spontaneous reports, real-world healthcare data, machine learning, natural-language processing, registries, distributed networks, and active surveillance. It examines methodological quality, primary-study overlap, concordance, contradiction, certainty, transferability, and implementation evidence. Scoping reviews are retained where they provide transparent mapping of emerging or heterogeneous fields, but their descriptive coverage is not interpreted as proof of effectiveness or readiness; explicit reporting of objectives, eligibility, charting, and evidentiary boundaries remains necessary [4]. The contribution is a review-level comparison of what each surveillance approach can support, where conclusions remain conditional, and how complementary evidence could be integrated without collapsing signal generation into causal confirmation or technical capability into clinical or regulatory readiness.

Umbrella-review protocol and eligibility criteria

The review question was structured around six objectives: to identify review-level evidence across major pharmacovigilance approaches; assess the methodological reliability of the included reviews; determine how overlapping primary studies affect the apparent evidence base; identify areas of convergence and contradiction; distinguish implementation evidence from analytical performance evidence; and define unresolved methodological and translational gaps. The unit of inclusion was the review rather than the individual primary study. Eligible evidence comprised systematic reviews, meta-analyses, scoping reviews, and systematic mapping reviews that examined a defined pharmacovigilance data source, surveillance method, validation problem, implementation context, or evidence-integration question. Methodological and operational companion papers were eligible only when they directly informed review appraisal, search reporting, overlap management, registry use, distributed analysis, implementation boundaries, or model maintenance.

Eligible reviews had to describe a reproducible or assessable evidence-selection process, state their review question or objective, identify the surveillance source or computational task under examination, and provide findings that could be mapped to a specific claim, table field, figure element, contradiction, limitation, or research priority. Relevant domains included spontaneous reporting, patient reporting, safety-signal detection, causal assessment, electronic health records, claims data, clinical narratives, natural-language processing, machine learning, knowledge engineering, knowledge graphs, registries, distributed networks, active surveillance, social media, subgroup harm, and model drift. Reviews were not required to evaluate every domain, but their conclusions

had to remain interpretable within a defined population, dataset, surveillance task, validation target, or implementation setting.

Non-peer-reviewed reports, conference-only publications, preprints without final journal publication, protocols without completed results, regulatory webpages, software documentation, database descriptions, policy documents, and promotional commentaries were excluded as evidence units. Reviews were also excluded when pharmacovigilance was incidental to their primary question, when methods or findings could not be mapped to the review taxonomy, or when their conclusions required extrapolation beyond the population, product, data source, or validation context examined. Primary studies were not incorporated into the principal synthesis and were not used to create de novo pooled estimates. Where a methodological or implementation companion paper was retained, its role was identified separately so that it could not enlarge the apparent volume of review-level evidence. Eligibility establishes relevance to the umbrella-review question; it does not confer causal validity or implementation readiness on the underlying findings.

Search strategy and selection of systematic reviews

Searches were designed for MEDLINE/PubMed, Embase, Scopus, Web of Science Core Collection, the Cochrane Database of Systematic Reviews, and Epistemonikos. Search concepts combined pharmacovigilance and drug-safety terminology with terms for adverse events, adverse drug reactions, safety signals, spontaneous reporting, real-world data, electronic health records, claims, registries, distributed networks, active surveillance, social media, natural-language processing, machine learning, artificial intelligence, knowledge engineering, knowledge graphs, causal assessment, external validation, implementation, model drift, and subgroup harm. Review-design terms included systematic review, meta-analysis, scoping review, mapping review, overview, and umbrella review. Search reporting was structured to preserve the database, platform, strategy, limits, and supplementary retrieval route required for reproducibility [3].

Selection was conducted at the review-report level. Records were first reconciled through DOI, title, author, journal, and publication-lineage information. Protocols were linked to completed reviews where possible, and preprint and journal versions were treated as one publication lineage rather than independent evidence. Corrected and original versions were examined to avoid duplicate inclusion. Titles and abstracts were assessed against the review-level eligibility criteria, followed by full-text evaluation of the review question, search methods, included evidence types, pharmacovigilance relevance, extractable findings, and decision context. Exclusion at full text was based on predefined methodological or topical criteria rather than disagreement with a review's conclusions. Data charting captured bibliographic details, review type, information sources, search coverage, primary-study scope, population, medicine or product scope, data source, surveillance method, analytical task, comparator, validation target, principal findings, methodological limitations, review-reported certainty, implementation evidence, and transferability conditions. Extraction also recorded whether findings concerned detection, prediction, validation, causal assessment, communication, workflow implementation, or patient outcomes. This distinction prevented retrospective model performance from being interpreted as prospective surveillance benefit and prevented reporting activity from being equated with better risk management. Review findings were synthesized narratively because the included reviews addressed heterogeneous questions and did not provide a common basis for de novo quantitative pooling.

