Pharmacophore

ISSN-2229-5402

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

PREDICTING BIOSIMILAR COMPARABILITY USING GLYCOSYLATION, CHARGE VARIANTS, POTENCY, AND STABILITY ATTRIBUTES

Oliver Schmidt[1]*, Lukas Weber[1], Jonas Richter[2]

1. ***Department of Computational Pharmacology and AI, Faculty of Medicine, University of Stuttgart, Stuttgart, Germany.***
2. ***Department of Intelligent Drug Systems, Faculty of Engineering, Technical University of Darmstadt, Darmstadt, Germany.***

ARTICLE INFO

Received:
19 January 2026
Received in revised form:
06 April 2026
Accepted:
07 April 2026
Available online:
28 April 2026

Keywords: Biosimilar comparability, Predictive modeling, Glycosylation, Charge variants, Potency, Stability

ABSTRACT

Biosimilar approval hinges on demonstrating analytical and functional similarity to the reference product across a panel of quality attributes. In current practice, comparability is often assessed by comparing individual attributes against predefined acceptance ranges. Univariate comparisons can overlook the correlation structure that links glycosylation, charge heterogeneity, potency, and stability. This creates the possibility that a batch may appear acceptable attribute by attribute while remaining atypical in the broader multivariate quality space. This manuscript proposes a predictive model for estimating the overall comparability of a biosimilar batch to its reference product. The model is designed to integrate glycosylation, charge variant, potency, and stability data while identifying the attributes most responsible for predicted dissimilarity. A gradient-boosted classification framework is conceptually trained on historical batch-level characterization data from reference and biosimilar development programs. Input features encode N-glycan profiles, charge variant distributions, relative potency, and forced-degradation stability behavior, with SHAP used to explain predictions.

Conceptually, the model would provide a single comparability score for each biosimilar batch. It would also generate an interpretable attribution profile showing which quality attributes contributed most strongly to any predicted deviation. Such a predictive tool could strengthen biosimilar development by providing a transparent, multivariate assessment of analytical similarity. It could help reduce the risk of failed comparability studies and support regulatory discussions with data-driven evidence.

To Cite This Article: Schmidt O, Weber L, Richter J. Predicting Biosimilar Comparability Using Glycosylation, Charge Variants, Potency, and Stability Attributes. Pharmacophore. 2026;17(2):73-82. https://doi.org/10.51847/00sL81Z4Dl

Introduction

The expanding use of biosimilars has increased the need for rigorous analytical and functional comparison against licensed reference biologics, particularly for monoclonal antibodies and other structurally complex proteins. Analytical similarity studies of trastuzumab, rituximab, pegfilgrastim, bevacizumab, and tocilizumab demonstrate that biosimilar development depends on coordinated evaluation of structural, physicochemical, and biological attributes [1, 2]. In such studies, N-glycosylation, charge heterogeneity, binding activity, cell-based potency, and degradation behavior are treated as central evidence streams for demonstrating that the proposed product is highly similar to its reference [3, 4]. A predictive model for biosimilar comparability must therefore be aligned with this head-to-head analytical paradigm rather than functioning as a generic product-quality classifier.

Current comparability assessments often rely on quality ranges, similarity intervals, or equivalence-style comparisons applied to individual critical quality attributes. This approach has practical value because it allows each attribute, such as acidic charge variants or high-mannose glycans, to be assessed against reference product variability [5, 6]. However, biologic product quality is not a collection of independent variables; glycosylation can influence effector function and clearance, while charge variants can reflect chemical modifications that also affect potency and stability [7, 8]. A univariate framework may therefore miss the multivariate quality pattern that defines whether a biosimilar batch is analytically close to the reference product as a whole.

Corresponding Author: Oliver Schmidt; Department of Computational Pharmacology and AI, Faculty of Medicine, University of Stuttgart, Stuttgart, Germany. E-mail: oliver.schmidt@gmail.com.

The increasing availability of batch characterization data creates an opportunity to treat comparability as a multivariate prediction problem. Machine learning has already been proposed for biopharmaceutical manufacturing, process control, and quality prediction, including applications in monoclonal antibody glycosylation, charge variant monitoring, and process analytical technology [9, 10]. Raman spectroscopy combined with machine learning has been explored for simultaneous prediction of charge variants [10], and broader reviews describe how artificial intelligence can support real-time quality prediction in bioprocessing [11, 12]. These developments suggest that interpretable models could complement established biosimilar similarity assessments by learning relationships among critical quality attributes rather than evaluating them in isolation.

This manuscript proposes a conceptual predictive model that integrates N-glycosylation profiles, charge variant distributions, potency assay outputs, and stability attributes to estimate the analytical comparability of a biosimilar batch. Gradient-boosted trees are well suited to this setting because tabular quality data commonly contain nonlinear relationships, feature interactions, and occasional missingness across development stages [13]. The model would be designed not only to produce a comparability probability but also to provide feature-level explanations, enabling scientists to distinguish whether a predicted deviation is driven by glycan composition, charge heterogeneity, potency behavior, or degradation susceptibility [14, 15]. In this way, predictive analytics could serve as an interpretable decision-support layer within the broader biosimilar totality-of-evidence framework.

