Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

EXPLAINABLE SURVIVAL MODELS FOR POST-MARKETING SAFETY SIGNAL PRIORITIZATION USING REPORTING AND LITERATURE EVIDENCE

Nguyen Van Nam[1], Tran Thi Hoa[1]*, Le Minh Duc[2]

1. ***Department of Computational Drug Discovery, Faculty of Pharmacy, Hanoi University of Science and Technology, Hanoi, Vietnam.***
2. ***Department of AI Pharmaceutical Systems, Faculty of Medicine, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam.***

ARTICLE INFO

Received:
16 January 2026
Received in revised form:
05 April 2026
Accepted:
09 April 2026
Available online:
28 April 2026

Keywords: Explainable AI, Survival analysis, Pharmacovigilance, Signal prioritization, DeepSurv, SHAP

ABSTRACT

Post-marketing safety surveillance must sift through large volumes of potential drug–adverse event associations, where signal timing is central to patient protection. Current prioritization tools often emphasize present evidence strength while giving less attention to when a signal may mature. Disproportionality analysis and manual review can help identify candidate associations, but they do not forecast the expected trajectory of signal maturation. They also provide limited transparency about which evidence streams most influence the urgency assigned to a signal. This article develops a conceptual explainable survival modeling framework for predicting time-to-signal confirmation or escalation for drug–adverse event pairs. The framework uses reporting trends, literature evidence, and product-specific covariates while decomposing each prediction into interpretable drivers. A deep survival model, such as DeepSurv, is proposed for longitudinal drug safety data with time-varying reporting and literature features updated over repeated surveillance intervals. SHAP values are used post hoc to attribute predicted hazard to specific features and evidence streams. Conceptually, the model could identify drug–event pairs that warrant earlier review than static prioritization approaches. For each prioritized signal, it would provide a transparent rationale, such as a rising reporting trajectory combined with emerging literature evidence accelerating the expected signal timeline. Explainable survival models could bring a more dynamic and auditable form of analytic support to post-marketing signal prioritization. Their greatest value lies in aligning time-to-event prediction with the evidentiary reasoning already used by pharmacovigilance experts.

To Cite This Article: Nam NV, Hoa TT, Duc LM. Explainable Survival Models for Post-Marketing Safety Signal Prioritization Using Reporting and Literature Evidence. Pharmacophore. 2026;17(2):83-92. https://doi.org/10.51847/rIT9k7poLT

Introduction

Post-marketing safety surveillance operates under persistent pressure to identify clinically meaningful drug–adverse event associations from noisy, rapidly accumulating evidence. Spontaneous reporting systems can support early detection, but they also contain reporting biases, duplicate narratives, missing information, and uneven case quality that complicate prioritization [1]. Recent work on artificial intelligence in pharmacovigilance emphasizes that automation should assist, rather than replace, expert review by helping safety teams manage large candidate signal queues [2]. In this setting, the core challenge is not only whether a signal exists, but also whether its evidentiary trajectory warrants timely escalation.

Traditional signal prioritization relies heavily on disproportionality measures, seriousness, novelty, and clinical judgment, but these approaches often provide a static view of evidence at a particular review date. Guidance on disproportionality analysis highlights the importance of interpreting reporting patterns carefully and transparently, especially when statistical signals may reflect confounding, reporting stimulation, or sparse data [3]. Machine learning methods have been proposed to support signal validation and prioritization, yet many such models classify or rank signals without explicitly modeling the time until escalation or confirmation [4]. Survival analysis offers a natural conceptual bridge because it represents signal maturation as a time-to-event process shaped by evolving evidence.

The recent development of explainable survival modeling creates an opportunity to combine temporal prediction with interpretable prioritization. DeepSurv extends the Cox modeling tradition through flexible nonlinear functions, allowing

Corresponding Author: Tran Thi Hoa; Department of Computational Drug Discovery, Faculty of Pharmacy, Hanoi University of Science and Technology, Hanoi, Vietnam. E-mail: tran.hoa@outlook.com

complex interactions to influence a predicted hazard while retaining the time-to-event framing [5]. DeepHit and Dynamic-DeepHit further demonstrate how neural survival models can represent competing risks and longitudinal covariate histories in clinical prediction settings [6, 7]. Explainability methods such as SurvSHAP and SurvSHAP(t) then provide mechanisms for attributing survival predictions to individual features over time, making model-based urgency more suitable for safety review [8, 9].

This article proposes an explainable survival model for post-marketing signal prioritization that integrates spontaneous reports, literature evidence, and product-specific context. The thesis is that drug–event pairs should be prioritized not only by current disproportionality or seriousness, but also by the predicted time until a signal reaches a reviewable threshold of evidence. Prior pharmacovigilance work on feature engineering, causality assessment, and individual case utility shows that structured and unstructured safety information can be transformed into decision-support features [10, 11]. The proposed framework therefore treats signal timing as an auditable prediction task in which each forecast is accompanied by an explanation of the evidence that accelerated or delayed the expected signal trajectory.