Methodological-quality and certainty assessment

Methodological quality was assessed according to review design. Systematic reviews and meta-analyses were evaluated using AMSTAR 2 critical domains, including protocol development, search adequacy, justification of exclusions, risk-of-bias assessment, synthesis methods, interpretation of bias, and investigation of publication bias where applicable [2]. The appraisal was domain-based rather than numerical. A review was not considered reliable solely because it reported an extensive search or appeared in a high-ranked journal, and an overall score was not created by assigning equal weight to items with different implications. Appraisal judgments were intended to qualify confidence in the review process, not to validate the surveillance methods or technologies examined.

Scoping and mapping reviews were assessed through criteria suited to their descriptive objectives. Evaluation focused on clarity of the review question, alignment between objectives and eligibility criteria, comprehensiveness and reproducibility of searching, transparency of selection and charting, treatment of source limitations, and separation of mapped activity from evidence of effectiveness. PRISMA-ScR informed reporting assessment, but it was not used as a numerical quality instrument [4]. Reviews that catalogued algorithms, databases, or

applications without external validation or implementation outcomes were treated as evidence of research activity and conceptual coverage rather than proof of pharmaceutical utility.

Certainty was interpreted at two levels. First, any certainty or confidence judgment formally reported by an included review was retained with its original scope and method. Second, the umbrella synthesis considered methodological quality, consistency, directness, validation context, primary-study overlap, and transferability when describing the confidence that could reasonably be placed in a conclusion. No unreported evidence grade was retroactively assigned. Convergence across reviews increased confidence only when the reviews addressed sufficiently similar questions and were not merely reproducing the same primary studies. Contradictory conclusions were examined for differences in populations, surveillance tasks, data provenance, outcome definitions, analytical choices, comparators, and validation targets. Methodological appraisal therefore constrained interpretation but did not convert observational associations, disproportionality signals, or computational outputs into confirmed causal or clinical conclusions.

Management of overlapping primary studies

Overlap occurs when multiple included reviews synthesize the same primary studies. Without explicit management, repeated evidence can appear to represent independent confirmation, inflate the perceived breadth of a domain, and disproportionately amplify frequently reviewed datasets. The synthesis therefore treated review agreement as potentially dependent until primary-study membership and review scope were considered [5].

Corrected covered area can summarize the concentration of shared primary studies, but its interpretation requires a sufficiently complete study-by-review citation matrix. A single value cannot capture differences in review questions, study quality, populations, outcomes, analytical methods, or publication dates. No overlap estimate was inferred when the necessary membership information was unavailable [6].

Static citation matrices and graphical displays were used conceptually to identify reviews drawing on common evidence and to distinguish broad overlap from concentration around a small group of influential studies. These displays support transparency but do not determine whether the shared findings are valid, biased, or transferable [7].

Pairwise interpretation was prioritized because overlap may be high within one surveillance domain and low across others. Spontaneous-reporting, electronic-health-data, computational, and active-surveillance reviews were therefore interpreted within their respective questions before cross-domain conclusions were drawn [8].

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for characteristics, methodological quality, and primary-study overlap of included reviews on modern pharmacovigilance across spontaneous reports, real-world data, machine learning.

Table 1. Characteristics, methodological quality, and primary-study overlap of included reviews on Modern Pharmacovigilance across Spontaneous Reports, Real-World Data, Machine Learning.

Evidence domain or review construct	Review question	Eligible evidence type	Key extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Spontaneous reporting	What information and signals arise from voluntary reports?	Systematic or scoping reviews	Reporter, event narrative, signal method, comparator	Transparent eligibility and signal definition	Characterize hypothesis generation	Underreporting and absent denominator	Signal does not establish incidence or causation
EHRs and claims	How are routinely collected data used for safety detection?	Systematic or scoping reviews	Exposure, outcome, design, confounding, validation	Reproducible phenotyping and design description	Assess longitudinal surveillance	Misclassification and residual confounding	Association remains design-dependent
NLP and machine learning	Which computational tasks are evaluated?	Systematic, scoping, or mapping reviews	Corpus, labels, task, comparator, external validation	Clear reference standard and validation setting	Compare extraction and prioritization	Dataset and metric heterogeneity	Performance is not implementation benefit

