Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2026, 15(2):94-103
https://doi.org/10.51847/isC5CoKSj7

Original Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

The Global Safety Knowledge Graph Linking Case Reports, Clinical Records, Scientific Literature, Regulatory Actions, and Product Changes

Amira Khalil[1]*, Sara Fathy[1], Mahmoud Adel[2], Hany Nabil[3], Rania Mostafa[4]

[1]Department of Global Safety Knowledge Graphs and Integration, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt.
[2]Department of Case Reports and Clinical Records, Faculty of Pharmacy, Cairo University, Cairo, Egypt.
[3]Department of Scientific Literature and Regulatory Actions, Faculty of Pharmacy, Mansoura University, Mansoura, Egypt.
[4]Department of Product Changes and Pharmacovigilance, Faculty of Pharmacy, Helwan University, Cairo, Egypt.

***Email:** amira.khalil@alexu.edu.eg*

ABSTRACT

Medicine-safety evidence is distributed across spontaneous case reports, clinical records, registries, observational studies, scientific publications, regulatory assessments, product labels, and manufacturing or product histories. These sources differ in structure, terminology, provenance, temporal coverage, jurisdiction, and evidentiary meaning, limiting the reconstruction of how an emerging concern develops into a safety signal, causal assessment, regulatory response, or product change. This original knowledge-graph architecture article proposes a Global Safety Knowledge Graph that represents source records, normalized concepts, evidence claims, inferred relationships, assessment states, regulatory actions, and product versions as distinguishable but connected objects. The architecture combines an evidence-source layer, semantic and identity-resolution services, a temporal-provenance graph core, analytical and conflict-management functions, and governed review and feedback processes. Product versions, manufacturing changes, and exposure contexts are represented separately from active ingredients so that product-quality or traceability hypotheses are not misattributed to an entire substance or class. Conflicting evidence is preserved through support, challenge, supersession, and unresolved-status relations rather than being overwritten during updating. Temporal representation distinguishes when an event occurred, when evidence entered the system, when an assessment changed, and when a regulatory or product action became effective. Validation is proposed across syntactic, semantic, identity-resolution, provenance, temporal, analytical, transportability, workflow, and decision-impact levels. The architecture may improve evidence traceability and support testable safety hypotheses, but it does not convert association into causation, eliminate source bias, establish regulatory acceptability, or justify automated clinical or regulatory decision-making. Its principal contribution is a unified, evidence-bounded model for connecting global medicine-safety knowledge while preserving uncertainty, provenance, temporal history, and human responsibility.

Key words: *Knowledge graph, Pharmacovigilance, Drug safety, Safety signal, Causal inference, Adverse event*
Received: 01 March 2026; **Revised:** 28 May 2026; **Accepted:** 03 June 2026

INTRODUCTION

Pharmacovigilance encompasses activities concerned with detecting, assessing, understanding, and preventing adverse effects and other medicine-related problems. Within that scope, an adverse event is an observation following exposure, an adverse drug reaction implies at least a suspected relationship, a safety signal is

information warranting further investigation, and causal assessment evaluates whether an observed relationship may be attributable to a medicine. These constructs should not be treated as interchangeable states of evidence [1].

Modern medicine-safety systems operate across a product lifecycle rather than through isolated signal-detection exercises. Reports are received, coded, assessed, combined with other evidence, communicated among institutions, and sometimes translated into risk-minimization measures, labeling changes, restrictions, or additional monitoring. These activities involve distinct professional, scientific, and regulatory responsibilities, and the progression from observation to action is neither automatic nor uniform across settings [2].

Global spontaneous-reporting resources provide extensive longitudinal evidence concerning suspected medicine-related harms, but their scale does not remove limitations arising from under-reporting, stimulated reporting, missing clinical context, variable terminology, duplicate submission, and differences among reporting systems. Report counts and graph connectivity therefore cannot be interpreted directly as incidence, comparative risk, or causal strength [3].

This article develops an original, non-empirical architecture for linking case reports, clinical records, scientific literature, regulatory actions, and product changes within a provenance-preserving temporal knowledge graph. The contribution is not a new safety database, predictive model, causal-estimation method, or regulatory standard. It is a proposed conceptual structure for keeping heterogeneous evidence connected without erasing its origin, uncertainty, jurisdiction, temporal validity, or decision status.

Fragmentation of global medicine-safety evidence

Spontaneous reports, electronic health records, registries, claims databases, observational studies, literature, and regulatory documents produce complementary forms of safety evidence. Spontaneous reports may reveal unusual clinical patterns, whereas longitudinal healthcare data can provide denominators, comparators, exposure histories, and covariate information. These sources answer different questions and contain different biases; combining them does not make them methodologically interchangeable [4].