Background

Post-Marketing Safety Signal Management

Post-marketing signal management is a staged process in which candidate drug–event associations are detected, reviewed, prioritized, validated, and either escalated, monitored, or dismissed. Signal validation studies show that machine learning can support classification decisions, but final interpretation remains grounded in pharmacovigilance expertise, clinical plausibility, and regulatory context [4]. Signal communication reviews also indicate that different stakeholders frame and transmit safety signals through diverse evidentiary and procedural pathways [12]. An explainable survival model should therefore be designed as a support tool for committees and safety scientists, not as an autonomous signal confirmation mechanism.

Spontaneous Reporting Systems and Literature as Evidence Streams

Spontaneous reporting systems such as FAERS and VigiBase provide broad post-marketing coverage, but their evidentiary value depends on careful interpretation of case characteristics, reporting behavior, and disproportionality patterns [1]. Literature evidence, including published case reports and safety reviews, can complement spontaneous reports by adding clinical detail, temporality, dechallenge or rechallenge information, and mechanistic discussion. Studies on routinely collected health data and signal validation show that external evidence streams can strengthen prioritization when they are aligned with pharmacovigilance questions [13]. A survival framework can treat each new report or publication as time-stamped evidence that changes the estimated trajectory of signal maturation.

Survival Analysis in Clinical and Pharmacovigilance Settings

Survival analysis provides a principled language for modeling time from an origin point to an event while accounting for observations that have not yet experienced the event. In the proposed safety setting, the origin may be the first report, market entry, or the first month in which a drug–event pair enters surveillance, while the event may be escalation, confirmation, or a regulatory action. Cox-type models remain useful because they express covariate effects on hazard in a transparent time-to-event framework, and neural extensions can relax linearity assumptions [5]. Although pharmacovigilance applications often focus on disproportionality rather than time-to-event prediction, the same survival principles can represent how safety evidence accumulates toward action.

Deep Survival Models and Their Explainability

DeepSurv uses a neural network to estimate a risk function analogous to the Cox proportional hazards model, making it suitable for nonlinear covariate relationships in time-to-event settings [5]. DeepHit extends survival modeling to competing risks, while Dynamic-DeepHit incorporates longitudinal information so that predictions can change as new covariate values arrive [6, 7]. For pharmacovigilance, these architectures are conceptually useful because signal evidence evolves through repeated reports, publications, and regulatory updates. SHAP-based survival explanation methods, including SurvSHAP and SurvSHAP(t), can then identify which covariates influence predicted survival or hazard at specific time points [8, 9].

Prior Work on Signal Prioritization and the Gap in Explainable Timing Prediction

Existing machine learning work in pharmacovigilance has addressed tasks such as adverse event coding, case report triage, causality support, and signal validation, demonstrating that automated models can help process safety information [14-16]. Other studies have explored feature engineering and case utility prediction in FAERS, showing that structured attributes can support review prioritization [10, 11]. However, many models remain static in the sense that they estimate current relevance, seriousness, or classification status rather than the expected time until a signal matures. The gap addressed here is an explainable timing model that predicts when escalation could occur and presents that prediction in terms of the reporting, literature, and product features that safety reviewers can inspect.

Model Development Overview

High-Level Prediction Pipeline

For each drug–adverse event pair, the proposed pipeline constructs a longitudinal record whose feature values are refreshed across repeated surveillance intervals. Reporting data, literature evidence, and product context are aligned to the same time axis so that the model can estimate a survival curve for remaining unconfirmed, un-escalated, or otherwise unresolved. This design follows the broader pharmacovigilance movement toward intelligent automation while retaining reviewer oversight and auditability [17]. The model output is not a final safety judgment but a predicted time-to-signal profile accompanied by feature attributions that explain why a pair is considered more or less urgent.

Figure 1 illustrates the proposed evidence-trajectory-to-signal-urgency architecture, showing how reporting evidence, literature evidence, and product context are transformed into censoring-aware survival predictions with SHAP-based explanations for pharmacovigilance review.