Active surveillance and registries	How is structured safety information collected?	Systematic or scoping reviews	Population, follow-up, ascertainment, governance	Defined population and outcome capture	Assess proactive surveillance	Variable scale and standardization	Active collection does not remove bias
Methodological quality	How credible is each review process?	Review-method papers and appraisal guidance	Search, protocol, exclusions, bias assessment, synthesis	Critical-domain assessment	Qualify confidence	Incomplete reporting	Appraisal does not validate the underlying method
Primary-study overlap	How independent are apparently concordant reviews?	Citation matrices and overlap methods	Review–study membership, scope, chronology	Complete and interpretable mapping	Prevent double counting	Membership often incompletely reported	Agreement cannot be assumed independent

Spontaneous-reporting evidence

Spontaneous-reporting systems remain central to pharmacovigilance because they can capture unexpected, rare, or clinically distinctive events across broad populations. Patient reports may add treatment experiences, symptom descriptions, and functional consequences that are incompletely represented in professional reports, although reporting completeness and clinical verification vary [9].

Signal-detection methods applied to spontaneous reports include disproportionality, Bayesian approaches, subgroup analyses, temporal methods, and computational extensions. Their usefulness depends on the safety question, database structure, coding conventions, duplicate management, masking, reporting patterns, and comparator definition. No method is uniformly superior across all signal-detection tasks [10].

The interpretability of disproportionality analyses is weakened when database versions, case definitions, deduplication rules, statistical thresholds, sensitivity analyses, or analytical choices are incompletely reported. Such deficiencies restrict reproducibility and can make apparently similar analyses difficult to compare [11].

A safety signal is information suggesting a potentially new or changed association that warrants further investigation. It is not equivalent to a confirmed adverse drug reaction. Signal communication may draw on spontaneous reports, mechanistic evidence, observational studies, trials, and expert assessment, but the evidentiary basis and causal strength vary [12].

Electronic health records and claims data

Routinely collected healthcare data can support longitudinal exposure assessment, denominator-based analyses, comparator selection, and evaluation of outcomes occurring in clinical practice. Nevertheless, systematic-review evidence shows substantial variation in signal-identification designs, data structures, analytical assumptions, and validation strategies, with no context-independent best method [13].

Electronic health records provide diagnoses, laboratory values, prescriptions, procedures, and clinical context, but data are generated for care rather than surveillance. Missingness, fragmented care, coding incentives, changing documentation, incomplete medication use, and uncertain outcome validity can produce systematic error even when datasets are large [14].

Clinical narratives contain symptoms, temporal relationships, treatment changes, and clinician reasoning that may not appear in structured fields. Reviews of adverse-event extraction show heterogeneous annotation schemes, vocabularies, corpora, outcomes, and performance measures, limiting direct comparison and pooled interpretation [15].

Natural-language processing and machine learning may improve identification of candidate events in electronic records, but performance can depend on local documentation and reference standards. External, temporal, and cross-institutional validation remain necessary before transportability or routine surveillance utility can be inferred [16].

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for review-level findings across principal evidence sources and surveillance approaches in modern pharmacovigilance across spontaneous reports, real-world data, machine learning..

Table 2. Review-level findings across principal evidence sources and surveillance approaches in Modern Pharmacovigilance across Spontaneous Reports, Real-World Data, Machine Learning.

Method, platform, or evidence category	Representative application	Reported strength	Validation context	Generalization concern	Contradictory evidence	Evidence gap	Review interpretation
Spontaneous reports	Early signal generation	Broad reach and sensitivity to unusual narratives [9]	Case review and later corroboration	Reporting behaviour varies	More reports may reflect publicity rather than risk	Denominators and exposure data	Foundational for hypotheses, not incidence
Disproportionality analysis	Drug–event prioritization	Scalable database screening [10]	Known signals or expert review	Thresholds and coding differ	Methods may rank signals differently	Transparent benchmarking	Task-specific prioritization tool
EHR surveillance	Longitudinal case finding	Clinical context and repeated observations [14]	Chart review or external outcome definitions	Institutional documentation differs	Richer data may introduce more missingness	Multisite validation	Useful when provenance and phenotypes are explicit
Claims analysis	Population-based association assessment	Defined enrollment and utilization history [13]	Epidemiological comparator designs	Limited clinical detail and coding error	Administrative efficiency may reduce phenotype specificity	Outcome confirmation	Supports evaluation, not automatic causality
Clinical-note NLP	Extraction of adverse-event information	Access to unstructured narratives [15]	Annotated clinical corpora	Labels and language vary by site	High internal performance may not transport	Prospective workflow evaluation	Promising extraction support
EHR-based ML	Candidate-event classification	Flexible modelling of complex features [16]	Internal or external test data	Dataset shift and label leakage	Performance varies across institutions	Temporal and external validation	Readiness remains context-dependent