Fragmentation also arises from incompatible identifiers, terminology systems, schemas, access rules, and data-governance arrangements. A medicine may be represented by an active substance, brand, authorization number, manufacturer, presentation, or local code, while an outcome may be encoded as a diagnosis, laboratory result, clinical narrative, adverse-event term, or study endpoint. Without maintained mappings and source-level provenance, apparent integration may conceal semantic disagreement rather than resolve it [5].

Organizational capability introduces a further layer of fragmentation. Pharmacovigilance institutions differ in staffing, technical infrastructure, reporting practices, analytical capacity, legal authority, and procedures for signal assessment or communication. Consequently, the absence of linked evidence may reflect limited surveillance capacity, incomplete access, delayed processing, or jurisdiction-specific practice rather than the absence of a medicine-safety concern [6].

The central problem is therefore not simply that safety information is stored in separate databases. Fragmentation is simultaneously technical, institutional, and procedural, so no single database or reporting reform can integrate the complete safety-evidence lifecycle [2, 4, 6]. The proposed graph interprets fragmentation as discontinuities among evidence objects, product identities, clinical contexts, time, provenance, and decisions. It may make these discontinuities inspectable, but it cannot eliminate under-reporting, correct all source biases, or equalize global surveillance capacity.

Proposed global safety knowledge graph

Knowledge graphs have been applied to pharmacovigilance tasks including adverse-event representation, literature integration, relationship discovery, and signal exploration, but existing implementations remain heterogeneous in their source coverage, semantics, validation, and intended uses [7]. The proposed Global Safety Knowledge Graph extends this work by connecting evidence generation to assessment, regulatory action, product history, and continuing surveillance while preserving distinctions among these processes. **Figure 1** presents the original conceptual synthesis for global safety knowledge-graph architecture, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

