Pharmacophore

ISSN-2229-5402

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

DEEP SURVIVAL MODELS FOR ONCOLOGY TRIAL DISCONTINUATION USING TARGET CLASS, TOXICITY SIGNALS, AND PROTOCOL COMPLEXITY

Ahmed El-Kholy[1]*, Nour Abdelrahman[1], Karim Hassan[2]

1. ***Department of Pharmaceutical AI Engineering, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt.***
2. ***Department of Computational Drug Systems, Faculty of Medicine, Ain Shams University, Cairo, Egypt.***

ARTICLE INFO

Received:
11 December 2024
Received in revised form:
28 March 2025
Accepted:
29 March 2025
Available online:
28 April 2025

Keywords: Deep survival analysis, Oncology trials, Trial discontinuation, Toxicity signals, Protocol complexity, Drug development risk

ABSTRACT

Oncology clinical trials face a high risk of discontinuation, creating financial loss, delaying evidence generation, and slowing patient access to effective therapies. Discontinuation also raises ethical concerns when patients, investigators, and sponsors invest effort in trials that may not reach interpretable endpoints. Existing trial risk models often rely on static regression frameworks and limited feature sets. They are poorly suited to model time-to-discontinuation, dynamic toxicity emergence, or non-linear interactions among target class, safety burden, and protocol design complexity. This article proposes a conceptual deep survival modeling framework for predicting oncology trial discontinuation over time. The model would use drug target class, toxicity signals, and protocol complexity features to generate trial-specific survival curves. The proposed architecture would encode structured trial registry variables, target-class representations, normalized safety signals, and protocol complexity measures as covariates. A deep survival model, such as a DeepSurv-style network or a discrete-time neural survival architecture, would estimate the conditional probability that a trial remains active over time. Conceptually, the model would produce survival functions for individual oncology trials, update risk estimates as new toxicity evidence becomes available, and identify features most associated with elevated discontinuation risk. These outputs would support portfolio triage, operational monitoring, and protocol redesign. A deep survival approach could improve oncology development risk analytics by combining time-to-event modeling with flexible feature learning. Such models could support earlier identification of vulnerable trials and more informed drug development decisions.

To Cite This Article: El-Kholy A, Abdelrahman N, Hassan K. Deep Survival Models for Oncology Trial Discontinuation Using Target Class, Toxicity Signals, and Protocol Complexity. Pharmacophore. 2025;16(2):22-31. https://doi.org/10.51847/FpJ2yXmJEk

Introduction

Oncology clinical trials are frequently exposed to attrition pressures that arise from scientific uncertainty, safety concerns, slow accrual, and shifting strategic priorities. Empirical analyses of oncology development have shown that early termination is not merely an administrative endpoint but a clinically and financially consequential event that can interrupt evidence generation for promising therapies [1]. Broader work on clinical trial failure emphasizes that discontinuation can occur across development phases, with drivers including insufficient efficacy, tolerability limitations, recruitment barriers, and sponsor decisions [2]. Because oncology trials often involve vulnerable patients and intensive monitoring, the ethical cost of avoidable discontinuation is substantial.

Traditional risk scoring in pharmaceutical portfolio management has often relied on static feasibility assessments, expert judgment, or regression-based estimates of success. Although clinical trial success-rate modeling has provided useful aggregate benchmarks for development strategy [3], static models usually do not represent the evolving hazard of discontinuation as trial evidence accumulates. Reviews of drug development failure have highlighted that poor predictive translation, safety liabilities, and inadequate trial design remain persistent obstacles across therapeutic areas [4]. A survival

Corresponding Author: Ahmed El-Kholy. Department of Pharmaceutical AI Engineering, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt. E-mail: ahmed.elkholy@outlook.com

modeling framework is therefore attractive because it treats discontinuation as a time-indexed event rather than as a fixed binary label.

Deep survival models provide a natural methodological bridge between time-to-event analysis and flexible representation learning. DeepSurv introduced a neural extension of Cox modeling that can learn non-linear covariate effects while preserving a survival-analysis objective [5], and DeepHit extended neural survival modeling to discrete-time competing-risk settings [6]. Dynamic-DeepHit further showed how longitudinal information can update survival predictions over time [7], suggesting a direct analogy for oncology trials in which safety signals and operational metrics emerge during trial conduct. General reviews of machine learning for survival analysis have emphasized that these methods can accommodate censoring, ranking objectives, and individualized risk curves [8].