Background

Biosimilar Comparability and Regulatory Expectations

Biosimilar comparability is grounded in the principle that analytical similarity provides the foundation for reducing residual uncertainty about clinical similarity. Published analytical assessments of proposed biosimilars typically combine physicochemical characterization, biological activity, purity, impurity, and stability measurements to support a totality-of-evidence argument [2, 16]. Studies comparing proposed biosimilars with reference trastuzumab, rituximab, pegfilgrastim, and bevacizumab illustrate how reference product variability is used as the benchmark for evaluating similarity [1, 3]. A predictive model should therefore be framed as a supplement to, rather than a replacement for, the established regulatory expectation that similarity be demonstrated across orthogonal analytical and functional methods.

Critical Quality Attributes of Biologics: Glycosylation, Charge Variants, Potency

N-glycosylation is a major critical quality attribute because glycan structures can influence Fc-mediated effector functions, pharmacokinetics, immunogenicity risk, and product consistency [7]. Glycoforms such as afucosylated, galactosylated, sialylated, and high-mannose species can be quantified through analytical workflows and transformed into model features that reflect the functional fingerprint of the molecule [15, 17]. Charge variant profiles, measured through approaches such as ion-exchange chromatography or capillary isoelectric focusing, capture acidic, main, and basic species that may arise from deamidation, oxidation, C-terminal lysine processing, or other modifications [5, 8]. Potency assays then provide an integrated functional readout, allowing the model to connect physicochemical variation with expected biological performance [18].

Stability as a Comparability Dimension

Stability is an essential comparability dimension because biosimilarity should extend beyond initial structural resemblance to include comparable behavior under storage, handling, and stress conditions. Forced-degradation studies can reveal oxidation, deamidation, aggregation, fragmentation, or charge-shift patterns that may not be obvious in release testing alone [19]. Comparative stability investigations of antibody formulations show how aggregate formation and degradation pathways can be examined with orthogonal analytical techniques to assess product robustness [20]. In a predictive comparability model, stability features would help distinguish batches that initially appear similar from those that would be expected to diverge under stress or during long-term storage.

Current Statistical Methods for Comparability Evaluation

Traditional comparability evaluation commonly uses attribute-specific acceptance ranges, tolerance intervals, tiered statistical testing, or expert-defined similarity criteria. These approaches are valuable because they provide transparent and auditable rules for determining whether individual quality attributes fall within the observed variability of reference product lots [16]. However, recent discussions on analytical similarity emphasize that individual bias and isolated attribute review can affect comparability conclusions, especially when many interrelated attributes are examined simultaneously [21]. Multivariate predictive modeling could address this limitation by learning a coordinated similarity space while preserving the interpretability required for scientific and regulatory review.

Machine Learning in Biopharmaceutical Quality and Process Control

Machine learning has become increasingly relevant to biopharmaceutical process development, manufacturing monitoring, and quality prediction. Reviews of machine learning in bioprocess optimization and control describe how data-driven models can support process understanding, real-time monitoring, and prediction of product quality from upstream and downstream variables [22]. In monoclonal antibody manufacturing, machine learning has been proposed for predicting glycosylation from media markers [15], optimizing culture conditions for glycan and charge variant control, and connecting process variables to

product quality within Quality-by-Design frameworks [9]. These applications provide a conceptual basis for extending machine learning from process control to biosimilar comparability prediction.

Model Development Overview

High-Level Prediction Pipeline

The proposed prediction pipeline begins with a biosimilar batch characterized across four coordinated evidence streams: N-glycosylation, charge variant distribution, potency, and stability. These inputs would be transformed into a standardized feature vector and passed to a trained classifier that estimates the probability that the batch belongs within the analytical similarity space of the reference product [14]. Because biosimilar comparability requires both a decision and an explanation, the model would pair the comparability score with SHAP-based attribution to identify which features increased or decreased confidence in similarity [9]. This structure mirrors published calls for interpretable quality prediction in biomanufacturing while adapting the objective to biosimilar analytical comparability [12].

Figure 1 presents the proposed interpretable multivariate architecture for transforming biosimilar batch-level glycosylation, charge variant, potency, and stability evidence into a transparent comparability probability and CQA-level attribution profile.

