Pharmacophore

ISSN-2229-5402

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

COUNTERFACTUAL AI FOR ANTIMICROBIAL DOSE OPTIMIZATION USING RENAL FUNCTION, MIC VALUES, AND PK/PD TARGETS

Maria Gonzalez[1]*, Javier Ruiz[1], Lucia Torres[2], Elena Ruiz[1]

1. ***Department of Drug Informatics and AI Systems, Faculty of Pharmacy, University of Granada, Granada, Spain.***
2. ***Department of Computational Pharmaceutical Analytics, Faculty of Medicine, University of Seville, Seville, Spain.***

ARTICLE INFO

Received:
21 November 2024
Received in revised form:
28 February 2025
Accepted:
03 March 2025
Available online:
28 April 2025

Keywords: Explainable AI, Counterfactual explanations, Antimicrobial dosing, Pharmacokinetics, Pharmacodynamics, Renal function

ABSTRACT

Antimicrobial dosing is a high-stakes clinical task where insufficient exposure can lead to therapeutic failure, while excessive exposure increases the risk of toxicity. Conventional dosing tools often reduce this complexity to broad rules that fail to capture patient-specific pharmacokinetics, pathogen susceptibility, or dynamic changes in renal function. Many clinical decision support systems provide recommended doses without explaining how these recommendations might vary under plausible alternative patient states, limiting clinicians' ability to assess robustness against uncertainties in renal function, MIC interpretation, or PK/PD target selection. This article proposes a counterfactual explainable AI (XAI) framework for antimicrobial dose optimization, designed not only to recommend a dose but also to generate interpretable "what-if" scenarios illustrating how dosing would change if renal function, MIC, or PK/PD goals differed. The framework integrates a predictive dosing model with a counterfactual generation engine guided by pharmacometric reasoning: the predictive component estimates antimicrobial doses based on renal function, MIC, patient covariates, and target-attainment objectives, while the counterfactual component perturbs clinically relevant inputs within plausible ranges. Conceptually, the system would return a recommended dose alongside alternative dosing scenarios linked to changes in renal clearance, pathogen susceptibility, or exposure targets, each accompanied by a rationale connecting the output to pharmacokinetic and pharmacodynamic principles rather than presenting the recommendation as an unexplained prediction. By providing transparent, clinically interpretable alternatives, this counterfactual XAI approach could support individualized pharmacotherapy, promote antimicrobial stewardship, and enhance the safe implementation of AI-assisted prescribing.

To Cite This Article: Gonzalez M, Ruiz J, Torres L, Ruiz E. Counterfactual AI for Antimicrobial Dose Optimization Using Renal Function, MIC Values, and PK/PD Targets. Pharmacophore. 2025;16(2):1-11. https://doi.org/10.51847/1e7suncaUs

Introduction

Precise antimicrobial dosing is clinically important because efficacy depends on achieving sufficient exposure against the pathogen while avoiding unnecessary toxicity. Vancomycin monitoring guidelines emphasize that exposure-guided dosing is needed for serious methicillin-resistant *Staphylococcus aureus* infections, particularly where nephrotoxicity and underexposure are competing concerns [1]. Beta-lactam optimization guidance similarly frames dosing as a PK/PD problem in which patient physiology, infection severity, and antimicrobial properties must be considered together [2]. Standard nomograms can provide useful starting points, but they are poorly suited to patients with dynamic renal function, altered volume of distribution, or organism susceptibility patterns that differ from assumptions embedded in static rules [3].

Electronic health records, therapeutic drug monitoring databases, and microbiology reports now provide information that could support more individualized antimicrobial dosing. Deep learning models using EHR-derived therapeutic drug monitoring data have been proposed for vancomycin dosing and monitoring, illustrating how longitudinal clinical information could inform precision pharmacotherapy [4]. Machine-learning approaches for initial and maintenance vancomycin dosing have also been explored across institutional settings, suggesting that data-driven dose estimation can complement population pharmacokinetic

Corresponding Author: Maria Gonzalez; Department of Drug Informatics and AI Systems, Faculty of Pharmacy, University of Granada, Granada, Spain. E-mail: maria.gonzalez@gmail.com

reasoning [5, 6]. However, clinicians may be reluctant to follow a model that recommends a dose without explaining how renal function, MIC, or PK/PD target assumptions influenced that recommendation [7, 8].