Figure 1 presents the original conceptual synthesis for umbrella evidence map of modern pharmacovigilance data sources, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

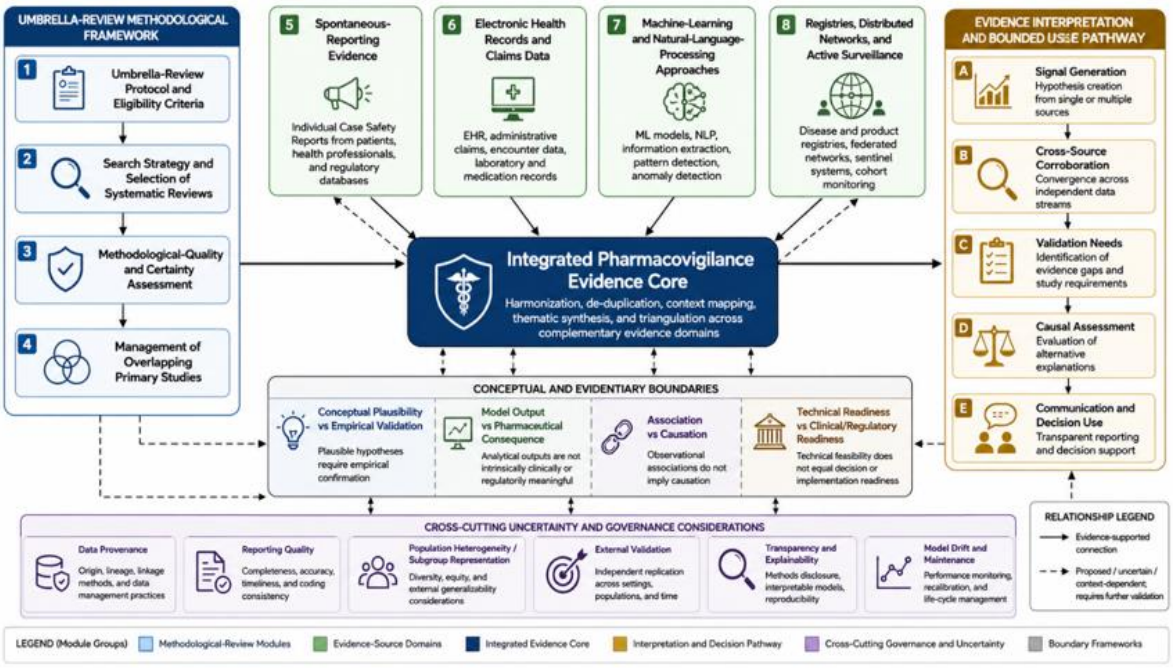

Figure 1. Umbrella Evidence Map of Modern Pharmacovigilance Data Sources

The figure is an original conceptual synthesis developed for "Modern Pharmacovigilance across Spontaneous Reports, Real-World Data, Machine Learning, and Active Surveillance: An Umbrella Review of Contemporary Evidence." Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, clinical recommendation, regulatory determination, or deployment-ready system.

Machine-learning and natural-language-processing approaches

Machine learning has been applied to adverse-event classification, relation extraction, duplicate detection, case prioritization, and signal ranking. These applications may support human review by organizing complex data, but their validity depends on representative labels, appropriate comparators, leakage control, and alignment between the modelling task and the pharmacovigilance question [17].

Knowledge-engineering approaches provide ontologies, semantic mappings, rule-based reasoning, and linked representations that can connect drugs, events, mechanisms, populations, and evidence sources. Review-level mapping indicates considerable methodological diversity, however, and semantic interoperability does not itself establish the clinical validity of an inferred relationship [18].

Contemporary artificial-intelligence applications span intake, coding, triage, literature screening, signal management, and report processing. Operational adoption is uneven, and evidence is concentrated in selected workflow components rather than prospectively validated end-to-end systems. Efficiency claims therefore require separate evaluation from accuracy, safety, and decision consequences [19].

Knowledge graphs may support evidence integration, provenance tracking, and relationship exploration. Their pharmacovigilance maturity remains constrained by incomplete data, uncertain edge semantics, heterogeneous ontologies, limited shared benchmarks, and sparse external comparisons. A graph-generated association remains a hypothesis until corroborated through appropriate evidence and causal assessment [20].

