*Available online* **www.ijpras.com**

**International Journal of Pharmaceutical Research & Allied Sciences, 2025, 14(3):173-183**

https://doi.org/10.51847/iQsnjR1EDE

**Original Article**

**ISSN : 2277-3657**
**CODEN(USA) : IJPRPM**

# Learning from Incomplete Potency Measurements in QSAR Modelling: A Methodological Framework for Censored Activities, Assay Detection Limits, and the Interpretation of Inactive Compounds

**David Chen[1]*, Julia Weber[2], Nadia Benali[3], Thomas Mwangi[4]**

*[1]Department of Pharmaceutical Sciences, National University of Singapore, Singapore.*
*[2]Department of Cheminformatics, Free University of Berlin, Berlin, Germany.*
*[3]Department of Pharmaceutical Chemistry, University of Tunis El Manar, Tunis, Tunisia.*
*[4]Department of Pharmaceutical Sciences, University of Nairobi, Nairobi, Kenya.*

***Email:** david.chen@nus.edu.sg*

## ABSTRACT

*Quantitative structure–activity relationship (QSAR) modelling commonly assumes that reported bioactivity values represent comparable numerical observations. However, experimental potency records frequently contain inequality-qualified measurements, assay detection limits, threshold-dependent inactivity labels, or missing endpoints. Conventional preprocessing may exclude these observations, replace their bounds with numerical values, or convert them into binary classes, thereby changing the information available for learning. This methodological article develops a boundary-condition extension of conventional QSAR reasoning by examining when scalar-response assumptions become inadequate and how incomplete measurements can retain inferential value. Drawing on published bioactivity database, assay-curation, censored-regression, uncertainty-modelling, and molecular prediction research, the analysis distinguishes experimental observability from molecular potency and pharmacological inactivity. A censoring-aware approach is developed around explicit observation operators, preserved inequality direction, assay-context information, likelihood contributions for bounded responses, and separation of measurement uncertainty from model uncertainty. Its applicability depends on reliable endpoint definitions, identifiable measurement limits, sufficiently compatible assays, and defensible assumptions about the observation process. Illustrative comparisons are proposed to distinguish gains from retained censoring information from improvements attributable to assay curation, model complexity, or data selection. The resulting contribution specifies when incomplete potency records support quantitative learning, when they justify only constrained classification, and when their interpretation remains unresolved. Censored observations should therefore be represented according to the information established by the experiment, with predictive advantages treated as hypotheses requiring comparative validation.*

**Key words:** *Quantitative structure–activity relationships, Censored potency measurements, Assay detection limits, Bioactivity data curation, Interval-censored regression, Pharmacological inactivity*

**Received:** 14 July 2025; **Revised:** 17 August 2025; **Accepted:** 22 August 2025

## INTRODUCTION

Quantitative structure–activity relationship modelling connects molecular descriptions to experimental endpoints in order to predict properties of untested compounds. Its apparent statistical simplicity depends on a demanding measurement assumption: each training response should have an interpretable relationship to the property being predicted. Bioactivity repositories contain substantial information about experimental drug–target relationships, yet individual records differ in source, endpoint, assay design, reporting convention, and data quality. ChEMBL

© 2025 The International Journal of Pharmaceutical Research and Allied Sciences (IJPRAS). **Open Access** - This article is under the **CC BY NC SA** license (https://creativecommons.org/licenses/by-nc-sa/4.0).

demonstrates both the scale and heterogeneity of contemporary bioactivity information [1]. BindingDB provides complementary experimentally derived binding measurements and provenance [2]. Curated collections such as Papyrus explicitly address differences among source records and their suitability for predictive modelling [3].

Large chemogenomic datasets also include categorical activity information and heterogeneous experimental endpoints [4]. Such records cannot always be reduced to a single continuous potency variable without additional assumptions. An inequality-qualified result establishes a permissible region for the underlying quantity; a classification establishes membership under a particular decision rule. Neither necessarily identifies an exact concentration–response parameter.

This distinction becomes important when a QSAR model is trained on IC50, Ki, or EC50 observations. IC50 is a concentration associated with a specified inhibitory response, Ki concerns inhibition or binding under an appropriate mechanistic definition, and EC50 describes a half-maximal effect relative to a defined response system. Their biological interpretation depends on assay conditions and pharmacological mechanism. A model cannot recover comparability merely by converting all concentrations to negative logarithmic units.

Existing workflows offer alternatives for handling incomplete records, including exclusion, classification, probabilistic uncertainty treatment, imputation, and censored regression. The methodological question is which information each choice retains and under what experimental conditions its resulting prediction remains defensible.

The central argument developed here is that conventional scalar-response QSAR reasoning operates within an implicit observability condition: a numerical training label must either be sufficiently measured or be represented by a statistical observation model consistent with its experimental reporting. The article examines how that condition changes model construction and interpretation. It advances a nonempirical methodological account, without claiming a newly validated algorithm or a universal improvement in predictive performance.