Counterfactual explainability offers a clinically intuitive way to address this limitation by asking how a model recommendation would change under plausible alternative patient states. Diverse counterfactual explanation methods have been developed to identify minimal changes in input features that alter a model output, which is directly relevant to dosing questions such as how a recommendation would change if renal clearance worsened or the MIC were higher [9]. In healthcare, explainability must also support accountability, clinical reasoning, and multidisciplinary communication rather than merely producing visually attractive explanations [7]. For antimicrobial dosing, counterfactuals are especially useful because they can translate model sensitivity into prescribing language that pharmacists and clinicians already use during dose adjustment.

The thesis of this article is that antimicrobial dose optimization should be modeled as both a prediction problem and an explanation problem. A counterfactual AI model could integrate renal function, MIC information, and PK/PD targets to recommend an initial or revised antimicrobial dose while simultaneously generating alternative dosing scenarios grounded in pharmacometric assumptions [10, 11]. Such a framework would align with the broader movement toward model-informed precision dosing while extending it with explainability that is tailored to clinical decision support [12, 13]. Rather than asking clinicians to trust a black box, the system would show how the recommendation is expected to change when clinically meaningful assumptions change.

Background

PK/PD Principles and Antimicrobial Dose Optimization

Antimicrobial PK/PD links dose to exposure and exposure to microbiological effect, commonly using indices such as AUC/MIC, peak concentration to MIC, or time above MIC. Vancomycin monitoring recommendations highlight AUC-guided dosing as a preferred conceptual basis for balancing efficacy and toxicity, especially when trough-only monitoring may be insufficiently informative [1]. Beta-lactam guidance emphasizes target attainment over the dosing interval, which makes the relationship between renal clearance, dosing frequency, and pathogen MIC central to dose design [2]. These principles imply that an explainable AI dosing system should not merely predict a dose, but should make explicit how exposure, MIC, and safety constraints shape the recommendation.

Renal Function and Its Impact on Antimicrobial Pharmacokinetics

Renal function is a major determinant of exposure for many antimicrobials, and changes in creatinine clearance or estimated glomerular filtration rate can alter the dose required to remain within a therapeutic window. Population pharmacokinetic analyses of vancomycin repeatedly identify renal function as a key covariate, supporting its role as a core feature in precision dosing models [14]. In critically ill patients, antimicrobial therapeutic drug monitoring is recommended partly because renal function, fluid balance, and organ support can change rapidly and unpredictably [3]. A counterfactual model should therefore treat renal function not as a fixed descriptor, but as a clinically dynamic variable whose plausible changes can be explored through what-if dosing scenarios.

MIC Values and Local Antibiograms

MIC values provide the susceptibility denominator for several PK/PD indices, but they are not perfectly stable measurements and may vary by method, organism, laboratory practice, and breakpoint interpretation. MIC-based dose adjustment has been described as clinically attractive but also vulnerable to misconceptions when MIC precision is overinterpreted or separated from uncertainty in PK/PD target attainment [15]. For this reason, a model may encode a measured MIC when available while also considering local MIC distributions or breakpoint categories when exact values are uncertain. Counterfactual explanations can make this uncertainty visible by showing how a dose recommendation would be expected to change if the organism's MIC were interpreted differently within clinically plausible bounds.

Machine Learning and Pharmacometrics

Machine learning has increasingly been applied to drug dosing, including models for vancomycin initial dose selection, trough prediction, AUC targeting, and therapeutic monitoring [5, 6, 16]. Hybrid pharmacokinetic and machine-learning approaches have also been proposed to improve individualization by combining mechanistic priors with data-driven updating [10]. More broadly, machine learning for dose individualization is viewed as a complement to pharmacometrics when models can incorporate nonlinear covariate relationships and complex clinical patterns [11]. The remaining gap is not simply predictive performance, but the ability to generate explanations that connect model recommendations to pharmacological reasoning clinicians can evaluate.