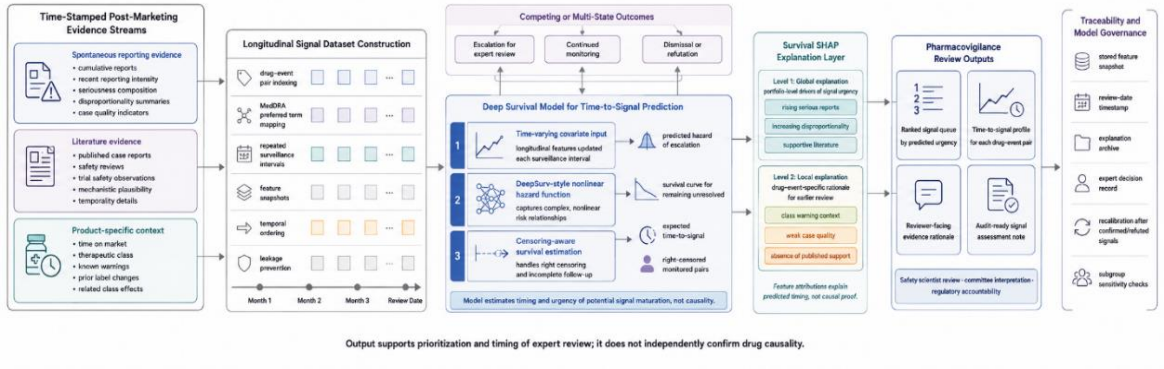

Figure 1. Explainable survival architecture for post-marketing safety signal prioritization.

Core Input Features

Reporting features would include cumulative case counts, recent reporting intensity, disproportionality summaries, seriousness indicators, and case-level quality signals derived from structured and narrative fields. Literature features would represent the presence and nature of published evidence, such as case reports, clinical trial safety observations, or reviews that discuss the drug–event association. Product-specific features would include time on market, therapeutic class, known warnings, prior label changes, and related class effects, reflecting the contextual reasoning used in signal assessment. Prior work on coding adverse reaction reports and identifying useful case reports supports the feasibility of converting heterogeneous safety evidence into model-ready covariates [11, 15].

Design Principles

The framework should be dynamic, censoring-aware, explainable, and aligned with signal management decisions. Dynamic modeling is needed because the meaning of a candidate signal changes when new cases, publications, or regulatory context emerge, as reflected in longitudinal survival architectures [7]. Censoring awareness is important because many monitored drug–event pairs will not have reached escalation or confirmation by the data-lock date. Explainability is essential because pharmacovigilance professionals must understand whether the model's prioritization follows credible clinical and evidentiary logic rather than opaque statistical artifacts [18].

Data Sources and Feature Engineering

Constructing a Longitudinal Signal Dataset

A longitudinal signal dataset would link drug substances to MedDRA preferred terms and aggregate spontaneous reports over repeated surveillance intervals. FAERS or VigiBase records can provide the reporting stream, while PubMed-indexed literature can provide time-stamped evidence describing suspected, observed, or reviewed drug–event relationships. The construction process should preserve temporal ordering so that evidence available after a review point is not used to explain predictions made before that point. Network and data-integration approaches in drug safety illustrate how structured safety entities can be linked across databases to support prioritization and signal characterization [19, 20].

Creating Time-Varying Covariates for Reporting and Literature Evidence

Time-varying covariates would summarize how evidence changes across surveillance intervals, rather than treating a drug–event pair as a static observation. Reporting covariates could represent cumulative reports, recent changes in reporting rate, seriousness composition, and changes in disproportionality signals, while literature covariates could represent newly appearing case reports or safety reviews. Artificial intelligence applications in pharmacovigilance increasingly depend on converting unstructured and structured data into features that can support downstream assessment [21, 22]. In an explainable survival model, these features are especially important because they become the units through which the model explains acceleration or delay in predicted signal timing.

Table 1 presents the proposed evidence-to-feature architecture through which reporting trends, literature evidence, and product context are transformed into time-varying covariates for explainable survival-based signal prioritization.

Pharmacophore, 17(2) 2026, Pages 83-92

Table 1. Evidence-to-survival-feature architecture for dynamic signal prioritization