Registries, distributed networks, and active surveillance

Intensive monitoring studies actively collect treatment and outcome information from defined populations. They may improve follow-up and temporal characterization compared with passive reporting, but review evidence shows substantial variation in recruitment, outcome ascertainment, duration, comparator use, and analytical purpose [21].

Active surveillance strategies used for antiviral agents include cohorts, sentinel systems, targeted monitoring, database analyses, and structured clinical follow-up. Their differing populations and ascertainment intensities limit simple comparison and prevent the evidence from supporting one universally preferred strategy [22].

Pregnancy pharmacovigilance illustrates the dependence of surveillance maturity on infrastructure. Reviews from low- and middle-income settings identify fragmented exposure data, incomplete outcome ascertainment, inconsistent definitions, loss to follow-up, and restricted linkage across maternal and infant records [23].

Registries can provide disease-specific longitudinal information, but registry existence does not guarantee pharmacovigilance readiness. Adverse-event variables, medicine exposure, seriousness, causality, follow-up, and reporting pathways may be absent or inconsistently standardized, limiting secondary safety use [24].

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for convergence, contradictions, certainty, and transferability across review-level findings.

Table 3. Convergence, contradictions, certainty, and transferability across review-level findings.

Cross-study pattern or configuration	Context or boundary condition	Mechanism or explanatory factor	Observed or reported outcome	Rival explanation	Certainty or rigour concern	Implication	Research priority
Spontaneous reports remain central	Early detection of unusual events	Broad, low-barrier reporting	Rich signal hypotheses [9]	Stimulated reporting	Missing denominators	Preserve as hypothesis source	Improve reporting quality and linkage
EHRs add longitudinal context	Complete local care capture	Structured and narrative records	More contextual case identification [14]	Documentation intensity	Outcome misclassification	Use validated phenotypes	Multisite replication

Active monitoring improves structure	Defined population and follow-up	Planned ascertainment	Better temporal characterization [21]	More intensive observation	Design heterogeneity	Interpret by protocol	Common outcome definitions
Strategy performance varies	Different medicines and settings	Task-specific data generation	No universally superior method [22]	Incomparable objectives	Limited head-to-head evidence	Avoid universal rankings	Prospective comparative evaluation
Pregnancy evidence is fragmented	Resource-constrained systems	Weak linkage and follow-up	Uncertain exposure–outcome capture [23]	Selection and referral patterns	Limited transferability	Strengthen infrastructure	Maternal–infant linkage
Registry readiness is uneven	Registries built for non-safety purposes	Missing PV variables	Restricted secondary safety use [24]	Local governance differences	Survey-based evidence	Assess fitness before use	Minimum PV data standards

Comparative performance and implementation evidence

Social-media surveillance may capture patient language, emerging concerns, and experiences that do not enter conventional systems. Reviews nevertheless report uncertain representativeness, ambiguous exposure, duplicate content, unverifiable clinical details, and limited evidence that social media can function as a standalone pharmacovigilance source [25].

Evaluation against consequential safety actions remains sparse. Studies differ in platforms, sampling, event definitions, time windows, and reference standards, and few demonstrate consistent anticipation of major warnings, labelling changes, or withdrawals. Negative findings may reflect both methodological weakness and the rarity of suitable validation events [26].

Interventions directed at professionals and patients can increase spontaneous adverse-reaction reporting, particularly through education, reminders, feedback, or simplified reporting routes. However, increased report volume does not necessarily improve report quality, signal specificity, sustainability, causal assessment, or patient outcomes [27].

Industry use of artificial intelligence indicates operational interest in automation, triage, coding, and signal workflows. Such adoption is implementation activity rather than proof of comparative effectiveness. Evaluation should include human oversight, error consequences, explainability, workflow integration, maintenance, and whether the technology improves safety decisions rather than only processing speed [28].

Evidence certainty, contradictions, and gaps

Across reviews, spontaneous reporting remains a foundational source of early safety information, but its limitations are consistent: underreporting, selective reporting, missing exposure denominators, variable completeness, duplicates, and susceptibility to publicity. These weaknesses constrain quantification without eliminating the source's value for hypothesis generation [29].

Transferability is especially uncertain where pharmacovigilance infrastructure, workforce, reporting culture, data systems, and regulatory capacity differ. Evidence from low- and middle-income settings shows that methods developed in data-rich environments may not operate similarly where medicine-use data, clinical records, and follow-up systems are fragmented [30].