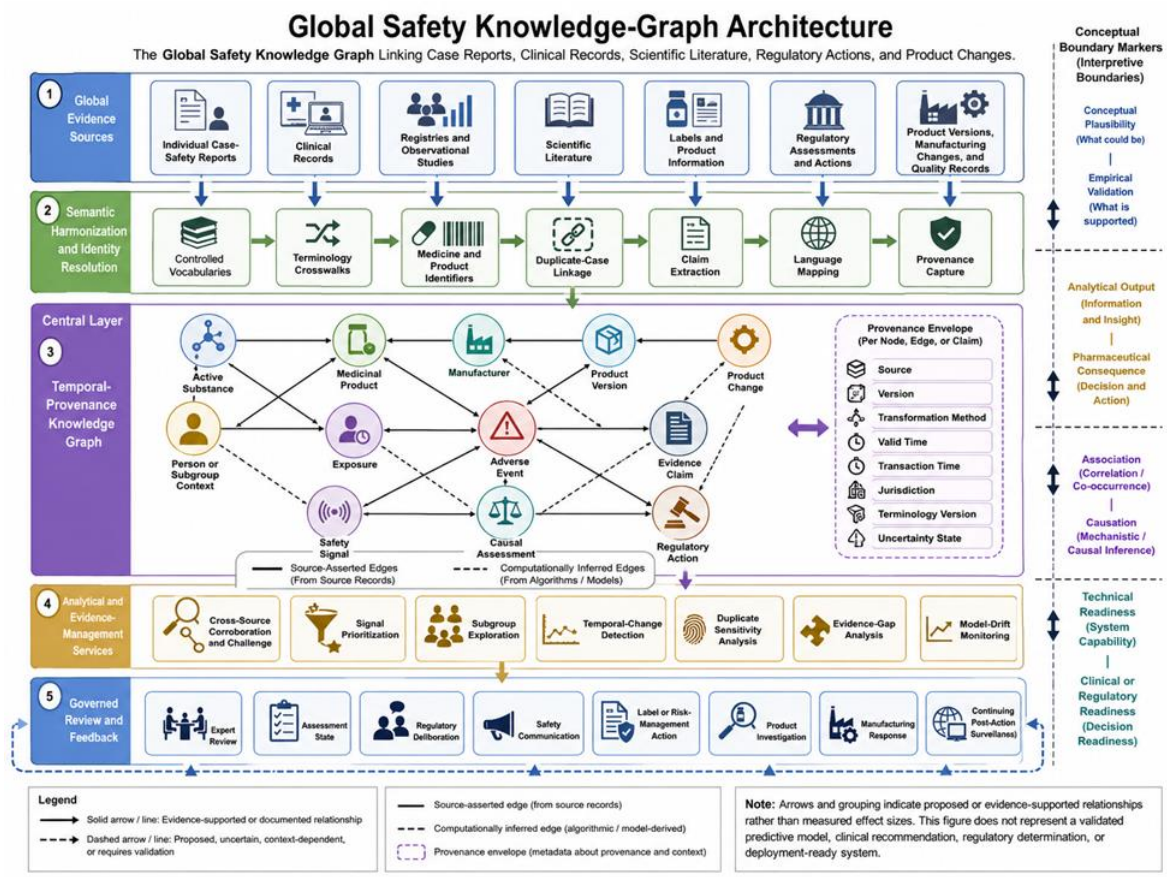

Figure 1. Global Safety Knowledge-Graph Architecture.

The figure is an original conceptual synthesis developed for "The Global Safety Knowledge Graph Linking Case Reports, Clinical Records, Scientific Literature, Regulatory Actions, and Product Changes." 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.

The graph core is intended to organize relationships rather than pronounce conclusions. Biomedical graph studies demonstrate that multi-relational structures can generate candidate adverse drug-reaction hypotheses and permit evaluation against clinical data, but an inferred link remains dependent on the construction of the graph and the quality of external validation [8]. Accordingly, computationally inferred edges would be visually and semantically separated from relationships explicitly asserted in source evidence.

Knowledge-graph construction also requires deliberate choices about entity definitions, relation types, ontology alignment, integration methods, negative evidence, and the downstream question being addressed [9]. A graph designed to retrieve supporting literature may differ substantially from one intended to assess duplicate reports, compare regulatory histories, or explore product-specific safety patterns. The architecture therefore does not assume that one universal graph representation can support every pharmacovigilance task.

The proposed architecture contains five connected layers. An evidence-source layer retains original records; a harmonization layer maps terminology and identities; a temporal-provenance graph core stores versioned assertions; an analytical layer supports prioritization, conflict detection, subgroup exploration, and drift monitoring; and a governed action layer records expert assessments, communications, regulatory decisions, product investigations, and subsequent surveillance. These layers are an original synthesis rather than an empirically established pipeline.

The graph would distinguish a reported relationship from a normalized relationship, an inferred relationship from a source assertion, and a signal hypothesis from a causal or regulatory conclusion. Its purpose is to improve traceability, permit cross-source questions, expose disagreement, and support reproducible evaluation. It does not confer truth on an edge, eliminate confounding, validate a safety signal, or authorize autonomous clinical, manufacturing, or regulatory action.

Entity, relation, and provenance model

A global graph must preserve the meanings that source systems assign to their data. Biomedical datasets differ in their definitions, collection processes, granularity, coding practices, and implicit assumptions; merely converting them to a common technical format does not establish semantic equivalence [10]. The proposed model therefore treats every normalized object as a mapped representation linked to, rather than substituted for, its original source record.

Controlled vocabularies are necessary for comparing adverse-reaction concepts across institutions, but terminology mapping must remain versioned and reversible. Multicenter pharmacovigilance work has shown that structured signal dictionaries and vocabulary-based mappings can improve consistency across heterogeneous electronic health records [11]. The proposed graph would record the source term, normalized concept, mapping method, terminology version, and any unresolved ambiguity for each transformation.

Provenance is defined here as the information required to reconstruct where an assertion originated, how it was transformed, when it was valid, and who or what produced it. FAIR design principles support the use of explicit, measurable metadata for findability, accessibility, interoperability, and reuse [12]. These properties are necessary for graph auditability, but they do not by themselves establish the accuracy, completeness, representativeness, or causal validity of the underlying evidence.

The original assertion model contains four linked objects: a source object, a normalized entity, an evidence assertion, and an assessment state. Each assertion would record its source version, creator or reporter type, extraction or transformation method, jurisdiction, valid time, transaction time, terminology version, and uncertainty status. Relations would include reports, records, supports, challenges, maps-to, refers-to, temporally-precedes, supersedes, assessed-as, acted-upon-by, and associated-with-product-version.

Identity and equivalence remain separate propositions. Two reports may concern the same case without being confirmed duplicates; two product codes may share an active substance without identifying the same manufacturer or formulation; and two outcome terms may be semantically related without denoting clinically equivalent events. The graph may organize these possibilities and preserve their supporting evidence, but identity, causal relevance, and decision significance require independent, intended-use-specific validation.