Evidence domain	Model-ready feature construction	Survival-model function	Expected influence on predicted signal timing	Reviewer-facing interpretation
Spontaneous report volume	Cumulative number of drug–event reports aggregated by surveillance interval	Establishes the baseline longitudinal visibility of the candidate association	Higher cumulative volume may shorten predicted time-to-escalation when accompanied by credible supporting evidence	The association has accumulated enough reporting activity to justify closer monitoring
Recent reporting acceleration	Change in reporting frequency over recent intervals compared with prior baseline	Captures temporal momentum in the evidence trajectory	A sharp recent increase may increase predicted hazard of signal maturation	The signal may be emerging rapidly rather than remaining stable
Disproportionality pattern	Interval-specific or smoothed disproportionality summaries, such as reporting odds or related measures	Provides statistical evidence of reporting imbalance	Persistent or rising disproportionality may shorten predicted time-to-review	The association is becoming statistically more visible within the reporting system
Seriousness composition	Proportion or count of serious outcomes, hospitalization, life-threatening events, or death reports	Weights the clinical urgency of the signal trajectory	Greater seriousness may increase prioritization even when total case volume is moderate	The potential public-health consequence justifies earlier expert review
Case quality indicators	Completeness of narratives, temporality, dechallenge/rechallenge information, duplicate likelihood, and reporter type	Distinguishes credible evidence from noisy reporting accumulation	Strong case quality may accelerate predicted escalation; weak case quality may delay it	The model is sensitive to whether reports contain assessable clinical detail
Literature case evidence	Time-stamped appearance of published case reports discussing the drug–event association	Adds external clinical documentation to the survival trajectory	New case literature may shorten predicted time-to-signal, especially when reporting trends are rising	Published clinical observations strengthen the rationale for review
Literature synthesis evidence	Safety reviews, trial safety discussions, mechanistic papers, or systematic summaries	Represents broader external corroboration beyond isolated cases	Supportive synthesis evidence may increase predicted hazard of escalation	The association is no longer limited to spontaneous reports alone
Product life-cycle context	Time since market entry, launch period, mature product status, or recent approval expansion	Helps separate early post-launch reporting volatility from persistent signal development	Early-market products may require different calibration than long-marketed drugs	Timing is interpreted relative to expected reporting behavior across the product life cycle
Known warning and label context	Existing warnings, contraindications, precautions, or prior label changes	Positions the association relative to established safety knowledge	Novel or label-inconsistent events may receive greater prioritization than already-characterized events	The model differentiates new concerns from known and already-managed risks
Therapeutic class and related effects	Class membership, known class effects, pharmacologic mechanism, and related product signals	Adds contextual plausibility and cross-product safety reasoning	Class-consistent evidence may accelerate review when paired with rising reports or literature	The model reflects pharmacovigilance reasoning about class-level safety patterns
Surveillance endpoint status	Escalation, confirmation, label update, regulatory communication, dismissal, or continued monitoring	Defines event occurrence or censoring for survival learning	Clear endpoint definition determines what the model learns as "signal maturation"	The prediction is only as meaningful as the operational endpoint definition
Censoring status	Drug–event pairs not escalated by the data-lock date are treated as right-censored	Prevents unresolved pairs from being incorrectly labeled as negative signals	Preserves information from monitored but unresolved associations	The model acknowledges uncertainty rather than forcing premature classification

Defining the Event and Censoring

The event of interest should be defined operationally as the first documented escalation, confirmation, label change, regulatory communication, or other prespecified signal management milestone. A drug–event pair that has not reached the event by the end of observation would be treated as right-censored, preserving its contribution without incorrectly labeling it as a negative signal. This definition is consistent with survival analysis principles and with pharmacovigilance processes in which many associations remain under monitoring for extended periods [12]. The event definition must be auditable because any ambiguity in what counts as confirmation would affect both model training and interpretation.

Explainable Survival Model Architecture

Model Choice – DeepSurv with Time-Varying Inputs

A DeepSurv-style architecture is appropriate because it extends the Cox model with a neural risk function while retaining a hazard-based interpretation of time-to-event prediction [5]. For signal prioritization, the model would ingest updated feature vectors that represent the state of reporting, literature evidence, and product context at each surveillance interval. Nonlinear interactions could then reflect patterns such as a modest disproportionality signal becoming more urgent when paired with serious cases and supportive literature. The model should remain conceptually subordinate to expert review, with SHAP explanations used to show how each input feature shaped the predicted hazard.

Handling Recurrent Events and Competing Risks

Some safety surveillance questions may involve competing signal outcomes, such as escalation for regulatory review, monitoring without action, or dismissal after expert assessment. DeepHit provides a survival modeling precedent for competing risks, allowing different event types to be represented without forcing them into a single undifferentiated endpoint [6]. Dynamic-DeepHit further supports the idea that updated longitudinal evidence can revise the predicted trajectory as new information becomes available [7]. In pharmacovigilance, a multi-state extension could conceptually represent transitions from detected association to validated signal, regulatory action, or continued surveillance.

SHAP decomposition for survival predictions

SHAP decomposition would translate a model's predicted hazard or survival estimate into feature-level contributions that reviewers can inspect. SurvSHAP adapts Shapley-style reasoning to survival models, while SurvSHAP(t) emphasizes time-dependent explanations that can show how feature influence changes across the predicted signal horizon [8, 9]. For a drug–event pair, these explanations could distinguish whether urgency is being driven by recent serious reports, increasing disproportionality, supportive literature, or product-class context. This explanatory layer is central to XAI in drug safety because it turns model-based prioritization into a traceable argument that can be compared with clinical and regulatory reasoning [23].