Counterfactual Explanations in Healthcare

Counterfactual explanations describe how a model output would change if selected input features were different, ideally using changes that are plausible, sparse, and meaningful to the decision maker. Explainable AI taxonomies distinguish counterfactuals from feature-attribution methods because counterfactuals are contrastive and action-oriented, making them well suited to clinical "what-if" reasoning [17]. In healthcare, counterfactuals must be evaluated for clinical validity, safety, and trustworthiness, because a mathematically valid feature perturbation may still be physiologically impossible or

therapeutically unsafe [7]. For antimicrobial dosing, this means counterfactual generation should be constrained by renal physiology, MIC uncertainty, dosing feasibility, and PK/PD target logic.

Model Development Overview

High-Level Prediction and Explanation Pipeline

For a given antimicrobial and patient, the proposed system would use renal function, MIC, antimicrobial identity, pathogen identity, patient covariates, and PK/PD target selection as inputs to estimate an individualized dose. The predictive component could resemble existing machine-learning vancomycin dosing models that use clinical covariates to emulate expert dosing or estimate exposure-guided treatment decisions [18, 19]. In parallel, the counterfactual engine would perturb one or more inputs within clinically plausible ranges and re-estimate the dose, thereby producing alternative recommendations linked to specific feature changes. The output would be a recommendation accompanied by explanations such as whether the dose is primarily sensitive to renal clearance, MIC assumptions, or the selected exposure target.

Core Input Features

Core input features should include renal function measures such as creatinine clearance or estimated glomerular filtration rate, MIC or MIC category, organism, antimicrobial class, patient weight, relevant illness features, dosing history, and target-attainment objectives. Vancomycin machine-learning studies have commonly used clinical and therapeutic drug monitoring variables to estimate concentrations or dosing needs, supporting the feasibility of representing dosing as a structured prediction task [20-22]. MIC should be represented in a way that preserves its role in PK/PD reasoning, either as a measured value, a breakpoint-derived category, or a distributional feature reflecting local susceptibility patterns [15]. The target can be encoded as a desired exposure or safety condition, allowing the same architecture to support efficacy-oriented and toxicity-aware counterfactual reasoning.

Design Principles

The proposed counterfactuals must be clinically plausible, pharmacometrically consistent, and interpretable to the prescribing clinician. Plausibility requires that feature perturbations respect biological and clinical constraints, such as avoiding impossible renal-function changes or unsupported MIC substitutions. Pharmacometric consistency requires the counterfactual dose to remain connected to exposure-response reasoning rather than being generated solely by an unconstrained machine-learning optimizer [10, 12]. Interpretability requires that the output use clinical language, explain the direction of dose change, and show which assumption drives the recommendation, reflecting concerns that explainable AI in health care must support real decision making rather than superficial transparency [8].

Data Sources and Feature Engineering

Construction of a Training Set from Therapeutic Drug Monitoring and EHR Data

A training set for this conceptual model would be constructed from linked dosing events, measured antimicrobial concentrations, renal function values, microbiology results, and clinical context captured in EHR and therapeutic drug monitoring systems. Prior vancomycin studies using EHR-derived data show how measured concentrations, dosing histories, and patient covariates can be transformed into inputs for prediction of exposure or dose requirements [4, 20]. Bayesian and hybrid pharmacokinetic approaches further support the use of observed concentrations to update individualized exposure estimates from prior assumptions [10]. The engineered outcome would not need to be framed as a single historical "correct dose," but could instead represent whether a candidate dose would be expected to align with a prespecified PK/PD and safety rationale.

Encoding Renal Function and MIC

Renal function should be encoded as both a continuous predictor and a clinically interpretable state, because clinicians reason about creatinine clearance trends as well as absolute renal impairment categories. Recent vancomycin prediction models have treated renal function and related laboratory features as central covariates for concentration or dose estimation, reinforcing their importance in data-driven antimicrobial dosing [23-26]. MIC can be encoded as a continuous value when reported, but when only susceptibility categories are available, the model could map the category to organism- and site-specific MIC distributions. This encoding strategy reflects the caution that MIC-based dosing should account for measurement and interpretive uncertainty rather than treating a single MIC value as exact [15].

PK/PD Target Definitions

PK/PD targets should be represented explicitly as model inputs and optimization constraints, because the "optimal" antimicrobial dose depends on the therapeutic objective being pursued. For vancomycin, exposure-guided monitoring recommendations emphasize AUC-based reasoning for serious infections, while toxicity avoidance remains a parallel constraint [1]. For beta-lactams and other antimicrobials, target definitions may involve time-dependent exposure goals that are influenced by renal clearance and dosing interval [2]. Encoding these targets directly enables counterfactual questions such as whether the recommended dose changes because the patient's renal function changed, because the MIC assumption changed, or because the clinician selected a more conservative safety target.