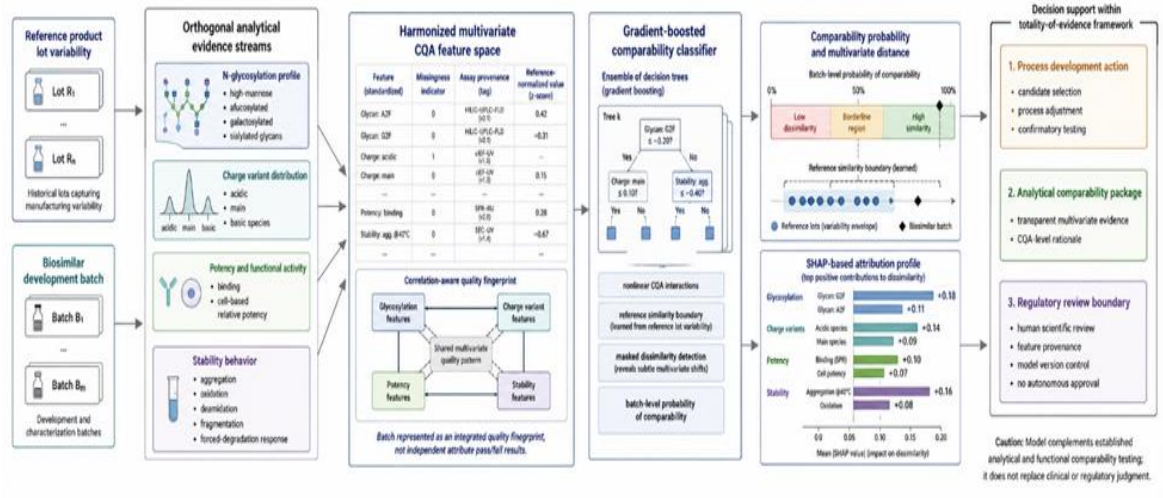

Figure 1. Interpretable Multivariate Biosimilar Comparability Architecture Integrating Glycosylation, Charge Variants, Potency, and Stability

Core Input Features

Core glycosylation features would include the relative abundance of key N-glycan species, such as agalactosylated, monogalactosylated, high-mannose, and sialylated glycans, because these structures can influence biological activity and product disposition [7]. Charge variant features would encode the continuous percentages of acidic, main, and basic peaks, reflecting the charge heterogeneity that has been studied in proposed trastuzumab and bevacizumab biosimilars [5, 6]. Potency features would summarize relative activity in binding or cell-based assays, while stability features would represent degradation tendencies such as aggregation, deamidation, and oxidation observed under defined stress conditions [18, 19]. By combining these variables, the model would represent the biosimilar batch as a multivariate quality profile rather than as a set of disconnected test outcomes.

Design Principles

The model should be multivariate, interpretable, and tolerant of occasional missing data because biosimilar development datasets often evolve as more complete characterization becomes available. It should be trained primarily to understand reference product variability and then evaluated against biosimilar development batches whose comparability status has been judged through expert analytical assessment [16, 21]. Quality-by-Design perspectives suggest that such a model should also support definition of a design space in which process adjustments are linked to expected product-quality behavior [17]. The final design principle is scientific transparency: model outputs must be explainable in terms of recognized critical quality attributes rather than treated as opaque algorithmic decisions [10, 12].

Data Sources and Feature Engineering

Compiling a Database of Biologic Batch Analyses

A comparability model would require a harmonized database of biologic batch analyses containing reference product lots, internal reference standard batches, and biosimilar development batches. Published analytical similarity studies show the types of batch-level characterization data that can inform such a database, including physicochemical profiles, functional assays, impurity assessments, and stability comparisons [1, 2]. Because analytical methods and reporting formats differ across programs, feature engineering would need to standardize units, normalize assay scales, and document method comparability

before model training [16]. The database should preserve product lineage and analytical context so that the model learns biosimilar-relevant variability rather than artifacts of laboratory method differences.

Feature Extraction from Glycosylation and Charge Variant Profiles

Complex glycosylation spectra should be reduced to biologically meaningful features that retain information about major and functionally relevant glycoforms. Reviews of monoclonal antibody N-glycosylation emphasize the importance of representing glycan categories associated with effector function, clearance, and product consistency, including galactosylated, afucosylated, high-mannose, and sialylated species [7]. Charge variant distributions can be encoded as percentages of acidic, main, and basic peaks, with additional derived features reflecting shifts in acidic or basic species relative to the reference product pattern [8]. This compact representation would allow the model to learn interactions between glycan composition and charge heterogeneity without requiring the full raw analytical signal for every batch.

Encoding Potency and Stability Endpoints

Potency should be encoded as a relative functional response against an appropriate reference standard, allowing the model to compare biological activity across batches while respecting assay context. Stability features would represent degradation behavior under defined stress or storage conditions, including the tendency toward aggregation, oxidation, deamidation, fragmentation, or charge redistribution [19, 20]. LC-MS-based assessment of charge variants and biological activity suggests that degradation-related species can have case-specific effects on stability and function, making these variables important for comparability prediction [18]. The model would therefore treat potency and stability endpoints as linked readouts that help translate physicochemical variation into expected functional and shelf-life relevance.

Table 1 defines how glycosylation, charge variant, potency, stability, reference-normalized context, and derived interaction features can be organized into a multivariate CQA representation for predictive biosimilar comparability assessment.

Table 1. Multivariate CQA Representation for Predictive Biosimilar Comparability Assessment