Individual case-safety reports

Individual case-safety reports should enter the graph as versioned evidence objects rather than definitive records of a causal medicine–event relationship. Duplicate identification requires probabilistic comparison of structured fields and narrative content because repeated submissions may lack shared identifiers or contain partially inconsistent information [13].

Case-level exploration can highlight terms, clinical features, or reporting patterns that warrant expert attention. Such computational prioritization is most defensible when reviewers can inspect the contributing reports, distinguishing features, and transformations rather than receiving an unexplained score [14].

Disproportionality measures and graph associations are vulnerable to confounding, reporting bias, stimulated reporting, competition among events, and other causal distortions. They may prioritize hypotheses but cannot independently establish that exposure caused an outcome [15].

The proposed ICSR subgraph therefore separates report version, suspected product, exposure, coded event, narrative evidence, reporter type, seriousness, outcome, follow-up status, duplicate cluster, and assessment state. An ICSR edge records a reported or inferred association; it does not estimate incidence, comparative risk, or individual causality.

Clinical records, registries, and observational studies

Longitudinal clinical data can contextualize case-report hypotheses by providing exposure histories, comparison groups, temporal ordering, and covariates. Their regulatory usefulness nevertheless depends on whether the source captures the population, treatment, outcome, and confounders required for the specific question [16].

Distributed surveillance can reduce unnecessary centralization by applying common definitions and analytic procedures across locally governed data. Adaptation of a shared data model must preserve local coding, coverage, and healthcare-system context rather than assuming that structural conformity establishes semantic equivalence [17].

OHDSI-compatible vocabularies and data structures illustrate how heterogeneous pharmacovigilance resources may be organized for reproducible analysis [18]. The proposed graph would link encounter, exposure episode,

phenotype definition, cohort, protocol, estimate, and analytic execution while preserving their source-specific limitations.

Clinical records are therefore neither a universal reference standard nor merely confirmatory evidence. A missing association may reflect limited power, incomplete outcome capture, exposure misclassification, or incompatible populations, while a detected association may remain confounded.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, components, relationships, and intended uses in the global safety knowledge graph linking case reports, clinical records, scientific.

Table 1. Definitions, components, relationships, and intended uses in The Global Safety Knowledge Graph Linking Case Reports, Clinical Records, Scientific.

Component or scientific dimension	Problem addressed	Inputs or determinants	Proposed mechanism or relationship	Expected contribution	Evidence required	Failure or uncertainty risk	Boundary statement
ICSR object	Reports may be duplicated, incomplete, or revised	Structured fields, narratives, follow-up versions	**Proposed:** versioned report nodes and probabilistic duplicate clusters	Case-level traceability	Adjudicated linkage evidence [13]	False merging or separation	A report is not incidence or causality
Case-review support	Large report sets impede inspection	Coded terms, narratives, distinguishing features	**Proposed:** explainable feature highlighting	More transparent prioritization	Expert-review evaluation [14]	Automation bias	Prioritization does not replace review
Causal boundary	Association may reflect reporting or confounding structures	Disproportionality, covariates, causal assumptions	**Proposed:** separate association, signal, and causal-assessment states	Reduced semantic overclaiming	Bias analyses [15]	Causal interpretation of non-causal edges	Graph linkage does not establish effect
Clinical-data context	Reports often lack denominators and longitudinal context	EHRs, registries, claims, study protocols	**Proposed:** linked exposure, phenotype, cohort, and estimate nodes	Cross-source assessment	Question-specific data-fitness evidence [16]	Misclassification and residual confounding	Clinical data are not an unbiased standard
Distributed surveillance	Data cannot always be centralized	Local datasets, common models, governed queries	**Proposed:** federated graph mappings	Reproducible multisite analysis	Cross-site semantic testing [17]	Structurally similar but semantically different data	Common format does not prove equivalence
Standardized analytics	Heterogeneous resources impede reproducibility	Shared vocabularies and analytic definitions	**Proposed:** versioned OHDSI-compatible mappings	Reusable safety queries	Replication across sources [18]	Vocabulary and implementation drift	Reproducibility remains intended-use specific
Provenance envelope	Transformations may obscure source meaning	Source, method, version, time, jurisdiction	**Proposed original construct:** provenance attached to every assertion	Auditability and reversibility	Reconstruction audits	Missing or incorrect lineage	Provenance supports inspection, not truth
Conflict state	Sources may disagree	Supporting, challenging, and superseding claims	**Proposed original construct:** preserve	Transparent uncertainty	Curated conflict cases	Premature reconciliation	Conflict is not resolved by aggregation alone

	disagreement as graph relations

Literature, labels, and regulatory actions

Scientific publications contribute claims rather than undifferentiated documents. The graph should link a claim to its population, design, comparison, outcome, limitations, and supporting evidence, while distinguishing publication from regulatory assessment. Signals pass through review and governance before they may produce communication or action [19].