Broader patient-safety reviews of artificial intelligence report substantial activity in prediction and detection but less evidence of prospective implementation, sustained performance, or patient-level impact. This imbalance means technical feasibility is better established than clinical consequence [31].

Pharmacovigilance-specific reviews similarly identify limited external validation, inconsistent benchmarks, explainability concerns, heterogeneous outcome definitions, and sparse evaluation in routine workflows. Apparent contradictions often arise because studies address different tasks, datasets, comparators, or validation targets rather than because one approach is universally effective or ineffective [32].

Table 4 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evidence gaps, implementation limitations, and priorities for integrated practice and research.

Table 4. Evidence gaps, implementation limitations, and priorities for integrated practice and research.

Evidence-maturity domain	Current evidence position	Methodological weakness	Translation barrier	Needed validation	Reporting priority	Decision implication	No-overclaiming boundary
Spontaneous-report transparency	Widely used but inconsistently documented [11]	Variable analytical reporting	Poor reproducibility	Independent replication	Database, coding, deduplication, thresholds	Treat outputs as signals	No incidence or causality claim
EHR and claims surveillance	Longitudinal evidence available [13]	Confounding and phenotype error	Site-specific data structures	Multidatabase validation	Provenance and design choices	Support signal evaluation	Association remains conditional
Clinical NLP	Strong extraction potential [15]	Heterogeneous corpora and labels	Limited transportability	Cross-site and temporal testing	Reference-standard construction	Assist case identification	Internal accuracy is insufficient
Pharmacovigilance AI	Growing task-level evidence [19]	Inconsistent benchmarks	Workflow and governance uncertainty	Prospective human–AI evaluation	Errors, overrides, maintenance	Use for bounded support	No autonomous readiness claim
Knowledge graphs	Emerging integration capability [20]	Uncertain semantics and completeness	Interoperability	External comparator studies	Provenance and edge meaning	Support hypothesis organization	Graph links are not causal proof
Registry and active monitoring	Structured collection is feasible [23]	Inconsistent variables and follow-up	Resource and linkage constraints	Fit-for-purpose validation	Exposure, outcome, and loss to follow-up	Context-specific use	No assumption of complete ascertainment
Social media	Adjunctive information source [26]	Weak reference standards	Representativeness and verification	Event-based external validation	Sampling and duplicate handling	Supplement established systems	No standalone replacement claim
Reporting interventions	Reporting may increase [27]	Heterogeneous intervention studies	Sustainability	Longer-term outcome evaluation	Report quality and downstream action	Improve participation cautiously	Volume is not patient benefit

Limitations of the umbrella review

Methodological-quality judgments depend partly on the appraisal framework used. AMSTAR 2 and ROBIS address related but nonidentical domains, and differences in scope or interpretation can produce differing judgments. Appraisal findings should therefore be reported transparently rather than treated as immutable classifications [33].

AMSTAR 2 has demonstrated validity but only moderate reliability under some conditions. This supports explicit decision rules, independent assessment, and resolution of disagreements. It also cautions against interpreting small differences in appraisal as precise measures of review credibility [34].

Overview methods are frequently underreported when reviews overlap, disagree, or contain problematic evidence. Incomplete primary-study membership and inconsistent review scope limited the extent to which dependence could be quantified, so overlap was interpreted qualitatively whenever reliable calculation was not possible [35]. The evidence domains were heterogeneous in review question, population, product, data source, comparator, and outcome. Review publication also lags primary research, particularly for rapidly changing computational approaches. The synthesis may therefore underrepresent very recent technical developments while appropriately avoiding selective incorporation of unsynthesized primary studies.

Priorities for integrated pharmacovigilance

Distributed pharmacoepidemiology networks can support reproducible analysis across institutions while data remain locally governed. Their value depends on common protocols, harmonized definitions, transparent analytic execution, quality checks, and interpretation of site-level heterogeneity rather than simple aggregation [36].

Patient registries could contribute more effectively to medicines evaluation when exposure, outcomes, timing, confounders, follow-up, and governance are fit for the intended safety question. Interoperability and common data elements are priorities, but registry suitability must be demonstrated rather than assumed [37].

Machine-learning surveillance requires maintenance after initial validation. Changes in populations, clinical practice, coding, documentation, technology, or outcome prevalence can alter input distributions and model performance. Monitoring for data drift, recalibration needs, and emerging subgroup errors should therefore accompany any prospective use [38].