Counterfactual AI Architecture for Dose Optimization

Predictive Model for Initial Dose Recommendation

The predictive layer could use a gradient-boosted tree, Bayesian regression model, Bayesian network, or hybrid pharmacokinetic-machine-learning model to estimate the dose expected to satisfy a selected PK/PD target under current patient conditions. Tree-based explanation methods have been used to move from local model behavior toward broader understanding of feature influence, making them suitable when the dosing model needs both predictive flexibility and interpretable structure [27]. Vancomycin initial-dose and AUC-targeting models show that machine learning can be framed around clinically relevant dosing objectives rather than generic concentration prediction alone [5, 6, 24]. The model output should be a dose recommendation expressed in a clinically usable dosing unit, but its interface should emphasize the assumptions and constraints behind the recommendation rather than presenting the value as an autonomous instruction.

Counterfactual Dose Generation

Starting from the factual input vector (x), the counterfactual generator would produce an alternative vector (x') by changing one or more clinically meaningful features while holding irrelevant or immutable features constant. For a predominantly renally cleared antimicrobial, a pharmacometric reference constraint could be expressed conceptually as (Dose_{new}=Dose_{baseline}\times\frac{CrCl_{new}}{CrCl_{baseline}}), with the model then explaining whether its recommendation follows or deviates from that renal-clearance expectation. Counterfactual generation methods such as diverse counterfactual explanations provide a basis for producing multiple plausible alternatives rather than a single nearest contrastive case [9]. In antimicrobial dosing, however, the search must be further constrained by safety, feasible dose formulations, MIC uncertainty, and PK/PD logic, aligning counterfactual generation with clinical pharmacology rather than unrestricted feature optimization.

Multi-Objective Counterfactual Generation

A multi-objective counterfactual engine should balance efficacy, toxicity avoidance, plausibility, actionability, and interpretability. Reinforcement-learning concepts for precision dosing suggest that dosing decisions can be viewed as sequential choices under therapeutic constraints, while model-informed AI emphasizes that pharmacological structure should guide how such choices are represented [12]. Counterfactual explanations in clinical decision support have also been linked to the need for plausibility and clinician trust, especially when the system is used to support high-risk treatment decisions. Therefore, each counterfactual dose scenario should indicate the changed assumption, the expected direction of exposure change, and the clinical trade-off, without implying that the model has proven superiority in the absence of formal prospective evaluation.

Figure 1 presents the proposed counterfactual AI architecture for antimicrobial dose optimization, showing how renal function, MIC interpretation, and PK/PD target assumptions are converted into a factual dose recommendation and clinically plausible what-if dosing alternatives.

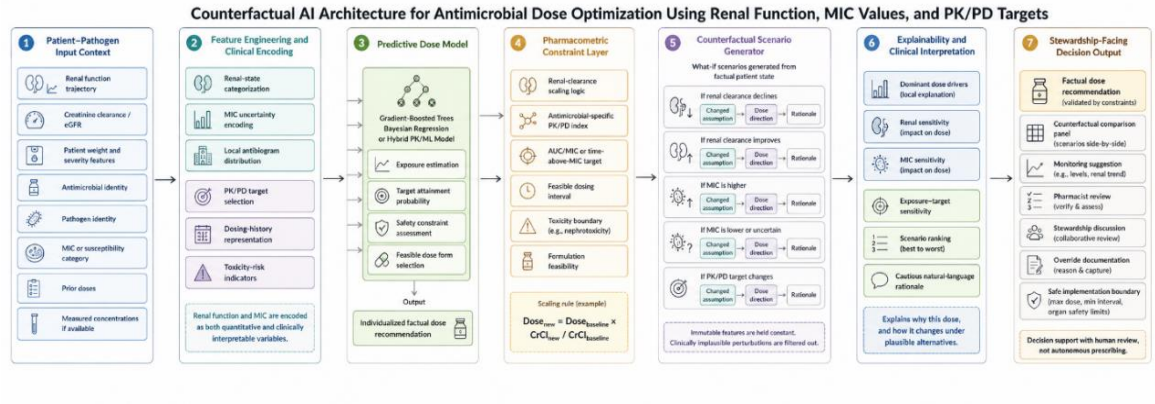

Figure 1. Counterfactual AI Architecture for Antimicrobial Dose Optimization Using Renal Function, MIC Values, and PK/PD Targets