Identifying Key Drivers of Signal Timing and Priority

Global Feature Importance for Signal Urgency

Global explanations would summarize which categories of evidence most often accelerate the predicted timing of signal escalation across the monitored portfolio. In this framework, recent reporting acceleration, seriousness patterns, and literature-supported clinical plausibility would be expected to emerge as important drivers when they consistently shift predicted hazard upward [1]. Systematic reviews of adverse drug reaction signal detection methods show that disproportionality, reporting behavior, and methodological assumptions must be interpreted together rather than treated as isolated indicators [24]. Global SHAP summaries would therefore support portfolio-level learning by showing which evidence patterns repeatedly make candidate signals appear more urgent.

Local Explanation for a Specific Signal of Interest

For an individual drug–event pair, local explanation would provide a signal-specific rationale rather than only a rank position on a triage list. A SHAP waterfall or equivalent explanation could show that a recent increase in serious reports, a supportive published case, and a relevant class warning each contributed to a higher predicted hazard of escalation [23]. This type of explanation is consistent with pharmacovigilance decision support work that emphasizes case utility and reviewer-facing interpretability [11]. The goal is not to claim confirmation, but to show why the model would recommend earlier expert attention for that pair.

Interaction between Reporting and Literature Evidence

The interaction between reporting evidence and literature evidence is especially important because neither stream is fully sufficient on its own. Disproportionality may identify an association that is statistically notable but clinically uncertain, while published case evidence may add temporality, biological plausibility, or differential diagnosis detail [3]. SHAP dependence analysis could conceptually reveal that literature support amplifies the effect of rising reporting trends, moving a borderline association into a higher-priority review category. This interpretation aligns with real-world signal validation approaches that combine multiple evidence streams to strengthen or weaken prioritization decisions [13].

Counterfactual Analysis for Signal Management

Counterfactual explanation would allow safety teams to ask how the predicted signal timeline might change under plausible future evidence scenarios. For example, the model could conceptually compare the current trajectory with scenarios in which reporting continues to increase, literature evidence remains absent, or a label update changes product context. Dynamic survival models support this kind of reasoning because they update risk estimates as longitudinal information changes [7]. In signal management, these counterfactuals would be most useful as scenario-planning aids rather than as deterministic forecasts.

Explainability Methods for Signal Management Teams

Dashboard for Signal Prioritization with Explanations

A signal prioritization dashboard would rank drug–event pairs by predicted urgency while displaying a concise explanation of the dominant evidence drivers. The interface should make clear whether the priority is driven by reporting acceleration, case seriousness, literature support, product class context, or a combination of these factors [21]. Prior work on artificial intelligence in pharmacovigilance cautions that automated outputs must be integrated into workflows in ways that preserve human accountability [18]. A useful dashboard would therefore present model reasoning as review support, not as a replacement for clinical and regulatory judgment.

Narrative Explanations for Signal Assessment Reports

Narrative explanations would translate model attributions into standardized prose suitable for inclusion in signal assessment documentation. For example, a generated explanation could describe how recent reporting patterns and literature evidence jointly increased the predicted urgency while noting which features reduced concern. Work on adverse reaction coding and automated case processing shows that language-oriented AI can assist pharmacovigilance documentation when outputs remain reviewable and traceable [14, 16]. In the proposed framework, the narrative should remain factual and evidentiary, avoiding causal overstatement unless supported by expert assessment.

Auditing Model Decisions against Expert Review

Auditability requires that each prediction, feature snapshot, and explanation be stored with the review date so that later reviewers can reconstruct why a signal was prioritized. This is particularly important because pharmacovigilance AI must be evaluated not only for technical behavior but also for alignment with safety governance and professional responsibility [25]. Crowdsourced and expert-labeled pharmacovigilance datasets show that reviewer judgment can be structured into training and evaluation resources, but disagreement and uncertainty must remain visible [26]. An explanatory audit trail would help identify cases where the model's reasoning diverges from expert interpretation in systematic or clinically meaningful ways.

Feedback Loop from Confirmed and Refuted Signals

A feedback loop would incorporate later signal outcomes, such as confirmation, refutation, continued monitoring, or regulatory action, into future model development. Machine learning for signal validation demonstrates that historical decisions can inform prioritization support, but the model must avoid treating past review patterns as unquestionable truth [4]. Feedback should therefore include outcome status, reviewer rationale, and whether new evidence changed the interpretation of the original signal. Over time, this process would be expected to refine both prediction and explanation quality while preserving expert oversight.

Integration Into Pharmacovigilance Workflows

Automated Weekly Signal Triage

In routine use, the model could update candidate signal priorities as new spontaneous reports, literature records, and product-context changes enter the surveillance environment. Intelligent automation in drug safety has been framed as a way to reduce manual burden and help specialists focus attention on cases or associations most likely to require expert review [17]. Recent industry perspectives similarly emphasize that pharmacovigilance AI should be embedded into existing operational processes rather than deployed as isolated technical tools [21]. A weekly triage workflow would therefore deliver ranked, explained signal candidates into the same platforms where safety teams already document assessment decisions.