The central thesis of this MDL article is that oncology trial discontinuation can be modeled as a trial-level survival problem using features that capture target class, toxicity burden, and protocol complexity. A trial involving a kinase inhibitor, immune-checkpoint agent, antibody-drug conjugate, or cytotoxic regimen may carry distinct discontinuation hazards because mechanism of action shapes expected efficacy and toxicity profiles [9]. Protocol design features, including endpoint burden, amendment likelihood, and eligibility complexity, provide a complementary operational signal that may alter the probability of remaining active [10, 11]. A deep survival model could learn interactions among these biological, safety, and design features to produce trial-specific survival curves rather than static go/no-go scores. **Table 1** summarizes how the proposed deep survival framing converts oncology trial discontinuation from a static portfolio outcome into a time-indexed prediction problem informed by biological, safety, and protocol-design signals.

Table 1. Value of a Deep Survival Framing for Oncology Trial Discontinuation Prediction

Modeling element	Conventional portfolio-risk view	Added value in the proposed MDL framework
Outcome definition	Trial failure or continuation treated as a fixed binary endpoint	Discontinuation modeled as a time-to-event process with censoring
Risk timing	Risk estimated mainly before or at trial launch	Hazard can be updated as trial conduct and evidence evolve
Feature interaction	Limited handling of non-linear relationships among trial attributes	Learns interactions among target class, toxicity burden, and protocol complexity
Clinical interpretation	Produces general feasibility or success-risk scores	Generates trial-specific survival curves that support monitoring and prioritization
Decision relevance	Supports broad go/no-go judgment	Supports earlier identification of trials needing safety review, protocol adjustment, or accrual intervention

Background

Oncology Trial Discontinuation: Causes and Consequences

Oncology trial discontinuation may result from lack of efficacy, safety concerns, inadequate enrollment, sponsor strategy, or external competition, and these causes can unfold at different points in the trial lifecycle. Studies of oncology trial termination have shown that the reasons for discontinuation are heterogeneous, making a single binary prediction target less informative than a time-to-event representation [1]. Immune-checkpoint inhibitor trials illustrate how rapidly expanding target classes can create both opportunity and discontinuation risk when evidence, competition, and toxicity profiles evolve simultaneously [12]. In this setting, a survival model should estimate not only whether a trial might discontinue but also when the discontinuation hazard is expected to intensify.

Role of Drug Target Class in Trial Risk

Drug target class is a clinically meaningful feature because mechanism of action influences expected response patterns, dose-limiting toxicities, biomarker requirements, and competitive positioning. Machine learning models that use drug bioactivity and mechanism-related features have shown how pharmacological representations can inform trial outcome prediction [13]. Oncology-specific modeling has also suggested that drug and disease attributes can help anticipate randomized trial outcomes when represented in a structured predictive framework [9]. For trial discontinuation, target-class embeddings could encode similarities among kinase inhibitors, monoclonal antibodies, immune-oncology agents, and antibody-drug conjugates without requiring a purely manual risk hierarchy.

Toxicity Signals in Clinical Development

Toxicity signals are central to oncology development risk because early adverse events, serious adverse events, dose-limiting toxicities, and treatment-related deaths can alter dose selection, enrollment, regulatory confidence, and sponsor willingness to continue. Reviews of drug development failure identify safety as a recurrent contributor to attrition and emphasize that toxicity must be interpreted jointly with efficacy and therapeutic context [4]. In oncology, premature discontinuation in the immune-checkpoint inhibitor era demonstrates that enthusiasm for a mechanism does not eliminate safety- or feasibility-related

termination risk [12]. A survival model could therefore treat emerging toxicity as a time-varying signal that updates the predicted probability of remaining active.

Protocol Complexity and Operational Burden

Protocol complexity adds an operational dimension to discontinuation risk because a trial with many endpoints, arms, visits, procedures, and eligibility restrictions can face slower enrollment, higher site burden, and greater amendment pressure. Benchmarking studies have documented growth in protocol design complexity and have linked complex designs to measurable burdens in development operations [10, 11]. More recent work on protocol amendment experience in oncology trials highlights how design burden can continue to reshape trial execution after activation [14]. A protocol complexity index in a survival model would therefore function as a structured proxy for operational fragility rather than as a purely administrative descriptor.

Deep Survival Analysis and Its Use in Healthcare

Deep survival analysis extends conventional time-to-event modeling by learning non-linear feature interactions, flexible latent representations, and individualized survival functions. DeepSurv demonstrated how neural networks can be trained with a Cox-style survival objective [5], while DeepHit and Dynamic-DeepHit introduced discrete-time and longitudinal approaches suitable for competing outcomes and updating risk [6, 7]. Reviews of deep survival learning have emphasized that these models are increasingly used when proportional hazards assumptions are restrictive or when high-dimensional covariates are important [15, 16]. For oncology trial risk analytics, the same modeling principles can be redirected from patient-level prognosis to trial-level discontinuation prediction.