Generating and Presenting Counterfactual Dose Scenarios

Clinically Plausible Perturbation Ranges

Clinically plausible perturbation ranges should be defined from observed physiology, recent patient trajectory, and institution-specific microbiology rather than from arbitrary mathematical intervals. Renal function could be varied within a range consistent with recent creatinine trends, acute kidney injury status, and expected drug clearance behavior, reflecting the importance of renal covariates in population pharmacokinetic models and therapeutic monitoring guidance [3, 14]. MIC perturbations should reflect measured susceptibility uncertainty, breakpoint interpretation, and local organism distributions, because MIC-based dose adjustment can be misleading when the apparent precision of the value is overinterpreted [15]. These

constraints ensure that counterfactual scenarios remain recognizable to clinicians as possible patient states rather than abstract feature manipulations.

Selecting the Most Actionable Counterfactuals

The system should rank counterfactuals by clinical usefulness, prioritizing scenarios that alter a modifiable decision, clarify uncertainty, or anticipate a plausible near-term change in patient status. Diverse counterfactual generation methods support the production of multiple alternatives, but antimicrobial dosing requires filtering those alternatives through clinical actionability, dosing feasibility, and safety logic [9]. For example, a renal-function counterfactual may be more actionable when the patient is clinically unstable, while an MIC counterfactual may be more relevant when susceptibility results are preliminary or local resistance patterns are uncertain. The goal is not to overwhelm the prescriber with all possible alternatives, but to select those that would most improve understanding of the recommended dose.

Table 1 consolidates the main counterfactual dimensions through which renal function, MIC uncertainty, PK/PD target selection, dose feasibility, and clinician override can be translated into clinically interpretable antimicrobial dosing scenarios.

Table 1. Counterfactual Dose-Reasoning Structure for Antimicrobial Optimization

Counterfactual dimension	Factual input state	Plausible counterfactual change	Expected dose-reasoning effect	Pharmacometric constraint	Clinical interpretation value
Renal function decline	Current creatinine clearance or eGFR supports the factual dose	Creatinine clearance decreases within a range consistent with recent renal trajectory or acute kidney injury risk	Dose amount or dosing frequency may need reduction to prevent excessive exposure	Counterfactual must respect renal-clearance scaling, antimicrobial elimination pathway, and toxicity boundary	Helps clinicians anticipate nephrotoxicity-sensitive dose adjustment before renal deterioration becomes severe
Renal function improvement	Current renal function suggests reduced or cautious dosing	Creatinine clearance improves or augmented renal clearance becomes plausible	Higher dose or shorter interval may be required to maintain target exposure	Change must remain biologically plausible and compatible with measured creatinine trend	Clarifies whether the recommendation is vulnerable to underexposure if renal clearance improves
MIC increase	Organism MIC is low, preliminary, or interpreted near a susceptibility threshold	MIC is modeled as higher within local susceptibility or breakpoint uncertainty	Higher exposure target or alternative dosing strategy may be needed to maintain PK/PD target attainment	MIC perturbation must reflect laboratory uncertainty, organism distribution, and antimicrobial-specific PK/PD index	Makes susceptibility uncertainty visible to prescribers and stewardship teams
MIC decrease or uncertain MIC	MIC is unavailable, inferred, or conservatively assumed	MIC is represented as lower or as a distribution rather than a fixed value	Dose may remain unchanged or become less aggressive if target attainment is robust	Counterfactual must avoid overinterpreting MIC precision and must consider local antibiogram data	Prevents false confidence in a single reported MIC value
PK/PD target intensification	Standard exposure target is selected	More aggressive AUC/MIC, %fT>MIC, or infection-severity target is selected	Dose intensity may increase, or monitoring may become more important	Target must match antimicrobial class, infection site, organism, and toxicity risk	Shows whether the model recommendation is driven by efficacy assumptions rather than renal function alone
Toxicity-aware target selection	Efficacy-oriented target is selected	Safety-conservative target is selected because of nephrotoxicity, neurotoxicity, or frailty concerns	Dose may be reduced, interval extended, or monitoring intensified	Safety constraint must be explicitly encoded and not treated as an afterthought	Supports individualized risk-benefit discussion rather than one-size-fits-all dosing
Feasible dose-form constraint	Model predicts a mathematically optimal dose	Dose is rounded to feasible vial size, infusion schedule, or institutional protocol	Recommended dose is translated into implementable clinical order	Dose must remain within formulary, infusion, and monitoring constraints	Bridges model output and real prescribing workflow
Clinician override scenario	Model recommendation is reviewed by	Clinician modifies dose because of severity, prior toxicity, source control, or clinical judgment	Override becomes an annotated decision case	Override should not automatically become a gold-standard label	Preserves accountability and supports future model governance