Regulatory decisions are temporally and jurisdictionally bounded. Comparative evidence indicates that the timing of safety-related labeling changes can differ among regulatory regions, requiring separate nodes for review date, decision date, effective date, jurisdiction, and label version [20].

Postapproval labeling changes are important lifecycle events, but they do not reveal the entire evidentiary or causal pathway underlying a decision [21]. The graph should therefore avoid interpreting a label revision as automatic confirmation of a particular computational signal.

Proposed relations include *supports*, *challenges*, *cites*, *assessed-by*, *effective-in-jurisdiction*, *communicated-through*, and *supersedes-label-version*. Regulatory actions would remain authoritative within their legal context without being converted into universal causal or class-wide conclusions.

Product versions, manufacturing changes, and exposure context

Safety patterns may sometimes indicate suspected product-quality problems, making product identity and distribution context relevant to pharmacovigilance investigation [22]. Such patterns generate hypotheses; confirmation requires independent quality, supply-chain, and exposure evidence.

Multisource biologic environments illustrate why active-substance identity alone may be insufficient. Manufacturer, brand, product version, and other traceability fields can determine whether a safety observation is attributable to a specific product rather than the broader substance class [23].

Product information may also represent immunogenicity and related safety attributes inconsistently [24]. A graph should preserve the exact document version, terminology, assay context, and jurisdiction instead of treating differently framed statements as equivalent facts.

The proposed product-lineage model links substance, authorization, manufacturer, brand, formulation, strength, presentation, product version, manufacturing change, distribution period, and batch or lot when legitimately available. Missing traceability must remain explicit rather than being replaced by an inferred identity presented as fact.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for testable propositions, evidence requirements, and validation criteria for the global safety knowledge graph linking case reports, clinical records, scientific.

Table 2. Testable propositions, evidence requirements, and validation criteria for The Global Safety Knowledge Graph Linking Case Reports, Clinical Records, Scientific.

Architecture layer or process stage	Core function	Information or material flow	Interaction with other components	Validation criterion	Potential failure mode	Human or experimental responsibility	Readiness boundary
Evidence ingestion	Preserve source records	Source to versioned object	Feeds harmonization	Lossless reconstruction	Truncated or altered evidence	Data steward audit	Ingestion success is not scientific validity
Identity resolution	Link cases and products	Candidate identity links	Supports ICSR and product lineage	Adjudicated precision and recall	False merge or separation	Expert adjudication [13]	Uncertain identity remains provisional
Semantic mapping	Normalize concepts	Source term to mapped concept	Enables cross-source queries	Meaning preserved across sites	Incorrect equivalence	Terminology specialists [11]	Mapping does not prove clinical equivalence

Clinical-data linkage	Contextualize signals	Exposure and outcome records to study objects	Tests case-derived hypotheses	Reproducible phenotype and cohort execution	Confounding or incomplete capture	Epidemiologists and clinicians [16]	Association is not causation
Regulatory linkage	Connect assessment to action	Signal to review, communication, and label version	Interacts with jurisdiction and time	Reconstructable decision history	Chronology mistaken for causality	Regulatory reviewers [19]	Graph does not determine action
Product lineage	Preserve product specificity	Product and manufacturing information to exposure context	Links quality investigation to reports	Verified product attribution	Ingredient-level misattribution	Quality and supply-chain experts [22]	Traceability hypothesis is not defect confirmation
Temporal updating	Retain changing evidence	New, corrected, challenged, or superseded assertions	Updates all graph layers	Complete version history	Silent overwriting	Governance and audit teams	Latest evidence is not automatically best
Drift monitoring	Detect changing data generation	Coding, reporting, population, and product changes	Triggers reassessment	Prespecified change detection and review	Undetected degradation	Model owners and independent validators	Technical monitoring does not authorize deployment

Evidence conflict, uncertainty, and temporal updating

Temporal methods can identify unexpected increases in reporting frequency, but such changes remain hypotheses requiring clinical and reporting-context review [25]. Increased reporting may reflect a real safety problem, stimulated attention, market expansion, coding changes, or altered reporting behavior.

Changes in mandatory reporting can modify the composition of incoming safety data and consequently alter apparent signal behavior [26]. The graph should therefore represent reporting rules, source coverage, and system changes as contextual events rather than background metadata.