*Incomplete bioactivity measurements as a pharmacoinformatics problem*

Bioactivity information originates in experiments before becoming a machine-learning target. An individual database record may combine a chemical structure, target identifier, endpoint definition, assay description, response qualifier, unit, and measured or reported value. PubChem's architecture illustrates the coexistence of chemical information and assay-derived activity outcomes [5]. Drug Target Commons likewise highlights the need for meaningful integration and curation of drug–target interaction evidence [6].

A potency record can be incomplete in several scientifically different ways. First, an experiment may report that a parameter exceeds the highest concentration at which it could be resolved. Second, a fitted parameter may fall below the assay's reliable quantification range. Third, only an interval may be justified. Fourth, an assay may classify a compound as inactive without establishing a continuous potency parameter. Finally, no relevant experiment may have been performed, leaving the endpoint genuinely missing. These situations encode different knowledge states and require different statistical treatment.

Consider a report of IC50 greater than a tested concentration $c$. On the concentration scale, the result supplies a lower bound. After conversion to $pIC50 = -\log10(IC50 \text{ expressed in molar concentration})$, it supplies an upper bound. Conversely, IC50 less than $c$ implies a lower bound on pIC50. Thus, ordinary descriptions of left- and right-censoring require an explicitly declared response scale. Reversing the inequality incorrectly can turn a weakly supported activity estimate into an apparently potent observation.

The existence of a concentration limit also requires experimental interpretation. Failure to achieve 50% inhibition within the tested range may justify a bound under an appropriate dose–response model, but an incomplete, nonmonotonic, cytotoxic, or technically compromised response curve may not identify the same underlying parameter. A missing IC50 estimate may consequently reflect limited exposure range, insufficient efficacy, assay interference, or unsuccessful fitting. Those alternatives cannot be distinguished from the inequality symbol alone.

Assay heterogeneity adds a separate source of uncertainty. Landrum and Riniker demonstrated that merging IC50 or Ki values from different assays can introduce substantial disagreement even where compounds and nominal targets correspond [7]. Matched molecular comparisons may be more robust to some inter-assay differences than absolute potency values, but this depends on the form of systematic variation [8]. Such evidence concerns measurement compatibility and does not itself demonstrate that censoring caused the discrepancies.

Biological context may also affect predictive relationships. Assay-aware modelling has examined the use of assay descriptors to differentiate experimental environments [9]. Combining molecular structures with bioassay descriptions provides another route for incorporating contextual information into prediction [10]. These

approaches identify useful sources of variation, although neither automatically reconstructs an unreported censoring threshold.

Incomplete potency must therefore be understood at two levels: the *experimental information state*, specifying what the assay establishes, and the *pharmaceutical interpretation*, specifying what that information can legitimately imply about molecular activity. A censored value may retain useful ordering information while providing insufficient evidence for an exact potency difference. An inactive designation can remain operationally useful without establishing an absence of molecular interaction.

*Existing evidence and methods for handling censored potency measurements*

QSAR preparation pipelines commonly transform heterogeneous experimental records into modelling-ready datasets. The AMPL platform demonstrates the importance of reproducible preprocessing, molecular representation, model assessment, and uncertainty considerations in a drug-discovery workflow [11]. Nevertheless, reproducible preprocessing does not automatically make discarded observations scientifically irrelevant.

One defensible strategy is to train quantitative models only on exact potency values. ProfhEX provides an explicit example of excluding inequality-qualified measurements during the construction of activity datasets [12]. This strategy avoids assigning artificial scalar values to censored observations, but it changes the training population if censoring is associated with weak activity, assay design, or the accessible concentration range. Complete-case learning can then describe the subset of measurable compounds more effectively than the intended chemical population.

A second strategy changes the response representation. Threshold-based activity classification can incorporate observations whose exact potency is unknown, provided that their reported bound permits unambiguous assignment relative to the chosen cutoff. Probabilistic Random Forest research has shown the relevance of measurement uncertainty near classification thresholds [13]. Its treatment of uncertain class membership should, however, be distinguished from a likelihood representing a known bounded continuous response.

A third strategy incorporates the reporting constraint directly into regression. Svensson and colleagues examined regression models capable of learning from censored labels while quantifying uncertainty [14]. This supplies a particularly relevant methodological precedent: censored observations can contribute information without being treated as exact measurements. The strength of that precedent remains conditional on the evaluated data, model families, and observation assumptions.

Related biological prediction problems further illustrate the importance of endpoint-specific representation. Deep-learning work on HIV sensitivity to monoclonal antibodies demonstrates quantitative and categorical prediction within a different biological system [15]. Its applicability to small-molecule QSAR is limited by differences in molecular features, experimental biology, and prediction objectives. Meanwhile, QSAR-based imputation provides a comparator for predicting unmeasured assay endpoints [16]. A missing endpoint and a censored endpoint are not equivalent: the former lacks a direct response observation, while the latter can constrain a latent response.