Table 2 shows a weekly AI-assisted pharmacovigilance triage workflow in which multimodal safety data are continuously integrated, ranked, and translated into explainable signal candidates that are directly delivered into existing safety management systems for expert review.

Table 2. Weekly AI-assisted pharmacovigilance signal triage workflow

Component of workflow	Data inputs	AI-driven processing step	Output to safety team	Value in routine use
Spontaneous adverse event reports	ICSRs, EHR-linked reports, patient submissions	NLP extraction of drugs, events, de-duplication	Normalized case records	Reduces manual case structuring burden
Literature surveillance	Published articles, case reports, preprints	Text mining + relevance scoring	Candidate safety signals with evidence links	Expands detection beyond reporting systems

Product-context monitoring	Label updates, batch changes, manufacturing variations	Context-aware risk re-scoring	Adjusted signal prioritization weights	Captures changing exposure and formulation risk
Signal prioritization layer	Combined multimodal evidence streams	Multi-factor ranking model (frequency, severity, novelty)	Ranked list of candidate signals	Focuses attention on highest-risk associations
Explainability module	Model outputs + feature attribution data	SHAP-like or rule-based explanation generation	Interpretable rationale per signal	Supports regulatory and clinical trust
Workflow integration	Prioritized outputs + explanations	Automated routing into safety database systems	Weekly triage dashboard in existing PV tools	Embeds AI into established pharmacovigilance operations

Supporting Regulatory Interactions and Labeling Decisions

Time-stamped explanations could support regulatory interactions by documenting what evidence was available when a signal was prioritized and why the model suggested increased urgency. Reviews of AI applications in signal management argue that explainable and traceable systems are particularly important when outputs may influence labeling, regulatory communication, or resource allocation. Broader discussions of regulatory integration also emphasize that AI tools in pharmacovigilance should be transparent enough for inspection by internal governance groups and external stakeholders. The model's role would be to organize and explain evidence trajectories, while final regulatory and labeling decisions would remain human responsibilities.

Evaluation Strategy

Predictive Performance for Signal Timing

Evaluation should assess whether the model appropriately orders drug–event pairs by expected time-to-signal while respecting censoring and temporal data structure. Survival modeling measures such as discrimination, calibration, and prediction error could be used conceptually, but they should be interpreted alongside baseline approaches such as static disproportionality-based prioritization [5]. Competing-risk evaluation may also be appropriate if the framework distinguishes escalation, dismissal, and continued monitoring as separate outcomes [6]. The purpose of evaluation is to determine whether the model supports better-timed review decisions, not to present isolated numerical claims.

Lead Time and Sensitivity

Lead-time evaluation should examine whether the model would have highlighted eventual safety signals earlier than existing review practices under a temporally faithful retrospective design. This assessment should avoid using future evidence at earlier prediction points, because leakage would create unrealistic confidence in the model's apparent usefulness. Signal detection methodology reviews emphasize that performance must be interpreted in light of reporting artifacts, case definitions, and the operational purpose of surveillance [24]. Sensitivity should therefore be considered together with false-positive burden and reviewer capacity, because a model that flags too many weak candidates may not improve real-world prioritization.

Explanatory Audit and User Acceptance

Explanatory evaluation should ask whether safety assessors find the model's rationales clinically plausible, evidence-based, and useful for deciding which signals require earlier attention. XAI methods for survival prediction can provide feature-level and time-dependent explanations, but the value of those explanations depends on whether experts can connect them to pharmacological, clinical, and regulatory reasoning [8, 9]. User acceptance should also examine whether explanations improve trust calibration, meaning that reviewers can recognize both strong and weak model rationales. This evaluation would be especially important in drug safety, where transparent reasoning is often as important as prediction itself [23].

Table 3 consolidates the evaluation, interpretability, and governance requirements needed to translate explainable survival prioritization from a conceptual model into an auditable pharmacovigilance decision-support workflow.