Apparent signal changes must be interpreted against causal bias, temporal reporting patterns, and changes in reporting rules [15, 25, 26]. Conflicting evidence should be preserved through *supports*, *challenges*, *does-not-replicate*, *uses-incompatible-definition*, and *supersedes* relations.

Proposed workflow states include observed, normalized, linked, prioritized, corroborated, challenged, under assessment, action-associated, superseded, reopened, and unresolved. These states describe evidence handling, not numerical certainty, causal probability, or regulatory readiness.

Validation, interoperability, and governance

Distributed safety systems require shared data structures, governed analyses, quality procedures, and auditable coordination across participating organizations [27]. The graph should therefore support federated operation where sensitive source data remain locally controlled while approved assertions or query results are exchanged.

Interoperability requires more than a common file format. Maintained registries of standards, repositories, vocabularies, and policies are needed to document how resources relate and change over time [28].

Validation is proposed across syntactic ingestion, semantic mapping, identity resolution, provenance completeness, temporal reconstruction, analytical performance, transportability, workflow usability, and prospective decision impact. Success at one level does not imply success at the next.

Governance should specify data stewardship, access control, terminology maintenance, model ownership, change control, audit rights, conflict escalation, human override, and retirement criteria. No analytical output should reach a higher-consequence use unless the validation and oversight requirements for that specific use have been met.

Limitations and implementation priorities

Large healthcare databases do not become valid for every safety question through scale alone. Data validity must be assessed against the intended exposure, outcome, population, comparison, and analytic purpose [29].

Models and mappings may also degrade when reporting practices, coding systems, populations, products, or clinical workflows change. Dataset shift therefore requires active monitoring and renewed assessment of transportability [30].

Figure 2 presents the original conceptual synthesis for temporal linkage of emerging evidence to regulatory and product changes, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