The information retained by these reporting conventions determines which statistical operations are justified. **Table 1** distinguishes measurable potency, partial bounds, categorical decisions, and absent observations.

**Table 1.** Experimental information retained by common forms of incomplete bioactivity reporting

| Reported observation | Experimental information established | Admissible potency interpretation | Information unavailable | Appropriate modelling representation |
|---|---|---|---|---|
| **Exact IC50, Ki, or EC50 estimate** | Numerical endpoint under specified assay and fitting conditions | Approximate measured potency, subject to error | Perfect biological comparability across assays | Continuous response with suitable error model |
| **IC50 greater than** *c* | Endpoint exceeds reported concentration boundary, if the parameter is identifiable | Lower bound on IC50; upper bound on pIC50 | Exact potency beyond the boundary | Censored continuous response |
| **IC50 less than** *c* | Endpoint lies below a stated measurement boundary | Upper bound on IC50; lower bound on pIC50 | Exact potency below the boundary | Censored continuous response |

| | | | | |
|---|---|---|---|---|
| **Potency bounded between *a* and *b*** | Parameter lies within an experimentally supported interval | Interval-constrained potency | Position within the interval | Interval-censored likelihood |
| **"Inactive" under a documented screening rule** | Failure to satisfy the specified activity criterion | Assay- and cutoff-specific categorical decision | Exact IC50 or absence of binding | Classification; bounded regression only if justified |
| **Missing or uninterpretable endpoint** | No reliable numerical or bounding response is established | No direct potency estimate | Endpoint value and potentially its measurement process | Missing-data treatment or exclusion with audit trail |

Notes: The table distinguishes recorded observation semantics from proposed statistical use. The interpretation of inequality records presumes that the endpoint and experimental limit are meaningful. For negative-log potency, inequality direction reverses. A flat response curve need not yield a valid censored IC50. The final column identifies methodological options, not empirically demonstrated superiority.

*Information loss, statistical bias, and pharmacological misinterpretation*

Information loss occurs whenever a transformation discards distinctions that could affect the model's estimand. Removing censored records can alter the distribution of molecular structures and potencies represented during training. Replacing a reported bound with its cutoff introduces artificial precision. Converting all responses to binary labels can erase quantitative ordering among compounds measured at different concentrations. These consequences differ mathematically, even where their resulting training tables have similar dimensions.

The practical importance of incomplete observations is consistent with research on data valuation in high-throughput screening, where the contribution of training examples depends on their relevance to subsequent prediction and selection decisions [17]. However, a compound's apparent training value cannot establish whether a missing or censored potency reflects a valid biological endpoint. Data quality must be assessed before usefulness is inferred from algorithmic influence.

An especially important rival explanation is experimental interference. Screening technologies can produce signals affected by compound properties or assay-specific artifacts, and research has developed methods for identifying compounds associated with such interference [18]. A compound labelled active or inactive under an unreliable assay configuration may therefore require further measurement, not simply a different treatment of censoring. Retaining an untrustworthy bound can propagate misleading information as readily as substituting an exact number.

Incorrect treatment of bounds may also distort structure–activity comparisons. Activity cliffs describe marked potency changes between structurally related compounds and are a recognized challenge for prediction [19]. If one member of a molecular pair is reported only as exceeding an assay boundary, its true potency difference from the other remains partially identified. A predicted cliff magnitude based on an arbitrary replacement value may therefore have no defensible quantitative interpretation.

Chemical identity introduces another independent source of disagreement. Automated QSAR-ready structure standardization addresses differences in molecular representations that can otherwise compromise model construction [20]. Yet standardized structures do not resolve whether measurements used different biological targets, endpoint definitions, or concentration ranges. Chemical curation and assay curation protect distinct aspects of the model's validity.

A useful distinction follows between *selection bias*, *response misrepresentation*, and *pharmacological misclassification*. Selection bias may arise when only compounds with numerically estimable potency are retained. Response misrepresentation arises when a bound is assigned an unsupported exact value. Pharmacological misclassification occurs when a screen-level decision is interpreted as proof of biological inactivity or when target inhibition is equated with a clinically meaningful effect.

These errors are not inevitable consequences of censoring. If all observations are measured within an adequate range, an exact-response analysis may be sufficient. If a binary decision is the actual research endpoint, classification can be appropriate. If assay limits or endpoint semantics are unknown, a formal censored likelihood may offer no reliable gain. The relevant test is whether a modelling choice preserves the information warranted by the measurement process while targeting a scientifically meaningful outcome.

*Development of a censoring-aware QSAR methodological framework*

A censoring-aware approach begins by defining the target of prediction before selecting the loss function. Let $Y$ denote latent potency on an explicitly declared scale, $X$ the molecular representation, and $A$ the assay context. A reported activity record is an observation of $Y$ through an experimental reporting process, potentially restricted by a threshold or interval. This separation allows the underlying potency model and the observation process to be specified independently.