pharmacist or prescriber	rather than a silent rejection

Counterfactual Visualisation

Counterfactuals should be presented in an interface that allows clinicians to compare the factual recommendation with plausible alternatives in a compact and clinically familiar format. Explainable AI methods such as SHAP can support visual summaries of feature influence, while counterfactual panels can show how specific changes in renal function, MIC, or target selection would be expected to change the dose [17, 27]. A what-if table, dose-sensitivity panel, or slider could display the changed feature, the revised recommendation, and a concise pharmacometric rationale without requiring the clinician to inspect model internals. The interface should emphasize interpretability and safety, because healthcare explanations are valuable only when they support clinical reasoning rather than simply increasing the amount of information shown [7, 8]. **Figure 2** illustrates a clinically interpretable counterfactual visualisation interface that compares the factual antimicrobial dose with plausible alternatives and translates these scenarios into antimicrobial stewardship decision support.

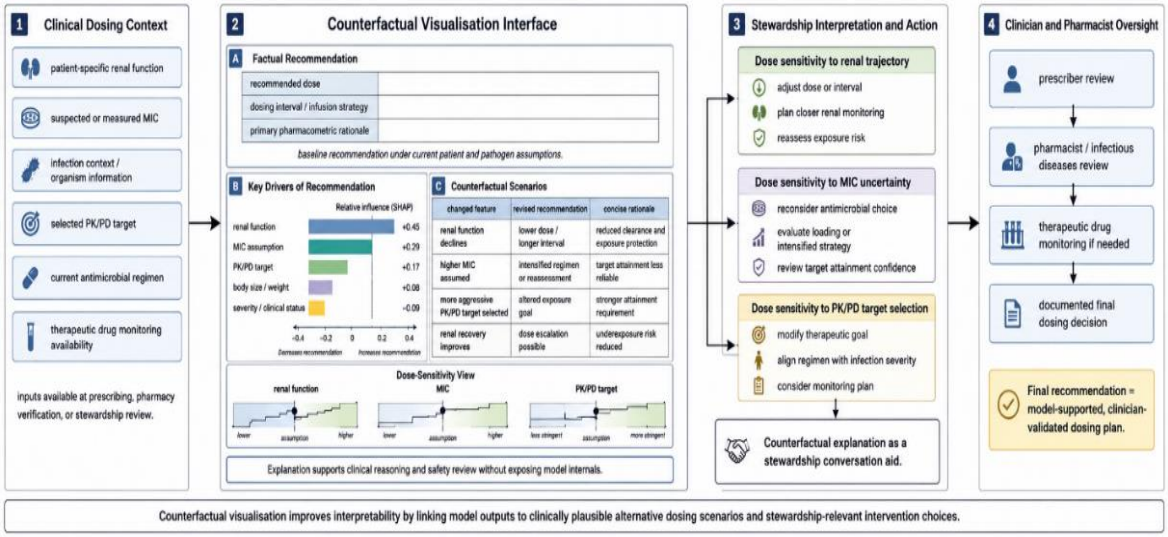

Figure 2. Counterfactual Visualisation Interface for Precision Antimicrobial Dosing and Stewardship Decision Support

Linking Counterfactuals to Stewardship Interventions

Counterfactual scenarios can support antimicrobial stewardship by showing when a dose recommendation is sensitive to organism susceptibility, renal trajectory, or target selection. Precision dosing software reviews emphasize that dosing tools are most useful when they fit into antimicrobial decision processes and provide outputs that can be interpreted by clinicians and pharmacists [13]. If a counterfactual suggests that a higher MIC assumption would make target attainment less reliable, the stewardship team could consider whether the antimicrobial, loading strategy, dosing interval, or monitoring plan remains appropriate. In this way, counterfactual explanation becomes a stewardship conversation aid rather than a stand-alone automated dosing directive.