An integrated approach should preserve provenance while combining complementary evidence. Spontaneous reports can initiate hypotheses; electronic healthcare data can support contextual and comparative evaluation; NLP can recover narrative evidence; knowledge structures can organize relationships; and active systems can improve targeted ascertainment. This proposed synthesis does not remove the need for external validation, causal assessment, human review, governance, or decision-specific evaluation [20].

CONCLUSION

Modern pharmacovigilance is best understood as a complementary evidence system rather than a competition among isolated data sources or technologies. Spontaneous reports remain valuable for early hypothesis generation; real-world healthcare data add longitudinal and comparative context; computational methods can support extraction, prioritization, and integration; and registries, distributed networks, and active surveillance can strengthen structured ascertainment. Their evidentiary maturity is nevertheless task-specific. Agreement among reviews does not necessarily represent independent confirmation, analytical performance does not establish pharmaceutical utility, and implementation activity does not demonstrate patient benefit. An integrated future therefore requires transparent provenance, explicit management of overlap, common validation standards, attention to subgroup and setting-specific harms, prospective workflow evaluation, and continuing monitoring after implementation. The present synthesis defines these relationships and boundaries but does not constitute a validated surveillance architecture, clinical recommendation, regulatory determination, or deployment-ready system.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. 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. doi:10.1136/bmj.n71
2. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358. doi:10.1136/bmj.j4008
3. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z
4. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med. 2018;169(7):467-73. doi:10.7326/M18-0850
5. Lunny C, Pieper D, Thabet P, Kanji S. Managing overlap of primary study results across systematic reviews: Practical considerations for authors of overviews of reviews. BMC Med Res Methodol. 2021;21(1):140. doi:10.1186/s12874-021-01269-y
6. Hennessy EA, Johnson BT. Examining overlap of included studies in meta-reviews: Guidance for using the corrected covered area index. Res Synth Methods. 2020;11(1):134-45. doi:10.1002/jrsm.1390
7. Bougioukas KI, Vounzoulaki E, Mantsiou CD, Savvides ED, Karakosta C, Diakonidis T, et al. Methods for depicting overlap in overviews of systematic reviews: An introduction to static tabular and graphical displays. J Clin Epidemiol. 2021;132:34-45. doi:10.1016/j.jclinepi.2020.12.004

8. Pérez-Bracchiglione J, Meza N, Bangdiwala SI, Niño de Guzmán E, Urrútia G, Bonfill X, et al. Graphical representation of overlap for overviews: GROOVE tool. Res Synth Methods. 2022;13(3):381-8. doi:10.1002/jrsm.1557
9. Inácio P, Cavaco A, Airaksinen M. The value of patient reporting to the pharmacovigilance system: A systematic review. Br J Clin Pharmacol. 2017;83(2):227-46. doi:10.1111/bcp.13098
10. 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
11. Khouri C, Revol B, Lepelley M, Mouffak A, Bernardeau C, Salvo F, et al. A meta-epidemiological study found lack of transparency and poor reporting of disproportionality analyses for signal detection in pharmacovigilance databases. J Clin Epidemiol. 2021;139:191-8. doi:10.1016/j.jclinepi.2021.07.014
12. Sartori D, Aronson JK, Norén GN, Onakpoya IJ. Signals of adverse drug reactions communicated by pharmacovigilance stakeholders: A scoping review of the global literature. Drug Saf. 2023;46(2):109-20. doi:10.1007/s40264-022-01258-0
13. 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
14. Davis SE, Zabotka 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
15. Modi S, Kasmiran KA, Mohd Sharef N, Sharum MY. Extracting adverse drug events from clinical notes: A systematic review of approaches used. J Biomed Inform. 2024;151:104603. doi:10.1016/j.jbi.2024.104603
16. 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
17. Lee CY, Chen YPP. Machine learning on adverse drug reactions for pharmacovigilance. Drug Discov Today. 2019;24(7):1332-43. doi:10.1016/j.drudis.2019.03.003
18. Natsiavas P, Malousi A, Bousquet C, Jaulent MC, Koutkias V. Computational advances in drug safety: Systematic and mapping review of knowledge engineering based approaches. Front Pharmacol. 2019;10:415. doi:10.3389/fphar.2019.00415
19. Warner J, Prada Jardim A, Albera C. Artificial intelligence: Applications in pharmacovigilance signal management. Pharm Med. 2025;39(3):183-98. doi:10.1007/s40290-025-00561-2
20. Hauben M, Rafi M, Abdelaziz I, Hassanzadeh O. Knowledge graphs in pharmacovigilance: A scoping review. Clin Ther. 2024;46(7):544-54. doi:10.1016/j.clinthera.2024.06.003
21. Torre C, Cary M, Borges FC, Ferreira PS, Alarcão J, Leufkens HGM, et al. Intensive monitoring studies for assessing medicines: A systematic review. Front Med (Lausanne). 2019;6:147. doi:10.3389/fmed.2019.00147
22. Ferreira-da-Silva R, Reis-Pardal J, Pinto M, Monteiro-Soares M, Sousa-Pinto B, Morato M, et al. A comparison of active pharmacovigilance strategies used to monitor adverse events to antiviral agents: A systematic review. Drug Saf. 2024;47(12):1203-24. doi:10.1007/s40264-024-01470-0
23. Shafi J, Virk MK, Kalk E, Carlucci JG, Chepkemoi A, Bernard C, et al. Pharmacovigilance in pregnancy studies, exposures and outcomes ascertainment, and findings from low- and middle-income countries: A scoping review. Drug Saf. 2024;47(10):957-90. doi:10.1007/s40264-024-01445-1
24. Plueschke K, Jonker C, Strassmann V, Kurz X. Collection of data on adverse events related to medicinal products: A survey among registries in the ENCePP Resources Database. Drug Saf. 2022;45(7):747-54. doi:10.1007/s40264-022-01188-x
25. 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
26. Lee JY, Lee YS, Kim DH, Lee HS, Yang BR, Kim MG. The use of social media in detecting drug safety-related new black box warnings, labeling changes, or withdrawals: Scoping review. JMIR Public Health Surveill. 2021;7(6). doi:10.2196/30137