Model Development Overview

High-Level Modeling Pipeline

The proposed modeling pipeline would define each oncology trial as an observational unit with a start date, an observed discontinuation or completion status, and a follow-up time subject to censoring. Static features would include target class, phase, sponsor type, disease setting, protocol design variables, and prior evidence, while time-updated features would include emerging toxicity signals and operational changes. Neural survival frameworks such as pseudo-value neural survival models [17] and deep Cox mixture models [18] illustrate how flexible architectures can estimate survival-related quantities without reducing the task to simple classification. The output would be a trial-specific survival curve representing the conditional probability that the trial remains active over successive time intervals.

Core Input Feature Groups

The first input feature group would represent drug target class through one-hot variables, ontology-derived labels, or learned embeddings that place related mechanisms near one another in latent space. The second group would summarize toxicity signals, including normalized rates of dose-limiting toxicities, serious adverse events, treatment-related discontinuations, and safety-related protocol modifications, interpreted relative to phase and population. The third group would quantify protocol complexity using endpoint count, arm structure, eligibility criterion burden, visit schedule density, and amendment-related indicators, reflecting benchmarks from protocol complexity research [19, 20]. Together, these feature groups would allow the model to learn whether particular combinations of mechanism, toxicity, and operational burden are associated with elevated discontinuation hazard.

Design Principles

The model should be designed around a time-to-event objective, because ongoing trials are right-censored and discontinued trials have event times that carry more information than a terminal label. Clinical trial outcome prediction systems such as HINT show how trial attributes, drug information, and disease context can be combined into machine learning pipelines for development decision support [21]. TrialBench further illustrates the importance of AI-ready multimodal trial datasets that can support standardized model development and comparison [22]. For oncology discontinuation, the design principles should therefore emphasize censoring-aware learning, modular feature engineering, dynamic toxicity updates, and interpretable outputs for portfolio decision-makers.

Data Sources and Feature Engineering

Compilation of a Trial-Level Survival Dataset

A trial-level survival dataset would combine structured registry data, published trial reports, sponsor disclosures, and linked safety evidence to define follow-up time, discontinuation status, and reason for termination. ClinicalTrials.gov provides core fields such as trial phase, start and completion dates, recruitment status, arms, interventions, endpoints, and eligibility criteria, but its quality limitations require careful curation [23]. Literature-derived safety tables and registry results can enrich toxicity features, while external commercial sources such as PharmaProjects or Citeline could support target-class and portfolio-history annotations when available. Predictive modeling studies of clinical trial termination demonstrate that registry-derived variables and text-derived features can be transformed into structured inputs for discontinuation risk analysis [24, 25].

Encoding Target Class and Toxicity

Target class could be encoded by manually curated categories or by ontology-based mapping from intervention names to mechanisms, with embeddings used when related mechanisms share biological or development-risk properties. Toxicity features could be extracted from early-phase publications, ClinicalTrials.gov results tables, or structured adverse-event summaries, then normalized by phase, tumor type, line of therapy, and exposure duration. Text-mining approaches to trial termination and latent topic modeling show that unstructured registry content can contribute meaningful predictors when transformed into analyzable representations [25, 26]. For oncology models, toxicity should be represented as both a baseline expectation for the target class and an emerging signal that can update predicted discontinuation risk.

Protocol Complexity Quantification

Protocol complexity can be quantified as a composite index that reflects the number of endpoints, number of treatment arms, eligibility criterion volume, visit intensity, procedure burden, and amendment-prone design elements. Machine learning analyses of large trial corpora have shown that clinical trials are becoming more complex, supporting the need for explicit complexity features rather than treating protocol burden as background noise [27]. The Chia corpus demonstrates that eligibility criteria can be annotated and structured at scale, making it feasible to extract inclusion and exclusion burden from trial text [28]. Missing protocol details would require transparent imputation strategies, because incompleteness in public registries can otherwise bias the learned relationship between design burden and discontinuation risk [23].

Table 2 defines the trial-level feature architecture required to convert oncology trial discontinuation into a censoring-aware survival prediction problem rather than a static classification task.

Table 2. Trial-Level Feature Architecture for Deep Survival Prediction of Oncology Trial Discontinuation