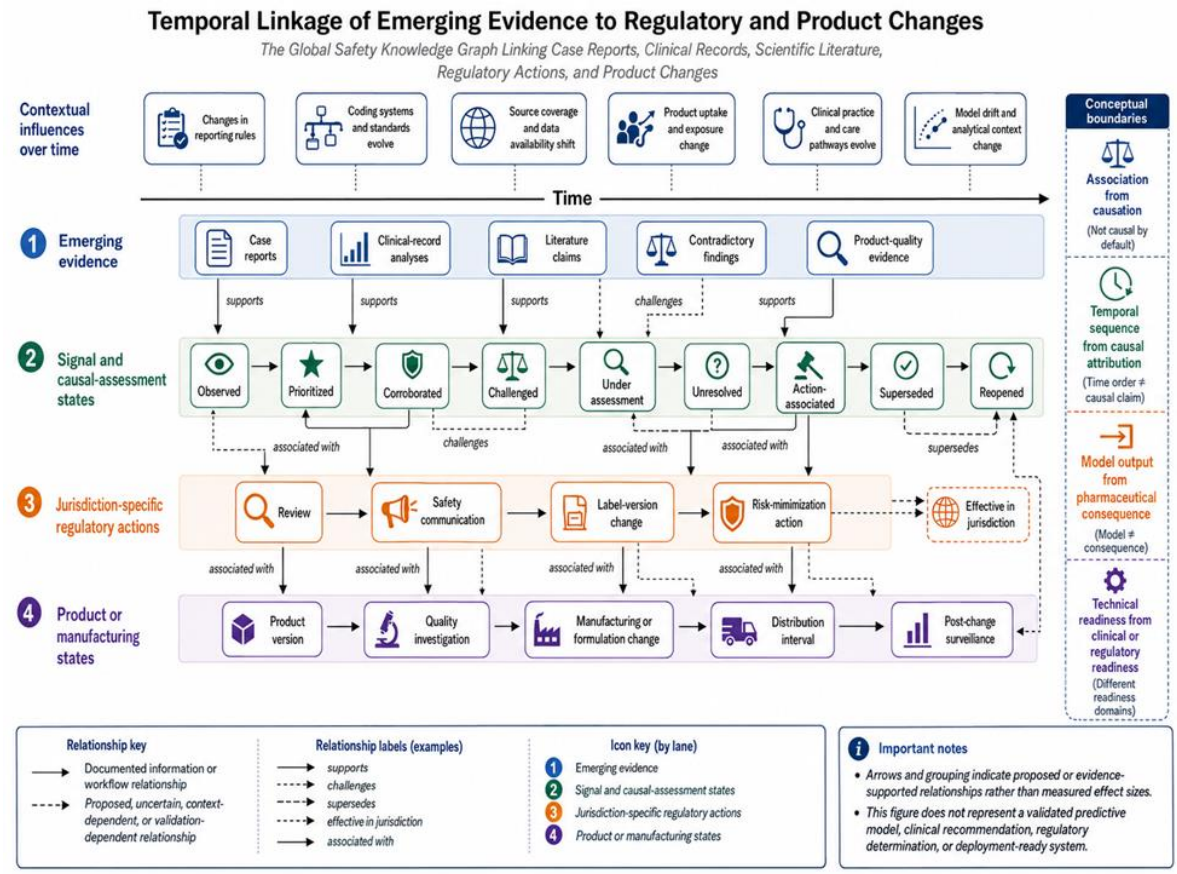

Figure 2. Temporal Linkage of Emerging Evidence to Regulatory and Product Changes.

The figure is an original conceptual synthesis developed for "The Global Safety Knowledge Graph Linking Case Reports, Clinical Records, Scientific Literature, Regulatory Actions, and Product Changes." 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.

Implementation should begin with lower-consequence functions such as source traceability, terminology mapping, duplicate review, evidence retrieval, and audit reconstruction. Signal prioritization or decision support should follow only after independent, intended-use-specific validation.

Principal priorities include multilingual terminology governance, privacy-preserving federation, product-identifier quality, provenance completeness, subgroup evaluation, drift monitoring, and prospective comparison with existing workflows. The architecture remains a research construct whose permissible use must remain narrower than its technical capacity to connect information.

CONCLUSION

The Global Safety Knowledge Graph is proposed as an evidence-bounded architecture for connecting case reports, clinical records, observational evidence, scientific claims, regulatory actions, and product histories without collapsing their distinct meanings. Its original contribution is an assertion-centered, temporal, provenance-preserving model that separates reported observations, inferred associations, safety signals, causal assessments, regulatory decisions, and product changes. The architecture may improve traceability, conflict inspection, and hypothesis evaluation, but it cannot correct all source biases, establish causality through graph structure, replace qualified review, or justify clinical, manufacturing, or regulatory action without use-specific validation and governance.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Beninger P. Pharmacovigilance: An overview. Clin Ther. 2018;40(12):1991-2004. doi:10.1016/j.clinthera.2018.07.012
2. Santoro A, Genov G, Spooner A, Raine J, Arlett P. Promoting and protecting public health: How the European Union pharmacovigilance system works. Drug Saf. 2017;40(10):855-69. doi:10.1007/s40264-017-0572-8
3. 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
4. Trifirò G, Sultana J, Bate A. From big data to smart data for pharmacovigilance: The role of healthcare databases and other emerging sources. Drug Saf. 2018;41(2):143-9. doi:10.1007/s40264-017-0592-4
5. Koutkias V. From data silos to standardized, linked, and FAIR data for pharmacovigilance: Current advances and challenges with observational healthcare data. Drug Saf. 2019;42(5):583-6. doi:10.1007/s40264-018-00793-z
6. Radecka A, Loughlin L, Foy M, Viana de Ferraz Guimaraes M, Macolic Sarinic V, Dimov Di Giusti M, et al. Enhancing pharmacovigilance capabilities in the EU regulatory network: The SCOPE Joint Action. Drug Saf. 2018;41(12):1285-302. doi:10.1007/s40264-018-0708-5
7. 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
8. Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, Broadbent M, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. Sci Rep. 2017;7(1):16416. doi:10.1038/s41598-017-16674-x
9. Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J. 2020;18:1414-28. doi:10.1016/j.csbj.2020.05.017
10. Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, Bender A, et al. A review of biomedical datasets relating to drug discovery: A knowledge graph perspective. Brief Bioinform. 2022;23(6). doi:10.1093/bib/bbac404
11. Lee S, Han J, Park RW, Kim GJ, Rim JH, Cho J, et al. Development of a controlled vocabulary-based adverse drug reaction signal dictionary for multicenter electronic health record-based pharmacovigilance. Drug Saf. 2019;42(5):657-70. doi:10.1007/s40264-018-0767-7
12. Wilkinson MD, Sansone SA, Schultes E, Doorn P, Santos LOBDS, Dumontier M. A design framework and exemplar metrics for FAIRness. Sci Data. 2018;5:180118. doi:10.1038/sdata.2018.118
13. Kreimeyer K, Menschik D, Winiecki S, Paul W, Barash F, Woo EJ, et al. Using probabilistic record linkage of structured and unstructured data to identify duplicate cases in spontaneous adverse event reporting systems. Drug Saf. 2017;40(7):571-82. doi:10.1007/s40264-017-0523-4
14. Juhlin K, Star K, Norén GN. A method for data-driven exploration to pinpoint key features in medical data and facilitate expert review. Pharmacoepidemiol Drug Saf. 2017;26(10):1256-65. doi:10.1002/pds.4285
15. 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
16. Donegan K, Owen R, Bird H, Burch B, Smith A, Tregunno P. Exploring the potential routine use of electronic healthcare record data to strengthen early signal assessment in UK medicines regulation: Proof-of-concept study. Drug Saf. 2018;41(9):899-910. doi:10.1007/s40264-018-0675-x
17. Huang K, Lin FJ, Ou HT, Hsu CN, Huang LY, Wang CC, et al. Building an active medical product safety surveillance system in Taiwan: Adaptation of the US Sentinel System common data model structure to the National Health Insurance Research Database in Taiwan. Pharmacoepidemiol Drug Saf. 2021;30(1):97-101. doi:10.1002/pds.5168