Explainability Methods for Clinicians

Global Explanation of Feature Influence on Dose

Global explanations should describe how the model generally uses renal function, MIC, antimicrobial identity, patient size, and PK/PD targets across the population for which it is intended. SHAP-based methods are useful for summarizing feature contributions in tree models and can help identify whether renal function and susceptibility measures dominate dose recommendations in expected ways [27]. Existing vancomycin machine-learning studies show that clinical covariates can be used to estimate dosing or concentration behavior, but global explanations would be needed to confirm that the model's learned relationships are clinically coherent [21-23]. Such summaries should be reviewed with pharmacists and infectious diseases clinicians to determine whether model behavior aligns with pharmacological expectations.

Local Explanation for an Individual Recommendation

Local explanations should focus on why the current patient received the factual recommendation and which assumptions most influenced that recommendation. A local explanation might identify renal function, recent concentrations, patient size, MIC category, or selected exposure target as the dominant drivers, using visual or textual formats that reflect the clinician's dosing workflow [4, 20]. Counterfactual sensitivity can then add a contrastive layer by showing which plausible feature changes would be expected to alter the dose meaningfully. This combination of feature attribution and counterfactual explanation is important because attribution can explain the current prediction, while counterfactuals explain how fragile or stable that prediction is under clinically relevant alternatives [9, 17].

Verbal Explanations Based on Counterfactual Logic

Natural-language explanations should translate model reasoning into the language of antimicrobial pharmacotherapy. For example, the system could state that the recommendation reflects reduced renal clearance, a susceptibility assumption near a relevant MIC threshold, and a selected exposure target, while also explaining how the dose would be expected to change if renal function improved or the MIC were interpreted differently. This approach is consistent with the view that explainability in healthcare must support communication, accountability, and clinician judgment rather than merely expose model mechanics [7]. The wording should avoid unsupported certainty and instead use cautious clinical language, such as "would be expected to," "could recommend," and "should be reviewed against measured concentrations."

Feedback Loop for Clinician Overrides

Clinician overrides should be captured as structured feedback when a prescriber or pharmacist rejects, modifies, or annotates a recommendation. Machine-learning dose individualization is most useful when it can learn from real clinical practice while remaining grounded in pharmacometric principles and governance processes [11]. Overrides can be stored as clinically meaningful counterfactual examples, such as cases where a clinician accepted a lower dose because of toxicity concern or selected an alternative regimen because of infection severity. However, these data should not be treated as automatically correct labels, because overrides may reflect local practice, workflow constraints, or uncertainty rather than a universally optimal dosing decision.

Integration into Clinical Workflow and Antimicrobial Stewardship

Point-of-Prescribing Decision Support

The counterfactual dosing model should be integrated at the point of prescribing or pharmacy verification so that the recommendation is available when the antimicrobial decision is being made. Clinical decision support for precision dosing is more likely to be useful when it fits into existing workflows and provides interpretable guidance rather than requiring clinicians to leave the prescribing environment [13]. A compact display could show the factual dose, the primary rationale, and access to a counterfactual view for renal function, MIC, or target assumptions. This design would support clinical review without positioning the system as a replacement for pharmacist judgment or therapeutic drug monitoring. **Figure 3** illustrates how a counterfactual precision-dosing model can be embedded within prescribing and pharmacy verification workflows to provide interpretable antimicrobial decision support at the point of care.