Feature domain	Representative variables	Survival-model role	Expected discontinuation-risk signal	Modeling implication
Trial time structure	Trial start date, last observed status date, completion date, discontinuation date, follow-up duration	Defines event time, censoring time, and survival horizon	Determines whether the trial contributes an observed event or right-censored follow-up	Must be encoded as a time-to-event problem rather than a binary termination label
Trial phase and development setting	Phase I, I/II, II, III, tumor type, line of therapy, biomarker-defined population	Contextual adjustment layer	Early-phase and late-phase trials may face different safety, efficacy, recruitment, and strategic risks	Stratification or interaction terms are needed to avoid comparing unlike trial contexts
Drug target class	Kinase inhibitor, immune-checkpoint agent, antibody–drug conjugate, monoclonal antibody, cytotoxic regimen, cellular therapy	Mechanism-informed baseline risk representation	Target class may shape toxicity burden, efficacy uncertainty, biomarker dependence, and competition intensity	One-hot, ontology-based, or embedding-based representation can capture similarities across mechanisms
Toxicity signal burden	Serious adverse events, dose-limiting toxicities, treatment-related discontinuations, treatment-related deaths, safety-related amendments	Time-updated risk driver	Emerging toxicity may increase discontinuation hazard before formal trial termination occurs	Should be modeled as both baseline expectation and longitudinal safety evidence
Protocol complexity	Endpoint count, eligibility restriction volume, number of arms, visit density, procedure burden, adaptive elements, amendment-prone features	Operational fragility proxy	Complex protocols may slow enrollment, increase site burden, and raise amendment pressure	Composite complexity indices can support non-linear interaction learning with phase and target class
Recruitment and operational status	Enrollment target, recruitment status changes, site activation pattern, accrual delay indicators, amendment timing	Dynamic operational covariate	Slow accrual or repeated operational changes may precede discontinuation	Can be represented through time-windowed covariates or sequence encoders
Sponsor and portfolio context	Sponsor type, prior target-class experience, competing internal programs, external therapeutic competition	Strategic-risk modifier	Trial discontinuation may reflect portfolio reprioritization rather than biological failure alone	Should be interpreted cautiously because strategic drivers may be under-reported
Termination reason category	Toxicity, efficacy failure, recruitment failure, business/strategic termination, external competition, unknown	Competing-risk endpoint	Different discontinuation causes imply different mitigation strategies	Cause-specific survival modeling is preferable when reason labels are reliable
Missingness and reporting quality	Missing toxicity tables, incomplete amendment records, inconsistent registry status, vague termination reasons	Bias and uncertainty indicator	Poor data quality may distort apparent associations between features and discontinuation	Missingness indicators, transparent imputation, and sensitivity analyses are required
Interpretability features	SHAP summaries, attention weights, feature-attribution	Explanation layer for portfolio users	Helps identify whether predicted risk is driven by	Explanations should be framed as predictive

trajectories, risk-contribution profiles	toxicity, protocol burden, target class, or operational signals	associations, not causal proof

Deep Survival Model Architecture

Neural Network Backbone

The neural network backbone would map trial-level covariates into a latent representation that captures non-linear interactions among target class, toxicity, and protocol complexity. A feed-forward DeepSurv-style architecture could be appropriate when the feature set is mostly static [5], while a transformer or recurrent encoder could be considered when longitudinal safety windows and protocol amendments are represented as temporal sequences. Contemporary reviews of artificial intelligence in trial design emphasize that flexible architectures can support feasibility assessment, eligibility optimization, and risk-informed development planning [29, 30]. Regularization, normalization, and conservative validation should be emphasized because trial-level datasets may be heterogeneous and vulnerable to overfitting.

Survival Output Layer

The survival output layer would determine how discontinuation time is represented and how censoring contributes to the learning objective. A Cox-style layer could output a continuous relative risk function when event times are recorded with sufficient temporal resolution, following the principle of neural Cox survival modeling [5]. A discrete-time output layer inspired by DeepHit could instead estimate the probability of discontinuation across ordered intervals and accommodate competing reasons for discontinuation more naturally [6]. Model choice should therefore depend on whether the analytic emphasis is continuous risk ranking, interval-specific discontinuation probability, or cause-specific survival curves.

Figure 1 illustrates the proposed deep survival modeling architecture for transforming oncology trial registry data, target-class features, toxicity signals, and protocol complexity measures into trial-specific discontinuation survival curves and decision-support outputs.