Table 3. Governance, evaluation, and interpretability framework for operational deployment

Deployment requirement	Analytical implementation	Pharmacovigilance value added	Key failure mode if neglected	Governance safeguard
Temporally faithful prediction	Train and evaluate predictions only using evidence available before each surveillance review date	Prevents inflated performance and supports realistic lead-time assessment	Future evidence leakage may make the model appear falsely useful	Maintain review-date feature snapshots and locked temporal validation protocols
Censoring-aware evaluation	Use survival-appropriate discrimination, calibration, and prediction-error measures	Reflects unresolved drug–event pairs without treating them as false signals	Static classification metrics may misrepresent monitored associations	Report survival-specific metrics alongside operational triage outcomes

Baseline comparator design	Compare survival prioritization against static disproportionality, seriousness-based triage, and expert-review baselines	Clarifies whether time-to-event modeling adds practical value	Model benefit may be overstated without realistic comparators	Predefine baseline strategies before retrospective evaluation
Lead-time assessment	Estimate whether eventual escalated signals would have been prioritized earlier	Measures the core operational promise of the framework	Earlier flagging may come at the cost of excessive false positives	Evaluate lead time together with reviewer workload and false-positive burden
Local explanation quality	Provide feature-level SHAP explanations for each prioritized drug–event pair	Helps reviewers understand why the model recommends earlier attention	Opaque urgency scores may reduce trust and hinder regulatory defensibility	Require explanation review before escalation decisions are documented
Global explanation monitoring	Track portfolio-level drivers of predicted urgency across products and event classes	Identifies systematic evidence patterns influencing prioritization	Model may rely too heavily on reporting artifacts or product popularity	Periodically audit global explanations by therapeutic area and event type
Counterfactual scenario use	Examine how predicted timing changes under plausible future evidence conditions	Supports planning for continued monitoring or intensified review	Counterfactuals may be misread as deterministic forecasts	Label all scenarios as decision-support simulations, not predictions of causality
Expert alignment assessment	Compare model rationales with safety scientist and committee reasoning	Evaluates whether explanations are clinically and regulatorily meaningful	Technically accurate predictions may still be operationally unhelpful	Use structured reviewer feedback to assess plausibility and actionability
Audit trail preservation	Store model version, feature snapshot, prediction date, explanation, and final human decision	Allows later reconstruction of why a signal was prioritized or deferred	Decisions may become difficult to justify during internal or external review	Maintain immutable prediction and explanation logs
Bias and reporting-artifact review	Examine performance across product age, therapeutic class, event seriousness, geography, and reporting source	Reduces the risk that the model prioritizes surveillance artifacts rather than safety relevance	Stimulated reporting, media attention, or duplicate reports may distort urgency	Conduct subgroup calibration, duplicate checks, and reporting-lag sensitivity analyses
Human oversight boundary	Position the system as review support rather than autonomous signal confirmation	Preserves professional and regulatory accountability	Model output may be mistaken for confirmed causal evidence	Require explicit expert sign-off for escalation, dismissal, or labeling action
Recalibration and learning cycle	Incorporate later confirmed, refuted, and monitored signal outcomes into model maintenance	Improves future prediction while learning from expert-reviewed outcomes	Historical review patterns may encode institutional bias or inconsistent decisions	Recalibrate with outcome labels plus reviewer rationale, not outcome labels alone

Limitations

Reporting Lag and Data Completeness

Reporting lag, duplicate reports, stimulated reporting, incomplete narratives, and uneven literature indexing can all distort the apparent timing of safety evidence. Because spontaneous reporting systems are not designed as incidence databases, an explainable survival model could learn reporting dynamics that partly reflect surveillance behavior rather than underlying drug risk [1]. Literature evidence may also appear after clinical concern has already emerged, which can complicate interpretation of whether publications are predictors, consequences, or parallel indicators of signal maturation. These limitations mean that model outputs should be treated as prioritization support rather than direct evidence of causality.

Generalizability across Products and Event Types

Model generalizability may vary across therapeutic areas, drug life-cycle stages, event seriousness, and the clinical recognizability of adverse reactions. Network-based pharmacovigilance studies suggest that safety associations can differ substantially by product class and event structure, which may require subgroup calibration or class-specific review strategies [19, 20]. Neural survival models may capture complex patterns, but they can also encode historical practice patterns that do not transfer well to novel products or rare events. For this reason, a global model should be complemented by expert review, periodic recalibration, and sensitivity analyses across clinically meaningful subgroups.

Conclusion

An explainable survival model for post-marketing safety signal prioritization would reframe signal management as a dynamic time-to-event problem. Instead of ranking drug–event pairs only by their current evidence strength, the model would estimate how quickly each pair may progress toward escalation or confirmation.

The main strength of this approach is its alignment with the way pharmacovigilance evidence actually accumulates over time. By combining reporting trends, literature evidence, and product-specific context, the model would provide not only a prioritization score but also a transparent account of the factors shaping that priority.

Important challenges remain before such a framework could be used operationally. Data timeliness, literature completeness, endpoint definition, model transportability, and prospective validation would all need careful governance within pharmacovigilance organizations.