CQA domain	Representative features encoded in the model	Biological or analytical meaning	Why univariate review may be insufficient	Contribution to multivariate comparability prediction
N-glycosylation profile	Relative abundance of agalactosylated, monogalactosylated, galactosylated, afucosylated, sialylated, and high-mannose glycans	Captures Fc-related structural features linked to effector function, clearance behavior, and product consistency	A glycan species may fall within an individual range while its joint pattern with potency or charge behavior remains atypical	Defines the biosimilar batch's glycan fingerprint relative to reference product variability and helps identify functionally meaningful glycan-driven dissimilarity
Charge variant distribution	Percent acidic species, main peak, basic species, acidic shift from reference mean, basic shift from reference mean	Reflects chemical or enzymatic modifications such as deamidation, oxidation, C-terminal lysine processing, or other charge-altering changes	Separate acidic and basic limits may miss unusual charge redistribution patterns that become meaningful when paired with potency or stability shifts	Allows the model to detect whether charge heterogeneity is consistent with the learned reference quality space or signals degradation-related divergence
Potency and functional activity	Relative binding activity, cell-based potency, receptor or target engagement readouts, normalized activity against reference standard	Provides an integrated functional readout connecting structural and physicochemical variation to biological performance	Potency alone may appear acceptable despite structural shifts that increase risk under stress or indicate process drift	Links physicochemical attributes to functional comparability and helps distinguish analytically different but functionally tolerated variation from concerning deviation
Stability and forced-degradation behavior	Aggregation tendency, oxidation, deamidation, fragmentation, charge redistribution under stress, storage-related degradation slope	Indicates whether initial similarity is preserved under handling, storage, and stress conditions	Release-time similarity may conceal future divergence in degradation susceptibility or formulation robustness	Adds temporal and stress-response information so the model can identify batches likely to depart from the reference profile over time
Reference-normalized batch context	Reference lot mean, lot-to-lot variability, similarity interval position, batch lineage, assay method provenance	Anchors each biosimilar batch against the observed analytical variability of the reference product	Attribute values interpreted without lot context may overstate or understate meaningful differences	Prevents generic quality classification by ensuring the prediction is calibrated to product-specific reference variability
Missingness and assay provenance indicators	Missing stability flag, early-stage assay flag, analytical platform identifier, method harmonization status	Documents incomplete or heterogeneous evidence across development stages	Ignoring missingness may bias predictions, while excluding incomplete batches may reduce	Enables transparent handling of evolving datasets while preserving traceability for scientific and regulatory review

			development-stage usefulness	
Derived interaction features	Glycan–potency consistency, charge–stability shift, degradation–potency relationship, combined CQA distance from reference centroid	Represents coordinated quality behavior across interdependent attributes	Interactions among CQAs may be invisible when each attribute is judged independently	Supports detection of masked dissimilarity when several small deviations collectively indicate atypical multivariate quality

Deep Kernel Model Architecture

Model Choice – Gradient-Boosted Trees

Gradient-boosted tree models, such as XGBoost or LightGBM, are conceptually appropriate for biosimilar comparability prediction because they can handle tabular datasets with nonlinear relationships among quality attributes. Machine learning applications in biopharmaceutical process development show that tree-based and ensemble methods are frequently useful when product-quality outcomes depend on interacting process and analytical variables [22]. In this setting, a gradient-boosted classifier could learn how combinations of glycosylation, charge variant, potency, and stability features map onto expert-defined comparability decisions [14]. Native compatibility with SHAP-style explanations also makes this architecture attractive for a regulated environment where the rationale for each prediction must be reviewed scientifically [9].

Input Feature Vector and Pre-processing

The input feature vector would combine continuous glycan abundances, charge variant percentages, relative potency measures, and stability descriptors in a single structured representation. Pre-processing would address missing stability data from early-stage batches through transparent imputation or missingness indicators, while continuous variables would be scaled or transformed only when doing so improves model stability without obscuring scientific meaning [11]. Data splitting should avoid leakage between closely related batches, analytical repeats, or product families, because leakage could make the model appear more generalizable than it would be in a true biosimilar assessment setting. Feature provenance should be retained throughout the workflow so that each prediction can be traced back to the underlying analytical method and batch context [16].

Output: Comparability Probability and Multivariate Distance

The primary model output would be a comparability probability representing the expected likelihood that a biosimilar batch belongs within the reference product's analytical similarity space. A secondary output could express multivariate distance from the learned reference quality region, allowing batches to be ranked by how closely their combined CQA profile resembles the reference product pattern [21]. This distance would not replace established attribute-specific similarity ranges but would provide an additional view of whether the batch is typical across correlated attributes [3]. In development practice, the combined probability and distance outputs could guide batch selection, process adjustment, and prioritization of confirmatory analytical testing.

Handling Multivariate Cqa Correlations And Comparability Ranges

Learning Correlations among Quality Attributes

The model would learn correlations among quality attributes by examining how glycosylation patterns, charge variant profiles, potency responses, and stability behavior vary together across reference and biosimilar batches. Culture-condition studies show that media and process factors can influence both glycosylation and charge variant profiles, suggesting that these attributes should not be modeled as independent signals [23]. Analytical comparability studies of recombinant monoclonal antibodies also demonstrate that multiple physicochemical and biological attributes must be interpreted together to understand product similarity [24]. By learning these dependencies, the model could construct a multivariate fingerprint in which a change in one attribute is evaluated in the context of compensating or reinforcing changes in others.

Defining Multivariate Comparability Boundaries

Instead of defining similarity only through separate univariate ranges, the model would learn a decision boundary in the combined critical quality attribute space. This boundary would be informed by reference product variability, biosimilar development knowledge, and expert comparability assessments derived from orthogonal analytical characterization [16]. Multivariate quality evaluation is especially relevant because approved biosimilar case studies show that similarity is supported by coherent patterns across structural, purity, potency, and stability attributes rather than by a single measurement [2, 3]. A learned boundary could therefore represent the acceptable region of holistic analytical similarity while still allowing each contributing attribute to be inspected individually.