The likelihood contribution then depends on what was observed. For an exact response $y$, it is the conditional density $f(y \mid X, A)$. For an observation establishing $Y \leq u$, it is the conditional cumulative probability $F(u \mid X, A)$. For $Y \geq l$, the contribution is $1 - F(l \mid X, A)$, with treatment of strict endpoints governed by the response distribution. An interval $l \leq Y \leq u$ contributes $F(u \mid X, A) - F(l \mid X, A)$. These expressions are methodological specifications; they do not constitute results from a newly fitted model.

Reliable implementation requires consistency between the experimental units and the chosen statistical response. For example, an IC50 concentration bound becomes an oppositely directed pIC50 bound after negative-log transformation. A reported threshold should retain its originating assay, units, and response qualifier. If no meaningful potency parameter was established, likelihood specification must not manufacture one.

Censored likelihoods also rely on assumptions about why the observation is bounded. A known assay limit may be incorporated conditionally on its recorded value. When censoring depends on unrecorded experimental failure, selectively chosen concentration ranges, or other unmodelled biological properties, ordinary conditional likelihoods may remain misspecified. Assay-dependent thresholds should consequently be treated as observed design characteristics where possible, with their absence recognised as an identifiability limitation.

Predictive uncertainty forms another component. Comparative research has demonstrated that uncertainty metrics can behave differently across chemical modelling tasks [21]. Explainable uncertainty methods investigate how prediction uncertainty relates to molecular inputs and model behaviour [22]. For censored QSAR, these considerations motivate separate reporting of uncertainty associated with model parameters or extrapolation, latent experimental responses, and censoring-induced partial observation. The separation is proposed as an analytical requirement; no specific decomposition is asserted to be universally identifiable.

Molecular representation must remain appropriate to the chemical question. Learned molecular features may offer predictive advantages under particular modelling settings, but representation studies demonstrate that their performance depends on data and evaluation design [23]. Neither sophisticated descriptors nor a highly flexible model can compensate for an incorrectly interpreted activity qualifier. Where assay descriptions are informative, conditional predictions may be modelled with assay covariates or suitable hierarchical effects. Where the dataset is too sparse to estimate these effects, a more restricted assay-specific model may be preferable.

This account produces a revised boundary condition for ordinary QSAR regression: an exact-response loss is scientifically interpretable when the supplied label is a justified point observation of the intended endpoint; otherwise, the likelihood or modelling objective must reflect the actual information state. **Table 2** compares the available strategies without presuming that the most elaborate option is necessarily optimal.

**Table 2.** Statistical treatment of incomplete potency records under alternative QSAR objectives

| Modelling treatment | Response supplied to learning | Information preserved | Condition for defensible use | Principal inferential vulnerability |
|---|---|---|---|---|
| **Exact-only regression** | Numerically resolved potency subset | Measured values and selected assay metadata | Selection of measurable values does not defeat intended inference, or inference is restricted to that subset | Loss of systematically censored compounds |
| **Boundary substitution** | Assay threshold treated as a number | Threshold location, but not uncertainty beyond it | Only an explicitly justified approximation or sensitivity analysis | Artificial exactness and biased gradients |
| **Threshold classification** | Active/inactive class under cutoff | Decision-relevant class membership | Class is unambiguous from the reported bound and cutoff | Loss of continuous potency resolution |
| **Probabilistic class modelling** | Probability-weighted labels | Uncertainty near classification cutoff | Class uncertainty has an interpretable specification | Does not generally model continuous censoring |

| | | | | |
|---|---|---|---|---|
| **Censored-response likelihood** | Density or tail/interval probability | Direction and extent of known response restriction | Endpoint, bounds, observation process and likelihood are valid | Misspecification or unidentified tails |
| **Missing-endpoint imputation** | Predicted response for unmeasured endpoint | Correlations among observed assay tasks | Missingness mechanism and model assumptions are justified | Imputation uncertainty mistaken for measurement |
| **Assay-aware censored model** | Bounded response conditional on assay features | Censoring information and recorded experimental context | Assay descriptors and sufficient data support estimation | Confounding, sparse assay coverage and context extrapolation |

Notes: Entries describe published methodological families and a proposed combined application. A documented bound supplies a likelihood constraint; an unmeasured endpoint does not. Censored-response modelling may be implemented using parametric, Bayesian, ensemble-based, or other estimators consistent with the observation model. No relative performance ranking is implied.

The resulting conceptual structure is branching because each information state permits a different modelling decision, while assay conditions can qualify several branches simultaneously. A lower or upper potency bound remains informative for a censored likelihood; an incomplete response curve may require additional assay assessment; and an unsupported inactivity label may permit only limited categorical inference. The cross-connections among observation, assay context, model choice, and downstream interpretation are represented in **Figure 1**.