The next step is to integrate explainable survival modeling into existing signal-management platforms as a decision-support layer. Collaboration between regulators, industry safety teams, clinicians, and methodologists would be essential to evaluate whether this approach improves the timeliness, transparency, and consistency of post-marketing safety decisions.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Cutroneo PM, Sartori D, Tuccori M, Crisafulli S, Battini V, Carnovale C, et al. Conducting and interpreting disproportionality analyses derived from spontaneous reporting systems. Front Drug Saf Regul. 2024;3:1323057.
2. Bate A, Luo Y. Artificial intelligence and machine learning for safe medicines. Drug Saf. 2022;45(5):403-5.
3. 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): explanation and elaboration. Drug Saf. 2024;47(6):585-99.
4. Imran M, Bhatti A, King DM, Lerch M, Dietrich J, Doron G, et al. Supervised machine learning-based decision support for signal validation classification. Drug Saf. 2022;45(5):583-96.
5. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18(1):24.
6. Lee C, Zame W, Yoon J, Van Der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. Proc AAAI Conf Artif Intell. 2018;32(1).
7. Lee C, Yoon J, Van Der Schaar M. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng. 2019;67(1):122-33.
8. Krzyziński M, Spytek M, Baniecki H, Biecek P. SurvSHAP(t): time-dependent explanations of machine learning survival models. Knowl Based Syst. 2023;262:110234.
9. Alabdallah A, Pashami S, Rögnvaldsson T, Ohlsson M. SurvSHAP: a proxy-based algorithm for explaining survival models with SHAP. In: 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA). IEEE; 2022. p. 1-10.
10. Kreimeyer K, Dang O, Spiker J, Muñoz MA, Rosner G, Ball R, et al. Feature engineering and machine learning for causality assessment in pharmacovigilance: lessons learned from application to the FDA Adverse Event Reporting System. Comput Biol Med. 2021;135:104517.
11. Muñoz MA, Dal Pan GJ, Wei YJ, Delcher C, Xiao H, Kortepeter CM, et al. Towards automating adverse event review: a prediction model for case report utility. Drug Saf. 2020;43(4):329-38.
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.
13. Gauffin O, Brand JS, Vidlin SH, Sartori D, Asikainen S, Català M, et al. Supporting pharmacovigilance signal validation and prioritization with analyses of routinely collected health data: lessons learned from an EHDEN network study. Drug Saf. 2023;46(12):1335.
14. Letinier L, Jouganous J, Benkebil M, Bel-Létoile A, Goehrs C, Singier A, et al. Artificial intelligence for unstructured healthcare data: application to coding of patient reporting of adverse drug reactions. Clin Pharmacol Ther. 2021;110(2):392-400.
15. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. Drug Saf. 2022;45(5):535-48.

16. Abatemarco D, Perera S, Bao SH, Desai S, Assuncao B, Tetarenko N, et al. Training augmented intelligent capabilities for pharmacovigilance: applying deep-learning approaches to individual case safety report processing. Pharm Med. 2018;32(6):391.
17. Bate A, Hobbiger SF. Artificial intelligence, real-world automation and the safety of medicines. Drug Saf. 2021;44(2):125-32.
18. Hauben M, Hartford CG. Artificial intelligence in pharmacovigilance: scoping points to consider. Clin Ther. 2021;43(2):372-9.
19. Barbieri MA, Abate A, Balogh OM, Pétervári M, Ferdinandy P, Ágg B, et al. Network analysis and machine learning for signal detection and prioritization using electronic healthcare records and administrative databases: a proof of concept in drug-induced acute myocardial infarction. Drug Saf. 2025;48(5):513-26.
20. Pétervári M, Benczik B, Balogh OM, Petrovich B, Ágg B, Ferdinandy P. Network analysis for signal detection in spontaneous adverse event reporting database: application of network weighting normalization to characterize cardiovascular drug safety. Drug Saf. 2022;45(11):1423-38.
21. 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.
22. Comfort S, Perera S, Hudson Z, Dorrell D, Meireis S, Nagarajan M, et al. Sorting through the safety data haystack: using machine learning to identify individual case safety reports in social-digital media. Drug Saf. 2018;41(6):579-90.
23. Ward IR, Wang L, Lu J, Bennamoun M, Dwivedi G, Sanfilippo FM. Explainable artificial intelligence for pharmacovigilance: what features are important when predicting adverse outcomes? Comput Methods Programs Biomed. 2021;212:106415.
24. 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.
25. Danysz K, Cicirello S, Mingle E, Assuncao B, Tetarenko N, Mockute R, et al. Artificial intelligence and the future of the drug safety professional. Drug Saf. 2019;42(4):491.
26. Gartland A, Bate A, Painter JL, Casperson TA, Powell GE. Developing crowdsourced training data sets for pharmacovigilance intelligent automation. Drug Saf. 2021;44(3):373-82.