Detecting Masked Dissimilarity

A key function of the model would be to detect batches that pass individual attribute limits but remain atypical in the multivariate quality space. For example, a batch with charge variant percentages inside individual ranges might still show an unusual joint pattern when evaluated alongside glycosylation and potency, particularly if acidic or basic species reflect degradation-related modifications [8, 18]. Raman-based chemometric quality evaluation illustrates how multivariate spectral

patterns can support out-of-specification identification beyond isolated measurements [25]. In biosimilar development, this kind of masked dissimilarity detection could help prevent false confidence when multiple small deviations collectively suggest a meaningful departure from the reference product profile.

Table 2 shows how batches that meet individual quality attribute specifications may still exhibit atypical multivariate patterns when attributes such as charge variants, glycosylation, and potency are evaluated jointly, revealing potential hidden deviations not detectable through univariate limits alone.

Table 2. Multivariate Quality Space Analysis for Detecting Masked Batch Dissimilarity in Biosimilar Development

Quality Attribute	Within Individual Specification Limits	Multivariate Pattern Observation	Potential Interpretation	Supporting Analytical Approach
Charge variants (acidic/basic species)	Yes	Correlated shift with glycosylation profile	Possible degradation-related modification or process drift	Ion-exchange chromatography, capillary electrophoresis
Glycosylation profile	Yes	Unusual co-variation with charge variants	Structural heterogeneity affecting higher-order quality space	LC–MS glycan mapping
Potency	Yes	Discrepancy when aligned with structural attributes	Functional impact masked by acceptable individual assay values	Cell-based bioassays
Acidic/basic species distribution	Yes	Joint deviation pattern across multiple attributes	Subtle degradation signature not reflected in single metrics	Charge variant profiling, peptide mapping
Raman spectral signature (chemometric profile)	Within tolerance band	Out-of-class multivariate clustering despite in-spec metrics	Global structural divergence from reference product	Raman spectroscopy + PCA/PLS modeling

Model Interpretability and Comparability Decisions

SHAP-Driven Attribute Attribution

For a borderline biosimilar batch, SHAP-driven attribution would identify which attributes most strongly influenced the predicted comparability score. A model might indicate, conceptually, that increased high-mannose glycan abundance reduced confidence in similarity because high-mannose structures are associated with altered clearance-related behavior and are recognized as important glycosylation features [7]. It might also show that a shift in acidic or basic charge variants contributed to predicted dissimilarity because charge heterogeneity can reflect chemical or enzymatic modifications relevant to stability and potency [5, 8]. Such explanations would make the prediction actionable by pointing process scientists toward upstream culture conditions, purification steps, or formulation variables that should be reviewed.

Supporting the Totality-of-Evidence Package

The model's explanations could support a totality-of-evidence package by showing how multivariate analytical similarity was assessed and why a given batch was considered comparable or at risk. Analytical similarity reviews emphasize that biosimilar evaluation requires integration of orthogonal physicochemical, structural, functional, and stability evidence rather than reliance on any single assay [16]. A transparent predictive model could supplement this evidence by mapping each batch into a learned comparability space and documenting the contribution of glycosylation, charge variants, potency, and degradation behavior to the predicted decision [14]. This use would be most appropriate as a scientific decision-support tool that strengthens interpretation of established comparability data rather than as an independent regulatory endpoint.

Integration Into Biosimilar Development And Regulatory Submissions

Early-Stage Risk Assessment of Biosimilar Candidates

During clone selection and process development, the model could be used to identify biosimilar candidates whose emerging quality profiles are expected to align closely with the reference product. Machine learning approaches for predicting antibody glycan quality from culture media markers show how early process information can be connected to downstream glycosylation attributes [15]. Optimization studies linking culture conditions and media components to glycosylation and charge variants further suggest that model-guided development could help steer candidate selection toward desired quality profiles. In this role, the comparability model would function as an early risk-assessment tool that helps prioritize candidates for deeper analytical and functional characterization.

Aiding in Lifecycle Management and Post-Approval Changes

The same modeling framework could also aid lifecycle management by evaluating whether post-change batches remain within the learned analytical similarity space. Comparability principles for recombinant monoclonal antibody therapeutics show that manufacturing changes require structured assessment of whether product quality remains consistent with prior material [24]. Machine learning in biopharmaceutical manufacturing has been proposed as an enabler of continuous monitoring and real-

time quality prediction, which is conceptually aligned with post-approval change management [11, 12]. A biosimilar comparability model could therefore provide continuity assurance by comparing new batches against the established multivariate quality profile while preserving conventional analytical testing as the primary evidence base.

Evaluation Strategy

Predictive Performance for Comparability Classification

The evaluation strategy would assess whether the model can distinguish batches judged comparable from those judged non-comparable by expert analytical review. Conceptual performance assessment could use classification metrics such as receiver-operating and precision-recall behavior, but any such metrics should be interpreted alongside scientific review of the underlying quality attributes rather than treated as standalone proof of biosimilarity [21]. Machine learning surveys in bioprocess development emphasize that model validation must address generalizability, data leakage, and the practical decision context in which predictions will be used. For biosimilar comparability, the most important evaluation question is whether the model's decisions are consistent with expert assessment across glycosylation, charge variant, potency, and stability evidence. **Table 3** shows a structured evaluation framework for biosimilar comparability models, integrating statistical performance metrics with expert-aligned assessment across critical quality attributes, generalizability testing, and decision-context validation.