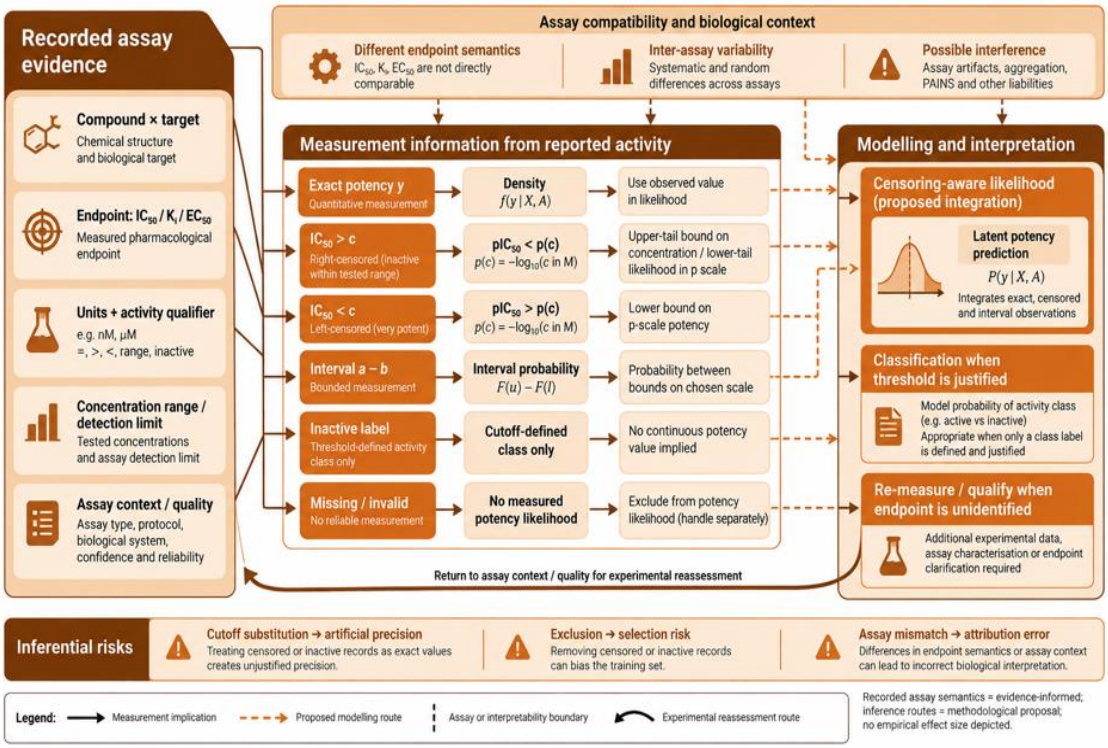



**Figure 1.** Potency information remains inferentially usable only when observation bounds, assay context, and modelling likelihood agree.

Branching information-preservation map showing how recorded assay evidence constrains QSAR model specification and pharmacological interpretation. The left region identifies the molecular target, endpoint, reporting qualifier, experimental concentration range, and assay context. The central region distinguishes exact potency, upper- and lower-bounded potency, interval observations, categorical inactivity, and missing or invalid measurements. Solid arrows represent observation semantics or mathematical implications conditional on valid measurement definitions; dashed arrows represent proposed modelling decisions requiring empirical validation. The right region separates continuous-response inference, justified classification, and circumstances requiring experimental reassessment. The upper band indicates that endpoint compatibility, assay variability, and technical

interference can qualify several pathways simultaneously. The lower band identifies the risks of artificial numerical precision, selection effects, and incorrect pharmacological attribution. A return connector represents further measurement or clarification when the endpoint remains unidentified. The figure does not imply that an inactive classification establishes the absence of binding, that a censored potency is an exact measurement, or that censoring-aware regression has universally superior predictive performance. IC50 denotes half-maximal inhibitory concentration; EC50, half-maximal effective concentration; Ki, inhibition constant; QSAR, quantitative structure–activity relationship.

*Implications for model construction, evaluation, and reporting*

Censoring-aware QSAR evaluation requires greater attention to the observation process than conventional point-prediction benchmarking. Each record should retain the original endpoint, numerical unit, inequality qualifier, tested concentration range, source assay, and transformation history. Molecular structure standardization and assay curation should precede response construction. Data splitting must prevent duplicate compounds, closely related experiments, or repeated measurements from leaking across training and test partitions.

The evaluation target should determine the scoring rule. Exact potency observations permit conventional continuous-response metrics when their assumptions are satisfied. Censored observations instead contribute constraints whose predictive probabilities can be evaluated through appropriate likelihood or interval-sensitive scoring procedures. A model may correctly assign high probability to the experimentally justified region while predicting a point estimate that differs from the recorded boundary. Treating the threshold as ground truth would incorrectly penalize this legitimate distinction.

Prediction intervals also require evaluation within a clearly defined applicability domain. Dynamic applicability-domain research has demonstrated local conformal prediction for compound–target affinity estimation [24]. Calibration investigations in structure–activity modelling further show why predictive confidence must be evaluated separately from average prediction accuracy [25]. Neither result guarantees coverage for censored QSAR predictions: calibration procedures must account for the censoring mechanism and the information used to construct their conformity scores.