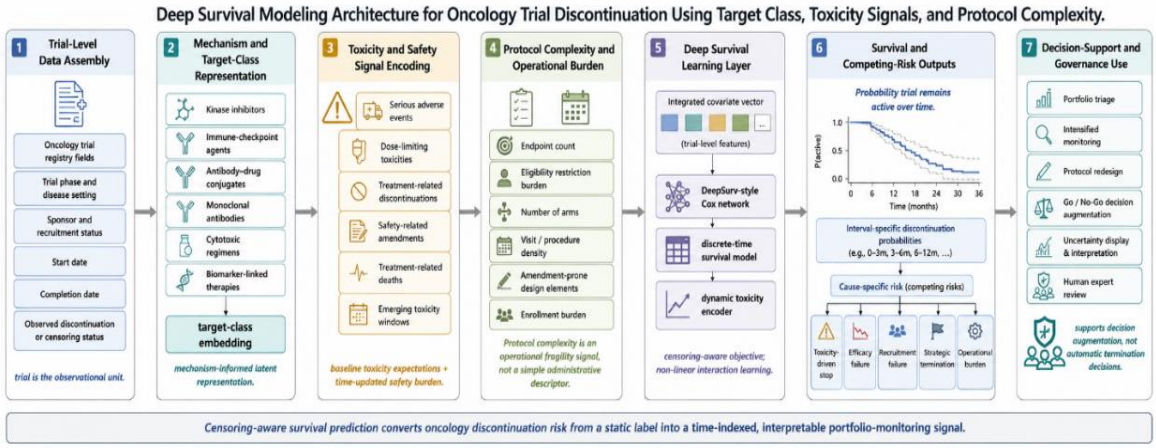

Figure 1. Deep Survival Modeling Architecture for Oncology Trial Discontinuation Using Target Class, Toxicity Signals, and Protocol Complexity

Loss Function and Training

The training objective should be censoring-aware and aligned with the survival output layer rather than optimized as ordinary classification. Cox-based formulations can use a partial-likelihood objective, while discrete-time models can combine survival likelihood with ranking terms that encourage clinically meaningful ordering of trial risk [6, 8]. Dynamic survival architectures support training with updated covariate histories, which is relevant when toxicity evidence accumulates during trial conduct [7]. A recent scoping review of artificial intelligence for clinical trial risk assessment underscores that such models should be developed as decision-support tools whose usefulness depends on interpretability, validation strategy, and alignment with real portfolio workflows [31].

Handling Competing Risks, Censoring, and Time-Varying Features

Competing Reasons for Discontinuation

Oncology trials may discontinue because of toxicity, inadequate efficacy, recruitment failure, strategic reprioritization, or external competitive changes, and these causes should not be collapsed into a single undifferentiated endpoint when the reason is available. DeepHit was designed for competing-risk survival prediction and therefore provides a conceptual template for estimating cause-specific discontinuation probabilities across time intervals [6]. In oncology development, a toxicity-driven stop should be interpreted differently from a business-driven termination, because each has different implications for target-class learning, protocol redesign, and portfolio triage. A cause-specific neural survival model could therefore distinguish whether elevated risk arises from safety liabilities, operational burden, or strategic vulnerability.

Right-Censoring of Ongoing Trials

Ongoing oncology trials at the analysis cutoff should be treated as right-censored observations rather than assumed to be successful or low risk. Survival learning frameworks are designed to use partial follow-up information, allowing active trials to contribute information up to their last observed status without requiring an observed discontinuation event [8]. Clinical trial duration modeling also highlights why elapsed follow-up time contains useful operational information even when the final outcome has not yet occurred [32]. In the proposed model, the predicted survival curve would estimate the probability that an active trial remains open beyond future time points while respecting the uncertainty introduced by censoring.

Incorporating Time-Varying Toxicity Signals

Toxicity should be modeled as a dynamic signal because early safety findings may emerge after enrollment begins and may change the discontinuation hazard before a formal termination decision is recorded. Dynamic-DeepHit provides a conceptual basis for updating survival predictions from longitudinal covariate histories rather than relying only on baseline inputs [7]. A DeepSurv-based clinical trial termination model has also illustrated how neural survival methods can be adapted to trial discontinuation prediction when structured trial features are available [33]. For oncology applications, recurrent, windowed, or attention-based encoders could incorporate toxicity summaries at successive follow-up periods and update the trial-specific survival curve accordingly.

Model Interpretability for Portfolio Decisions

Global Feature Importance via SHAP and Attention

Interpretability is essential because trial discontinuation predictions can affect resource allocation, sponsor governance, and the design of future oncology programs. Interpretable machine learning for early trial termination has shown that predictive models are more useful when feature contributions can be examined rather than treated as opaque risk scores [34]. In the proposed framework, global explanations could identify whether target class, toxicity burden, eligibility complexity, endpoint count, or phase contributes most strongly to predicted discontinuation patterns. Attention weights, integrated gradients, or SHAP-style summaries would be used conceptually to support model inspection, while avoiding the claim that any single feature has a universal causal effect.

Trial-Specific Risk Explanation

A trial-specific explanation would translate the survival curve into decision-relevant language for clinical operations, safety governance, and portfolio committees. For example, the model could indicate that discontinuation risk is being influenced by an emerging serious-adverse-event pattern, a heavily restricted eligibility profile, or a multi-arm design that increases operational burden, while grounding each explanation in the trial's observed feature values. Machine learning models developed for oncology trial outcomes and prostate cancer phase III trials demonstrate the value of trial-specific structured predictors for development risk assessment [9, 35]. Such explanations should be framed as risk-support statements rather than deterministic reasons for discontinuation.