Table 3. Evaluation framework for model-based biosimilar comparability assessment

Evaluation component	Description	Methods / metrics	Interpretation in biosimilarity context
Batch comparability classification	Distinguishing expert-judged comparable vs non-comparable batches	Receiver operating characteristic (ROC-AUC), precision–recall (PR-AUC), accuracy, F1-score	Quantifies discriminative ability, but not sufficient alone for regulatory or scientific equivalence
Alignment with expert analytical review	Agreement between model outputs and expert decisions	Cohen's kappa, concordance rate, confusion matrix stratified by attribute	Measures consistency with analytical scientists' judgment across multiple evidence layers
Critical quality attributes (CQAs) consistency	Evaluation across glycosylation, charge variants, potency, stability	Attribute-level error rates, multi-task performance metrics	Ensures model respects biologically meaningful quality signals rather than only global classification
Generalizability	Performance on unseen processes, batches, or manufacturing sites	External validation sets, cross-site validation, temporal split testing	Assesses robustness across manufacturing variability and prevents overfitting to specific datasets
Data leakage control	Ensuring no information overlap between training and evaluation sets	Leakage audits, group-wise splitting (by batch/process), feature provenance checks	Prevents inflated performance estimates and ensures realistic predictive capability
Decision-context validation	Suitability of predictions for real comparability decisions	Decision curve analysis, threshold sensitivity analysis, expert-in-the-loop review	Evaluates whether model outputs support actionable and scientifically valid decisions
Uncertainty characterization	Quantification of confidence in predictions	Calibration curves, Brier score, predictive entropy	Identifies cases requiring additional experimental or analytical confirmation

Assessment of Multivariate Sensitivity

The model should be evaluated for its ability to detect intentionally introduced differences in quality profiles, such as altered glycosylation, shifted charge variants, or increased degradation susceptibility. Machine learning-enhanced analytical methods have shown how subtle product-quality differences can be captured from complex measurement data, including size-based heterogeneity and biophysical properties [13, 26]. Sensitivity testing would therefore examine whether the model flags coordinated deviations that may be missed by separate univariate comparisons. Such evaluation would help determine whether the model contributes meaningful multivariate insight rather than merely reproducing existing attribute-by-attribute rules.

Case Studies Using Known Biosimilars

Retrospective case studies using known biosimilars would provide a practical way to evaluate whether the model behaves consistently with established comparability conclusions. Published similarity assessments of pegfilgrastim, trastuzumab, rituximab, tocilizumab, and bevacizumab provide examples of the analytical evidence patterns that a model should recognize as supportive of biosimilarity [1, 2]. These cases could be used conceptually to verify that the model classifies well-characterized biosimilar profiles as comparable while reserving lower confidence for profiles with meaningful glycosylation, charge, potency, or stability deviations [3, 4, 6]. The goal would not be to replace the conclusions of these studies, but to test whether the predictive framework reproduces their integrated analytical logic.

Table 4 outlines the interpretability, validation, and governance requirements needed to position predictive biosimilar comparability modeling as a transparent decision-support layer within the totality-of-evidence framework.

Table 4. Interpretability, Validation, and Governance Framework for Biosimilar Comparability Prediction

Framework component	Core analytical question	Recommended implementation	Evidence or output generated	Decision-use value	Governance safeguard
Comparability probability	Does the biosimilar batch fall within the learned analytical similarity space of the reference product?	Train a gradient-boosted classifier using reference product lots, internal standards, and expert-reviewed biosimilar development batches	Batch-level probability of comparability with confidence-oriented interpretation	Supports early candidate triage, process adjustment, and prioritization of confirmatory testing	Treat the probability as decision support, not as an autonomous biosimilarity determination
Multivariate distance	How far is the batch from the reference product's coordinated CQA profile?	Estimate distance from the learned reference quality region using combined glycan, charge, potency, and stability features	Ranked batch proximity to reference variability and identification of borderline profiles	Helps distinguish ordinary lot variability from atypical multivariate deviation	Maintain attribute-level review so multivariate distance does not obscure individual CQA failures
SHAP attribution	Which CQA domains drive predicted similarity or dissimilarity?	Generate feature- and domain-level SHAP explanations for each batch prediction	Ranked contribution profile for glycosylation, charge variants, potency, and stability	Makes model outputs scientifically actionable for analytical and process teams	Require expert review of explanations before using them in development or regulatory communication
Masked dissimilarity testing	Can the model detect batches that pass individual limits but remain atypical jointly?	Simulate or retrospectively identify coordinated small deviations across glycan, charge, potency, and stability features	Sensitivity profile showing whether joint deviations are flagged despite attribute-level acceptability	Demonstrates added value beyond univariate similarity ranges	Document test scenarios and avoid overclaiming clinical relevance from analytical deviations alone
Leakage and generalizability control	Is the model learning true comparability patterns rather than batch lineage, replicate structure, or product-family artifacts?	Split data by batch lineage, product family, development campaign, or analytical run where appropriate	Validation results under leakage-resistant data partitions	Improves credibility of predictive performance estimates	Preserve traceable data partitions and prohibit random splits that mix closely related batches across train and test sets
Reference lot representativeness	Is the learned similarity boundary based on adequate reference product variability?	Curate reference lots across manufacturing periods, expiry windows, assay runs, and relevant product presentations	Documented reference product variability map	Reduces risk of overly narrow or overly broad comparability boundaries	Require periodic review as additional reference lot data become available
Analytical method harmonization	Are feature values comparable across laboratories, platforms, and time?	Normalize units, align assay scales, document method changes, and include provenance indicators	Harmonized feature dictionary and method-comparability record	Prevents method artifacts from being misread as product-quality differences	Version-control feature definitions and retain raw-to-modeled data traceability
Lifecycle monitoring	Can the model support post-change or post-approval comparability review?	Apply the same model framework to new batches after manufacturing changes, formulation updates, or process refinements	Ongoing comparability trend reports against the learned multivariate quality space	Supports continuity assurance while conventional testing remains primary	Revalidate or recalibrate the model after major process, assay, or product-context changes
Regulatory communication boundary	How should the model be positioned in a biosimilar evidence package?	Present the model as an interpretable analytical decision-support layer within the totality-of-evidence framework	Transparent model rationale, CQA attribution, validation summary, and limitations	Strengthens explanation of multivariate analytical similarity without replacing established requirements	Explicitly state that the model does not substitute for orthogonal analytical testing, functional assays, clinical assessment, or regulatory judgment