External validation should test simultaneous changes in chemical structure, assay conditions, and reporting thresholds. Research on out-of-distribution molecular prediction illustrates that splitting strategy changes what generalization assessments reveal [26]. Applicability-domain studies similarly demonstrate that performance conclusions depend on which compounds are considered suitable for prediction [27]. For incomplete potency records, a particularly informative design would hold out compounds from assays having distinct measurement limits while separately examining chemical novelty. This would clarify whether reported improvements reflect better learning from bounds or transfer across experimental contexts.

**Table 3** specifies the validation evidence required before different claims about incomplete-activity modelling can be considered defensible.

**Table 3.** Validation claims for QSAR predictions trained on incomplete potency measurements

| Intended validation claim | Required held-out evidence | Evaluation operation | Failure signal requiring qualification |
|---|---|---|---|
| **Quantitative potency prediction** | Independently measured exact endpoints | Error and calibration assessment on compatible assays | Good classification but unreliable concentration estimates |
| **Correct treatment of bounds** | Held-out inequality or interval records with documented limits | Probability assigned to experimentally admissible region | Predictions repeatedly inconsistent with known bounds |
| **Generalization across assay limits** | Assays differing in tested ranges or reporting thresholds | Stratified assessment by concentration-limit context | Performance dependent on training-set threshold distribution |
| **Preservation of inactivity information** | Documented screening calls and selected confirmatory experiments | Compare constrained and discarded-label strategies | Inactivity inferred without adequate measurement support |
| **Molecular-domain reliability** | Structurally distinct held-out compounds | Applicability-domain and distribution-shift analysis | Apparent confidence on unsupported chemical regions |

| | | | |
|---|---|---|---|
| **Practical benefit of additional experiments** | Prospectively acquired measurements | Compare remeasurement strategies and subsequent model updating | Uncertainty reduction without relevant decision improvement |

Notes: The table presents proposed validation requirements informed by established QSAR evaluation, applicability-domain, and calibration literature. It does not report completed A2 experiments. Numerical acceptance criteria must be selected prospectively for the intended modelling and pharmaceutical decision context.

*Illustrative methodological scenarios and testable propositions*

Consider a hypothetical screening collection in which weakly active compounds frequently receive upper bounds on pIC50 because their IC50 values exceed the highest tested concentration. Removing these records may enrich the training set for more potent molecules. Replacing every inequality with the assay limit instead produces artificial ties at that boundary. A censored likelihood can retain their partial ordering relative to the limit, provided that the recorded concentration range and endpoint are valid.

Proposition 1: When censoring is associated with latent potency and assay limits are accurately recorded, a correctly specified censored-response model should exhibit better calibration for bounded observations than an otherwise comparable model trained on substituted threshold values. This prediction could be weakened by likelihood misspecification, limited sample size, or a practically irrelevant proportion of censored records.

A second scenario concerns assay-dependent concentration ranges. Compounds appearing inactive in one assay may receive numerical potencies in another without any change in intrinsic molecular structure. An analysis that controls both endpoint compatibility and censoring threshold may distinguish information recovery from assay-associated variation.

Proposition 2: When experiments employ heterogeneous concentration limits, benefits attributed to censored learning should diminish or reverse if the observation model omits relevant assay effects. Matched assay comparisons and held-out experimental contexts would discriminate this explanation from gains attributable solely to additional labelled compounds.

A third scenario concerns experimental acquisition. Uncertainty-based active-learning research shows the relevance of selecting informative observations for molecular-property generalization [28]. Bayesian active-learning approaches provide related methodology for prioritizing experiments in drug discovery [29].

Proposition 3: Where the uncertainty surrounding a bounded response materially affects compound ranking, targeted remeasurement may be more valuable than acquiring another exact observation from an already well-characterized response region. Testing this requires comparing subsequent decision performance, not merely the reduction in model uncertainty.

These propositions are methodological hypotheses. Their verification requires controlled datasets, documented observation rules, prespecified comparators, and independent experimental evidence.

*Discussion and future methodological research*

The argument extends conventional QSAR methodology by making the experiment's ability to identify a response part of the conditions under which prediction is scientifically interpretable. This revision has consequences for benchmark construction. MoleculeNet established influential practices for comparing molecular-learning datasets, representations, and evaluation procedures [30]. A censoring-aware benchmark would additionally need to preserve source qualifiers, concentration limits, assay provenance, and the distinction between absent and partially observed labels.

Such benchmarks should compare censoring-aware likelihoods against complete-case regression, threshold substitution, justified classification, and imputation, while holding molecular features and evaluation partitions constant where feasible. Differences between methods must be interpreted together with assay quality, censoring prevalence, model flexibility, and the completeness of reporting metadata.