Integration Into Drug Development Decision Support

Portfolio Prioritisation and Monitoring

A deep survival model could support portfolio prioritisation by ranking active oncology trials according to their predicted probability of remaining active over clinically relevant future periods. Rather than replacing expert judgment, the model would identify trials where toxicity signals, protocol burden, or target-class history warrant intensified monitoring or protocol reassessment. HINT demonstrates how structured drug, disease, and trial information can be integrated into clinical-trial-outcome prediction systems [21], and broader artificial intelligence reviews describe similar opportunities for improving trial design and execution [29]. In portfolio use, the model's value would come from making risk visible early enough for operational or scientific mitigation.

Go/No-Go Decision Augmentation

Go/no-go decisions in oncology development require balancing expected efficacy, safety tolerability, operational feasibility, and the probability that a trial will survive long enough to generate interpretable evidence. A survival curve for trial continuation could complement separate efficacy or probability-of-success models by representing the time-dependent feasibility of reaching decision-enabling milestones. Clinical trial success-rate modeling provides aggregate development benchmarks [3], while drug-development failure reviews emphasize that attrition reflects interacting scientific, safety, and execution factors rather than a single isolated cause [4]. The proposed model would therefore function as a decision-augmentation layer that contextualizes efficacy expectations within discontinuation risk.

Evaluation Strategy

Survival Prediction Performance

Evaluation should focus on whether the model ranks and times discontinuation risk appropriately, using survival-specific metrics rather than ordinary classification accuracy. Concordance, time-dependent discrimination, and prediction-error

measures are commonly discussed in survival analysis because they account for event timing and censoring [8]. Deep survival reviews emphasize that performance assessment should match the modeling target, especially when predictions are individualized survival functions rather than static labels [15]. For oncology discontinuation, these measures should be applied conceptually to assess whether higher-risk trials are identified earlier and whether predicted survival trajectories are clinically plausible. **Table 3** outlines a compact evaluation plan for assessing whether the deep survival model provides reliable, time-aware, and decision-relevant discontinuation predictions rather than merely ranking oncology trials by static risk.

Table 3. Compact Evaluation Framework for Oncology Trial Discontinuation Survival Prediction

Evaluation dimension	What it assesses	Why it adds value to this study
Discrimination	Whether higher-risk trials discontinue earlier than lower-risk trials	Tests whether the model ranks trial discontinuation risk appropriately
Time-dependent accuracy	Whether prediction quality changes across follow-up time	Captures whether risk prediction remains useful as trials progress
Prediction error	Difference between predicted and observed survival outcomes	Assesses the practical reliability of individualized survival curves
Calibration	Agreement between predicted and observed discontinuation probabilities	Prevents overconfident or misleading portfolio decisions
Uncertainty display	Width and stability of confidence or uncertainty bands	Helps decision-makers distinguish strong risk signals from uncertain forecasts

Calibration and Uncertainty

Calibration is necessary because a model that ranks trials correctly may still overstate or understate the probability of discontinuation at a given time point. Neural survival methods that output survival probabilities should therefore be evaluated with calibration plots, uncertainty bands, and subgroup checks across phase, target class, sponsor type, and tumor setting [16]. Pseudo-value approaches to neural survival modeling illustrate one pathway for estimating survival quantities that can be compared with observed time-dependent outcomes [17]. For decision support, uncertainty should be displayed alongside risk curves so that portfolio teams do not interpret model outputs as precise forecasts.

Prospective Validation and Comparison against Baseline

Prospective validation should be designed around temporal generalization, such as training on earlier oncology trials and evaluating predictions on later trials whose outcomes were not available during model development. Baseline comparisons should include simpler Cox models, logistic termination models, and interpretable machine learning approaches using the same target-class, toxicity, and complexity features [24, 34]. TrialBench underscores the importance of standardized AI-ready datasets for fair comparison across clinical trial prediction methods [22]. A deep model should be considered useful only if it provides better decision-relevant survival characterization, interpretability, or updating behavior than these simpler alternatives.

Table 4 consolidates the model-design, validation, interpretability, and governance requirements needed to make deep survival predictions usable for oncology portfolio decision support.

Table 4. Model Design, Validation, and Decision-Use Framework for Oncology Trial Discontinuation Survival Modeling