18. Yu Y, Ruddy KJ, Hong N, Tsuji S, Wen A, Shah ND, et al. ADEpedia-on-OHDSI: A next generation pharmacovigilance signal detection platform using the OHDSI common data model. J Biomed Inform. 2019;91:103119. doi:10.1016/j.jbi.2019.103119
19. Potts J, Genov G, Segec A, Raine J, Straus S, Arlett P. Improving the safety of medicines in the European Union: From signals to action. Clin Pharmacol Ther. 2020;107(3):521-9. doi:10.1002/cpt.1678
20. Fukuda K, Narukawa M. Decision and timing of safety-related labeling changes for new drugs in Japan: Comparison with the United States and the European Union. Pharmacoepidemiol Drug Saf. 2023;32(12):1331-40. doi:10.1002/pds.5664
21. Mostaghim SR, Gagne JJ, Kesselheim AS. Safety related label changes for new drugs after approval in the US through expedited regulatory pathways: Retrospective cohort study. BMJ. 2017;358. doi:10.1136/bmj.j3837
22. Trippe ZA, Brendani B, Meier C, Lewis D. Identification of substandard medicines via disproportionality analysis of individual case safety reports. Drug Saf. 2017;40(4):293-303. doi:10.1007/s40264-016-0499-5
23. Sagi S, Cohen HP, Woollett GR. Pharmacovigilance of biologics in a multisource environment. J Manag Care Spec Pharm. 2017;23(12):1249-54. doi:10.18553/jmcp.2017.23.12.1249
24. Borrega R, Cruz JP, Taylor P, Gonçalves J. Analysis of immunogenicity data in the product information of biological drugs: A need to report immunogenicity data systematically. BioDrugs. 2019;33(6):683-91. doi:10.1007/s40259-019-00387-w
25. Pinheiro LC, Candore G, Zaccaria C, Slattery J, Arlett P. An algorithm to detect unexpected increases in frequency of reports of adverse events in EudraVigilance. Pharmacoepidemiol Drug Saf. 2018;27(1):38-45. doi:10.1002/pds.4344
26. Candore G, Monzon S, Slattery J, Piccolo L, Postigo R, Kurz X, et al. The impact of mandatory reporting of non-serious safety reports to EudraVigilance on the detection of adverse reactions. Drug Saf. 2022;45(1):83-95. doi:10.1007/s40264-021-01137-0
27. Brown JS, Mendelsohn AB, Nam YH, Maro JC, Cocoros NM, Rodriguez-Watson C, et al. The US Food and Drug Administration Sentinel System: A national resource for a learning health system. J Am Med Inform Assoc. 2022;29(12):2191-200. doi:10.1093/jamia/ocac153
28. Sansone SA, McQuilton P, Rocca-Serra P, Gonzalez-Beltran A, Izzo M, Lister AL, et al. FAIRsharing as a community approach to standards, repositories and policies. Nat Biotechnol. 2019;37(4):358-67. doi:10.1038/s41587-019-0080-8
29. Rawson NSB, D'Arcy C. Healthcare databases for drug safety research: Data validity assessment remains crucial. Drug Saf. 2018;41(9):829-33. doi:10.1007/s40264-018-0673-z
30. Subbaswamy A, Saria S. From development to deployment: Dataset shift, causality, and shift-stable models in health AI. Biostatistics. 2020;21(2):345-52. doi:10.1093/biostatistics/kxz041