Drug-discovery prioritization offers a consequential application. Machine-learning research on high-throughput screening has demonstrated the importance of separating genuine biological responses from assay interference during hit prioritization [31]. Censoring-aware modelling could complement this objective by identifying compounds whose potency remains insufficiently resolved, although assay-interference detection and censored-response inference address distinct uncertainties.

Future investigations should establish whether explicitly retained bounds improve decision calibration, whether gains persist under prospective assay transfer, and whether remeasuring censored compounds produces meaningful changes in compound selection. Negative findings would be valuable: where censoring contributes little additional information or its reporting mechanism is unreliable, simpler methods may remain preferable.

*Limitations and applicability boundaries*

This article presents a methodological synthesis and boundary-condition extension, not an empirically validated QSAR algorithm. Its likelihood formulations assume that the latent endpoint has a meaningful definition and that a recorded bound corresponds to an interpretable experimental restriction. These conditions may fail for incomplete dose–response curves, heterogeneous endpoints, assay interference, or unrecorded selection mechanisms.

A correctly specified censored likelihood cannot overcome unidentified biological mechanisms, severe data scarcity, or structural extrapolation beyond the training domain. If assay thresholds are missing or selected through an unobserved process associated with response values, the assumed observation probabilities may be incorrect. Likewise, an inactivity label without a recoverable decision rule cannot establish a continuous potency interval.

The conceptual extension is most applicable when measurement qualifiers and assay contexts are sufficiently documented to support a coherent observation model. Whether this approach improves prediction, calibration, or compound prioritization remains an empirical question.

## CONCLUSION

Incomplete potency observations can retain scientifically useful information when their experimental meaning is preserved. An inequality-qualified IC50, Ki, or EC50 record establishes a constraint, whereas a screening inactivity label establishes a context-dependent decision and a missing endpoint may supply no direct potency evidence. These information states warrant different modelling assumptions.

The principal methodological implication is that QSAR regression should align its response likelihood with what the assay actually establishes. This requirement protects against artificial numerical precision while permitting partial measurements to contribute to learning. Its practical benefits depend on assay compatibility, reliable limits, identifiable observation mechanisms, and independent validation. Censoring-aware QSAR is therefore a justified methodological option under explicit conditions, with superiority over simpler alternatives remaining a falsifiable proposition.

**ACKNOWLEDGMENTS:** None

**CONFLICT OF INTEREST:** None

**FINANCIAL SUPPORT:** None

**ETHICS STATEMENT:** None

## REFERENCES

1. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024;52(D1):D1180-92. doi:10.1093/nar/gkad1004
2. Liu T, Hwang L, Burley SK, Nitsche CI, Southan C, Walters WP, et al. BindingDB in 2024: A FAIR knowledgebase of protein-small molecule binding data. Nucleic Acids Res. 2025;53(D1):D1633-44. doi:10.1093/nar/gkae1075
3. Béquignon OJM, Bongers BJ, Jespers W, IJzerman AP, van der Water B, van Westen GJP. Papyrus: A large-scale curated dataset aimed at bioactivity predictions. J Cheminform. 2023;15(1):3. doi:10.1186/s13321-022-00672-x
4. Sun J, Jeliazkova N, Chupakhin V, Golib-Dzib JF, Engkvist O, Carlsson L, et al. ExCAPE-DB: An integrated large scale dataset facilitating big data analysis in chemogenomics. J Cheminform. 2017;9(1):17. doi:10.1186/s13321-017-0203-5
5. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1):D1373-80. doi:10.1093/nar/gkac956
6. Tang J, Tanoli ZUR, Ravikumar B, Alam Z, Rebane A, Vähä-Koskela M, et al. Drug Target Commons: A community effort to build a consensus knowledge base for drug-target interactions. Cell Chem Biol. 2018;25(2):224-9.e2. doi:10.1016/j.chembiol.2017.11.009