27. Paudyal V, Al-Hamid A, Bowen M, Hadi MA, Hasan SS, Jalal Z, et al. Interventions to improve spontaneous adverse drug reaction reporting by healthcare professionals and patients: Systematic review and meta-analysis. Expert Opin Drug Saf. 2020;19(9):1173-91. doi:10.1080/14740338.2020.1807003
28. Kassekert R, Grabowski N, Lorenz D, Schaffer C, Kempf D, Roy P, et al. Industry perspective on artificial intelligence/machine learning in pharmacovigilance. Drug Saf. 2022;45(5):439-48. doi:10.1007/s40264-022-01164-5
29. Alomar M, Tawfiq AM, Hassan N, Palaian S. Post marketing surveillance of suspected adverse drug reactions through spontaneous reporting: Current status, challenges and the future. Ther Adv Drug Saf. 2020;11:2042098620938595. doi:10.1177/2042098620938595
30. Kiguba R, Olsson S, Waitt C. Pharmacovigilance in low- and middle-income countries: A review with particular focus on Africa. Br J Clin Pharmacol. 2023;89(2):491-509. doi:10.1111/bcp.15193
31. Bates DW, Levine D, Syrowatka A, Kuznetsova M, Craig KJT, Rui A, et al. The potential of artificial intelligence to improve patient safety: A scoping review. NPJ Digit Med. 2021;4(1):54. doi:10.1038/s41746-021-00423-6
32. Salas M, Petracek J, Yalamanchili P, Aimer O, Kasthuril D, Dhingra S, et al. The use of artificial intelligence in pharmacovigilance: A systematic review of the literature. Pharm Med. 2022;36(5):295-306. doi:10.1007/s40290-022-00441-z
33. Pieper D, Puljak L, González-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. J Clin Epidemiol. 2019;108:26-33. doi:10.1016/j.jclinepi.2018.12.004
34. Lorenz RC, Matthias K, Pieper D, Wegewitz U, Morche J, Nocon M, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. J Clin Epidemiol. 2019;114:133-40. doi:10.1016/j.jclinepi.2019.05.028
35. Lunny C, Brennan SE, Reid J, McDonald S, McKenzie JE. Overviews of reviews incompletely report methods for handling overlapping, discordant, and problematic data. J Clin Epidemiol. 2020;118:69-85. doi:10.1016/j.jclinepi.2019.09.025
36. Platt RW, Platt R, Brown JS, Henry DA, Klungel OH, Suissa S. How pharmacoepidemiology networks can manage distributed analyses to improve replicability and transparency and minimize bias. Pharmacoepidemiol Drug Saf. 2020;29 Suppl 1:3-7. doi:10.1002/pds.4722
37. 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
38. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: Implications and potential remedies. Br J Radiol. 2023;96(1150):20220878. doi:10.1259/bjr.20220878