Limitations

Dependence on Reference Product Lot Data Availability

The model's definition of comparability would depend strongly on the breadth and representativeness of reference product lot data. If the reference dataset is narrow, the learned similarity space may underrepresent normal manufacturing variability and could lead to overly conservative predictions [16]. Conversely, poorly harmonized data from different laboratories or analytical platforms could expand the learned space in ways that obscure meaningful product differences [21]. This limitation means that dataset curation, method standardization, and expert review remain essential prerequisites for responsible model use.

Not a Substitute for Clinical Assessment

The proposed model addresses analytical comparability and should not be interpreted as a direct predictor of clinical similarity, safety, or immunogenicity. Although glycosylation, charge variants, potency, and stability can influence biological behavior, clinical outcomes depend on additional factors that cannot be inferred from quality attributes alone [7, 8]. Totality-of-evidence reasoning requires that analytical similarity be integrated with pharmacological, clinical, and immunogenicity considerations as appropriate for the product and regulatory pathway [16]. The model should therefore be positioned as an analytical decision-support tool rather than as a substitute for the broader biosimilar evaluation framework.

Conclusion

A machine learning model for predicting biosimilar comparability could provide a structured way to integrate glycosylation, charge variant, potency, and stability attributes into one interpretable assessment. By treating the biosimilar batch as a multivariate quality profile, the model would help shift comparability thinking from isolated attribute checks toward holistic similarity assessment. The proposed framework is conceptual and should be evaluated as a scientific decision-support system rather than as an autonomous approval tool. Its central value would lie in improving how complex analytical evidence is organized, interpreted, and communicated.

The strongest feature of this approach is its alignment with the analytical similarity pillar of biosimilar development. It uses attributes already central to biosimilar comparability studies, but evaluates them in a way that reflects their interdependence. Interpretability is also essential because model explanations can show whether predicted dissimilarity is driven by glycosylation, charge heterogeneity, potency, stability, or a combination of these factors. This makes the framework more useful for process scientists, analytical teams, and regulatory reviewers.

Several challenges remain before such a model could be adopted in routine development or regulatory communication. Reference product lot data may be limited, analytical platforms may differ across organizations, and regulatory acceptance of machine learning-based comparability arguments will require transparency and careful validation. The model would also need governance procedures for data curation, version control, feature documentation, and expert review. These challenges are manageable, but they require deliberate collaboration between analytical scientists, data scientists, quality teams, and regulators.

Pre-competitive data sharing could accelerate development of robust biosimilar comparability models. Shared datasets containing harmonized glycosylation, charge variant, potency, and stability attributes would help define broader and more reliable similarity spaces. Collaboration among biosimilar developers, reference product manufacturers, academic groups, and regulators could make predictive analytics more trustworthy for biologic quality evaluation. With appropriate validation and transparent interpretation, multivariate predictive modeling could become a valuable complement to established biosimilar comparability practice.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Brokx S, Scrocchi L, Shah N, Dowd J. A demonstration of analytical similarity comparing a proposed biosimilar pegfilgrastim and reference pegfilgrastim. Biologicals. 2017;48:28-38.
2. Lee J, Kang HA, Bae JS, Kim KD, Lee KH, Lim KJ, et al. Evaluation of analytical similarity between trastuzumab biosimilar CT-P6 and reference product using statistical analyses. MAbs. 2018;10(4):547-71.
3. Xu Y, Xie L, Zhang E, Gao W, Wang L, Cao Y, et al. Physicochemical and functional assessments demonstrating analytical similarity between rituximab biosimilar HLX01 and the MabThera®. MAbs. 2019;11(3):606-20.
4. Miao S, Fan L, Zhao L, Ding D, Liu X, Wang H, et al. Physicochemical and biological characterization of the proposed biosimilar tocilizumab. Biomed Res Int. 2017;2017:4926168.
5. Dakshinamurthy P, Mukunda P, Kodaganti BP, Shenoy BR, Natarajan B, Maliwalave A, et al. Charge variant analysis of proposed biosimilar to Trastuzumab. Biologicals. 2017;46:46-56.
6. Ruppen I, Beydon ME, Solís C, Sacristán D, Vandenheede I, Ortiz A, et al. Similarity demonstrated between isolated charge variants of MB02, a biosimilar of bevacizumab, and Avastin® following extended physicochemical and functional characterization. Biologicals. 2022;77:1-5.
7. Shrivastava A, Joshi S, Guttman A, Rathore AS. N-Glycosylation of monoclonal antibody therapeutics: A comprehensive review on significance and characterization. Anal Chim Acta. 2022;1209:339828.