7. Landrum GA, Riniker S. Combining IC50 or Ki values from different sources is a source of significant noise. J Chem Inf Model. 2024;64(5):1560-7. doi:10.1021/acs.jcim.4c00049
8. Nelen J, Pérez-Sánchez H, De Winter H, Van Rompaey D. Matched pairs demonstrate robustness against inter-assay variability. J Cheminform. 2025;17(1):8. doi:10.1186/s13321-025-00956-y
9. Schoenmaker L, Sastrokarijo EG, Heitman LH, Beltman JB, Jespers W, van Westen GJP. Toward assay-aware bioactivity model(er)s: Getting a grip on biological context. J Chem Inf Model. 2025;65(13):7013-23. doi:10.1021/acs.jcim.5c00603
10. Schuh MG, Boldini D, Sieber SA. Synergizing chemical structures and bioassay descriptions for enhanced molecular property prediction in drug discovery. J Chem Inf Model. 2024;64(12):4640-50. doi:10.1021/acs.jcim.4c00765
11. Minnich AJ, McLoughlin K, Tse M, Deng J, Weber A, Murad N, et al. AMPL: A data-driven modeling pipeline for drug discovery. J Chem Inf Model. 2020;60(4):1955-68. doi:10.1021/acs.jcim.9b01053
12. Lunghini F, Fava A, Pisapia V, Sacco F, Iaconis D, Beccari AR. ProfhEX: AI-based platform for small molecules liability profiling. J Cheminform. 2023;15(1):60. doi:10.1186/s13321-023-00728-6
13. Mervin LH, Trapotsi MA, Afzal AM, Barrett IP, Bender A, Engkvist O. Probabilistic random forest improves bioactivity predictions close to the classification threshold by taking into account experimental uncertainty. J Cheminform. 2021;13(1):62. doi:10.1186/s13321-021-00539-7
14. Svensson E, Friesacher HR, Winiwarter S, Mervin L, Arany A, Engkvist O. Enhancing uncertainty quantification in drug discovery with censored regression labels. Artif Intell Life Sci. 2025;7:100128. doi:10.1016/j.ailsci.2025.100128
15. Dănăilă VR, Buiu C. Prediction of HIV sensitivity to monoclonal antibodies using aminoacid sequences and deep learning. Bioinformatics. 2022;38(18):4278-85. doi:10.1093/bioinformatics/btac530
16. Yang B, Chung Y, Yang AY, Yuan B, Chen T, Yu X. QComp: A QSAR-based imputation framework for drug discovery. J Chem Inf Model. 2025;65(15):7862-73. doi:10.1021/acs.jcim.5c00059
17. Hesse J, Boldini D, Sieber SA. Machine learning-driven data valuation for optimizing high-throughput screening pipelines. J Chem Inf Model. 2024;64(21):8142-52. doi:10.1021/acs.jcim.4c01547
18. David L, Walsh J, Sturm N, Feierberg I, Nissink JWM, Chen H, et al. Identification of compounds that interfere with high-throughput screening assay technologies. ChemMedChem. 2019;14(20):1795-802. doi:10.1002/cmdc.201900395
19. Tamura S, Miyao T, Bajorath J. Large-scale prediction of activity cliffs using machine and deep learning methods of increasing complexity. J Cheminform. 2023;15(1):4. doi:10.1186/s13321-022-00676-7
20. Mansouri K, Moreira-Filho JT, Lowe CN, Charest N, Martin T, Tkachenko V, et al. Free and open-source QSAR-ready workflow for automated standardization of chemical structures in support of QSAR modeling. J Cheminform. 2024;16(1):19. doi:10.1186/s13321-024-00814-3
21. Rasmussen MH, Duan C, Kulik HJ, Jensen JH. Uncertain of uncertainties? A comparison of uncertainty quantification metrics for chemical data sets. J Cheminform. 2023;15(1):121. doi:10.1186/s13321-023-00790-0
22. Yang CI, Li YP. Explainable uncertainty quantifications for deep learning-based molecular property prediction. J Cheminform. 2023;15(1):13. doi:10.1186/s13321-023-00682-3
23. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370-88. doi:10.1021/acs.jcim.9b00237
24. Oršolić D, Šmuc T. Dynamic applicability domain (dAD): Compound-target binding affinity estimates with local conformal prediction. Bioinformatics. 2023;39(8):btad465. doi:10.1093/bioinformatics/btad465
25. Friesacher HR, Engkvist O, Mervin L, Moreau Y, Arany A. Achieving well-informed decision-making in drug discovery: A comprehensive calibration study using neural network-based structure-activity models. J Cheminform. 2025;17(1):29. doi:10.1186/s13321-025-00964-y
26. Fooladi H, Vu TNL, Mathea M, Kirchmair J. Evaluating machine learning models for molecular property prediction: Performance and robustness on out-of-distribution data. J Chem Inf Model. 2025;65(19):9871-91. doi:10.1021/acs.jcim.5c00475
27. Grenet I, Merlo K, Comet JP, Tertiaux R, Rouquié D, Dayan F. Stacked generalization with applicability domain outperforms simple QSAR on in vitro toxicological data. J Chem Inf Model. 2019;59(4):1486-96. doi:10.1021/acs.jcim.8b00553

28. Yin T, Panapitiya G, Coda ED, Saldanha EG. Evaluating uncertainty-based active learning for accelerating the generalization of molecular property prediction. J Cheminform. 2023;15(1):105. doi:10.1186/s13321-023-00753-5
29. Masood MA, Kaski S, Cui T. Molecular property prediction using pretrained-BERT and Bayesian active learning: A data-efficient approach to drug design. J Cheminform. 2025;17(1):58. doi:10.1186/s13321-025-00986-6
30. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-30. doi:10.1039/C7SC02664A
31. Boldini D, Friedrich L, Kuhn D, Sieber SA. Machine learning assisted hit prioritization for high throughput screening in drug discovery. ACS Cent Sci. 2024;10(4):823-32. doi:10.1021/acscentsci.3c01517