Analytical component	Recommended design choice	Why it strengthens the framework	Key evaluation criterion	Decision-support implication
Prediction target	Time to trial discontinuation, with active trials treated as right-censored	Preserves timing information and avoids misclassifying ongoing trials as successes or failures	Correct handling of censoring and follow-up duration	Produces a survival curve instead of a single termination score
Survival model family	DeepSurv-style Cox model, discrete-time neural survival model, or dynamic survival architecture	Allows non-linear interactions among target class, toxicity, and protocol complexity	Concordance, time-dependent discrimination, prediction error, and calibration	Enables ranking of vulnerable trials across clinically meaningful time horizons
Competing-risk structure	Cause-specific or discrete-time competing-risk model when termination reason is available	Separates toxicity-driven discontinuation from efficacy, recruitment, operational, or strategic discontinuation	Cause-specific calibration and discrimination	Supports targeted mitigation rather than generic risk escalation
Time-varying toxicity handling	Windowed, recurrent, attention-based, or dynamic survival encoder	Allows new safety evidence to update	Improvement in dynamic prediction	Flags trials whose risk profile changes before formal discontinuation

		survival predictions during trial conduct	accuracy after toxicity updates	
Protocol complexity modeling	Composite index plus individual components such as endpoint count, arms, eligibility burden, visit density, and amendment signals	Captures operational burden as a structured risk source	Subgroup calibration by phase, tumor setting, and protocol burden level	Identifies redesign opportunities before operational fragility becomes irreversible
Baseline comparison	Cox proportional hazards model, logistic termination model, random survival forest, and interpretable machine-learning baseline	Tests whether deep learning adds value beyond simpler methods	Incremental gain in discrimination, calibration, updating behavior, or interpretability	Prevents overuse of deep models when simpler models are sufficient
Interpretability strategy	Global feature importance plus trial-specific risk explanation	Makes predictions usable for portfolio committees, safety teams, and protocol designers	Stability and plausibility of feature attributions across validation folds	Converts model output into actionable monitoring or redesign rationales
Calibration and uncertainty	Time-specific calibration curves, uncertainty intervals, subgroup checks, and sensitivity analyses	Prevents overconfident survival estimates in sparse or biased datasets	Calibration across phase, target class, sponsor type, and tumor setting	Encourages cautious use of model output in high-stakes development decisions
Prospective validation	Temporal validation using earlier trials for training and later trials for evaluation	Tests real-world generalizability under changing oncology development conditions	Prospective discrimination, calibration drift, and external validity	Determines whether the model is suitable for live portfolio monitoring
Governance boundary	Human-in-the-loop review, audit trail, model-use documentation, and non-automatic decision policy	Protects against treating predictive risk as deterministic evidence of failure	Documentation completeness, review consistency, and decision traceability	Positions the model as decision augmentation, not autonomous trial termination

Limitations

Under-Reporting and Data Quality

The proposed framework would be constrained by incomplete and uneven reporting of clinical trial status, safety outcomes, protocol amendments, and reasons for discontinuation. Data quality concerns at ClinicalTrials.gov demonstrate that registry fields may be missing, inconsistent, or difficult to harmonize across studies [23]. Negative or terminated oncology trials may be especially vulnerable to sparse publication of detailed toxicity tables, creating selection bias in linked safety features. Any implementation should therefore treat the model as dependent on careful data curation rather than as a purely automated registry-mining system.

Causal Attribution of Discontinuation

A second limitation is that discontinuation labels may not identify the true causal pathway leading to trial termination. A trial recorded as stopped for business reasons may have been influenced by emerging toxicity, weak efficacy, slow enrollment, or competitive displacement, and these factors may be difficult to disentangle retrospectively. Text-mining and topic-modeling studies of termination reasons show that public records can contain useful signals but may still require interpretation and harmonization before modeling [25, 26]. The model should therefore be presented as predictive rather than causal unless supported by stronger adjudication of discontinuation reasons.

Conclusion

A deep survival model for oncology trial discontinuation would frame trial failure risk as a time-to-event problem rather than a static classification task. By combining target class, toxicity signals, and protocol complexity, the model could generate trial-specific survival curves that describe the evolving probability of remaining active.

The main strength of this approach is its alignment with how development risk unfolds over time. It could incorporate baseline protocol and mechanism features, update predictions when safety data emerge, and provide interpretable outputs for clinical development teams and portfolio committees.

Important challenges remain in data quality, cause-of-discontinuation labeling, and the treatment of competing risks. Prospective validation would be essential before such a model could be trusted in real-time governance or investment decisions.