8. Beck A, Nowak C, Meshulam D, Reynolds K, Chen D, Pacardo DB, et al. Risk-based control strategies of recombinant monoclonal antibody charge variants. Antibodies. 2022;11(4):73.
9. Walsh I, Myint M, Nguyen-Khuong T, Ho YS, Ng SK, Lakshmanan M. Harnessing the potential of machine learning for advancing "quality by design" in biomanufacturing. MAbs. 2022;14(1):2013593.
10. Nitika N, Keerthiveena B, Thakur G, Rathore AS. Convolutional neural networks guided Raman spectroscopy as a process analytical technology (PAT) tool for monitoring and simultaneous prediction of monoclonal antibody charge variants. Pharm Res. 2024;41(3):463-79.
11. Nikita S, Thakur G, Jesubalan NG, Kulkarni A, Yezhuvath VB, Rathore AS. AI-ML applications in bioprocessing: ML as an enabler of real time quality prediction in continuous manufacturing of mAbs. Comput Chem Eng. 2022;164:107896.
12. Rathore AS, Nikita S, Thakur G, Mishra S. Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. Trends Biotechnol. 2023;41(4):497-510.
13. Shrivastava A, Mandal S, Pattanayek SK, Rathore AS. Rapid estimation of size-based heterogeneity in monoclonal antibodies by machine learning-enhanced dynamic light scattering. Anal Chem. 2023;95(21):8299-309.
14. Shatat SM, Al-Ghobashy MA, Fathalla FA, Abbas SS, Eltanany BM. Coupling of Trastuzumab chromatographic profiling with machine learning tools: A complementary approach for biosimilarity and stability assessment. J Chromatogr B. 2021;1184:122976.
15. Lakshmanan M, Chia S, Pang KT, Sim LC, Teo G, Mak SY, et al. Antibody glycan quality predicted from CHO cell culture media markers and machine learning. Comput Struct Biotechnol J. 2024;23:2497-506.
16. Nupur N, Joshi S, Gulliarme D, Rathore AS. Analytical similarity assessment of biosimilars: global regulatory landscape, recent studies and major advancements in orthogonal platforms. Front Bioeng Biotechnol. 2022;10:832059.
17. Puranik A, Saldanha M, Chirmule N, Dandekar P, Jain R. Advanced strategies in glycosylation prediction and control during biopharmaceutical development: Avenues toward industry 4.0. Biotechnol Prog. 2022;38(5):e3283.
18. Singh SK, Kumar D, Malani H, Rathore AS. LC–MS based case-by-case analysis of the impact of acidic and basic charge variants of bevacizumab on stability and biological activity. Sci Rep. 2021;11(1):2487.
19. Beaumont VA, Liu L, Shi H, Rouse JC, Kim HY. Application of NMR and chemometric analyses to better understand the quality attributes in pH and thermally degraded monoclonal antibodies. Pharm Res. 2023;40(10):2457-67.
20. Bana AA, Sajeev N, Halder S, Masi HA, Patel S, Mehta P. Comparative stability study and aggregate analysis of Bevacizumab marketed formulations using advanced analytical techniques. Heliyon. 2023;9(9):9.
21. Rathore AS, Joshi S, Nupur N, Saxena N, Bhattacharya S, Roy S. Taking the individual bias out of examining comparability of biosimilars: A case study on monoclonal antibody therapeutics. Int J Biol Macromol. 2023;227:124-33.
22. Mondal PP, Galodha A, Verma VK, Singh V, Show PL, Awasthi MK, et al. Review on machine learning-based bioprocess optimization, monitoring, and control systems. Bioresour Technol. 2023;370:128523.
23. Gangwar N, Priyanka P, Rathore AS. Achieving charge variant profile of innovator molecule during development of monoclonal antibody based biosimilars–use of media components. Biochem Eng J. 2022;182:108438.
24. Ambrogelly A, Gozo S, Katiyar A, Dellatore S, Kune Y, Bhat R, et al. Analytical comparability study of recombinant monoclonal antibody therapeutics. MAbs. 2018;10(4):513-38.
25. Massei A, Falco N, Fissore D. Use of Raman spectroscopy and PCA for quality evaluation and out-of-specification identification in biopharmaceutical products. Eur J Pharm Biopharm. 2024;200:114342.
26. Wu IE, Kalejaye L, Lai PK. Machine learning models for predicting monoclonal antibody biophysical properties from molecular dynamics simulations and deep learning-based surface descriptors. Mol Pharm. 2024;22(1):142-53.