Progress will require collaborative data sharing across sponsors, public registries, safety databases, and published trial-result sources. Larger and more representative training sets would make it possible to benchmark trial risk prediction models and improve the reliability of oncology development decision support.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Zhang E, DuBois SG. Early termination of oncology clinical trials in the United States. Cancer Med. 2023;12(5):5517-25.
2. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun. 2018;11:156-64.
3. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019;20(2):273-86.
4. Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B. 2022;12(7):3049-62.
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. In: 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. 2020;67(1):122-33.
8. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. ACM Comput Surv. 2019;51(6):1-36.
9. Schperberg AV, Boichard A, Tsigelny IF, Richard SB, Kurzrock R. Machine learning model to predict oncologic outcomes for drugs in randomized clinical trials. Int J Cancer. 2020;147(9):2537-49.
10. Getz KA, Campo RA. Trial watch: trends in clinical trial design complexity. Nat Rev Drug Discov. 2017;16(5):307-8.
11. Getz KA, Campo RA. New benchmarks characterizing growth in protocol design complexity. Ther Innov Regul Sci. 2018;52(1):22-8.
12. Khunger M, Rakshit S, Hernandez AV, Pasupuleti V, Glass K, Galsky MD, et al. Premature clinical trial discontinuation in the era of immune checkpoint inhibitors. Oncologist. 2018;23(12):1494-9.
13. Murali V, Muralidhar YP, Königs C, Nair M, Madhu S, Nedungadi P, et al. Predicting clinical trial outcomes using drug bioactivities through graph database integration and machine learning. Chem Biol Drug Des. 2022;100(2):169-84.
14. Botto E, Smith Z, Getz K. New benchmarks on protocol amendment experience in oncology clinical trials. Ther Innov Regul Sci. 2024;58(4):645-54.
15. Wiegrebe S, Kopper P, Sonabend R, Bischl B, Bender A. Deep learning for survival analysis: a review. Artif Intell Rev. 2024;57(3):65.
16. Chen GH. An introduction to deep survival analysis models for predicting time-to-event outcomes. Found Trends Mach Learn. 2024;17(6):921-1100.
17. Zhao L, Feng D. Deep neural networks for survival analysis using pseudo values. IEEE J Biomed Health Inform. 2020;24(11):3308-14.
18. Nagpal C, Yadlowsky S, Rostamzadeh N, Heller K. Deep cox mixtures for survival regression. In: Mach Learn Healthc Conf; 2021. p. 674-708.
19. Getz K, Smith Z, Kravet M. Protocol design and performance benchmarks by phase and by oncology and rare disease subgroups. Ther Innov Regul Sci. 2023;57(1):49-56.
20. Willigers BJ, Wiesiolek AE, Potter B, Pszczółkowska A, Kristensson C, Presz V, et al. Development of a protocol complexity tool: a framework designed to stimulate discussion and simplify study design. BMC Med Res Methodol. 2025;25(1):205.
21. Fu T, Huang K, Xiao C, Glass LM, Sun J. Hint: Hierarchical interaction network for clinical-trial-outcome predictions. Patterns. 2022;3(4).
22. Chen J, Hu Y, Cai M, Lu Y, Wang Y, Cao X, et al. TrialBench: Multi-modal AI-ready datasets for clinical trial prediction. Sci Data. 2025;12(1):1564.
23. Chaturvedi N, Mehrotra B, Kumari S, Gupta S, Subramanya HS, Saberwal G. Some data quality issues at ClinicalTrials.gov. Trials. 2019;20(1):378.
24. Elkin ME, Zhu X. Predictive modeling of clinical trial terminations using feature engineering and embedding learning. Sci Rep. 2021;11(1):3446.
25. Follett L, Geletta S, Laugerman M. Quantifying risk associated with clinical trial termination: A text mining approach. Inf Process Manag. 2019;56(3):516-25.

26. Geletta S, Follett L, Laugerman M. Latent Dirichlet Allocation in predicting clinical trial terminations. BMC Med Inform Decis Mak. 2019;19(1):242.
27. Markey N, Howitt B, El-Mansouri I, Schwartzenberg C, Kotova O, Meier C. Clinical trials are becoming more complex: a machine learning analysis of data from over 16,000 trials. Sci Rep. 2024;14(1):3514.
28. Kury F, Butler A, Yuan C, Fu LH, Sun Y, Liu H, et al. Chia, a large annotated corpus of clinical trial eligibility criteria. Sci Data. 2020;7(1):281.
29. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends Pharmacol Sci. 2019;40(8):577-91.
30. Zhang B, Zhang L, Chen Q, Jin Z, Liu S, Zhang S. Harnessing artificial intelligence to improve clinical trial design. Commun Med (Lond). 2023;3(1):191.
31. Teodoro D, Naderi N, Yazdani A, Zhang B, Bornet A. A scoping review of artificial intelligence applications in clinical trial risk assessment. npj Digit Med. 2025;8(1):486.
32. Cho J, Xu Q, Wong CH, Lo AW. Predicting clinical trial duration via statistical and machine learning models. Contemp Clin Trials Commun. 2025;45:101473.
33. Qi H, Yang W, Zou W, Hu Y. A clinical trial termination prediction model based on denoising autoencoder and deep survival regression. Quant Biol. 2024;12(2):205-14.
34. Kavalci E, Hartshorn A. Improving clinical trial design using interpretable machine learning based prediction of early trial termination. Sci Rep. 2023;13(1):121.
35. Beacher FD, Mujica-Parodi LR, Gupta S, Ancora LA. Machine learning predicts outcomes of phase III clinical trials for prostate cancer. Algorithms. 2021;14(5):147.