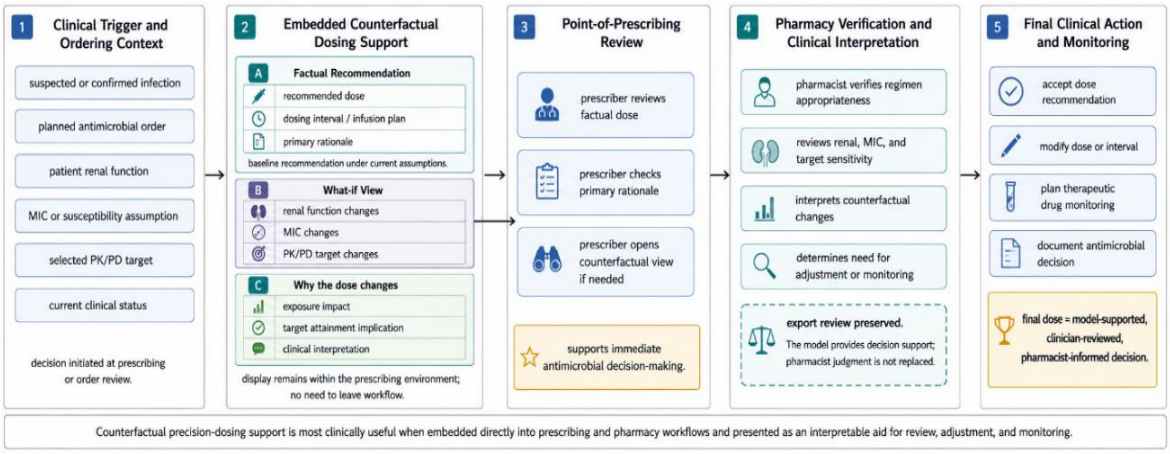

Figure 3. Workflow integration of a counterfactual precision-dosing model into prescribing and pharmacy verification, showing how factual recommendations and what-if dosing alternatives can support interpretable antimicrobial decisions without replacing pharmacist judgment or therapeutic drug monitoring.

Supporting Antimicrobial Stewardship Rounds

During antimicrobial stewardship rounds, counterfactual explanations could help pharmacists and infectious diseases clinicians explain why a dose adjustment, monitoring plan, or regimen change is being recommended. Guidance on antimicrobial therapeutic drug monitoring highlights the need for expert interpretation in critically ill patients, where drug exposure is shaped by dynamic physiology and uncertain infection characteristics [3]. Counterfactuals can make this expertise more visible by showing how renal decline, MIC uncertainty, or altered PK/PD goals would change the therapeutic reasoning. This can reinforce stewardship credibility because the recommendation is presented as a transparent pharmacological argument rather than a black-box alert.

Table 2 presents the proposed clinical functions of counterfactual dosing support at the point of prescribing, pharmacy verification, and antimicrobial stewardship review.

Table 2. Counterfactual Decision-Support Functions Across Prescribing and Antimicrobial Stewardship Workflows

Workflow location	Decision-support function	Counterfactual information displayed	Intended clinical user	Practical value for antimicrobial dosing
Point of prescribing	Provides dose recommendation while the antimicrobial order is being entered	Factual dose, suggested alternative dose, and primary rationale for adjustment	Prescriber	Supports timely dose selection without requiring the clinician to leave the prescribing environment
Pharmacy verification	Allows pharmacist review before medication release	"What would change if renal function declined, MIC increased, or PK/PD target changed?"	Clinical pharmacist	Helps verify whether the recommended dose remains appropriate under plausible clinical uncertainty
Renal function review	Tests dose sensitivity to changing kidney function	Counterfactual scenarios based on creatinine clearance, acute kidney injury, or renal recovery	Prescriber and pharmacist	Makes renal-dose adjustment logic transparent and clinically reviewable
MIC and pathogen uncertainty	Examines whether the dose remains adequate under different susceptibility assumptions	Alternative recommendations under higher MIC or uncertain organism susceptibility	Infectious diseases clinician and stewardship pharmacist	Supports antimicrobial adequacy assessment when microbiology information is incomplete or evolving
PK/PD target selection	Compares recommendations under different pharmacodynamic goals	Dose changes under alternative exposure targets, such as concentration- or time-dependent targets	Pharmacist, infectious diseases clinician	Connects dosing recommendations to pharmacological reasoning rather than generic alert logic
Antimicrobial stewardship rounds	Supports explanation of dose adjustment, monitoring, or regimen change	Plain-language counterfactual explanation of why the recommendation changes under different assumptions	Stewardship team	Improves credibility by presenting the recommendation as an interpretable pharmacological argument
Therapeutic drug monitoring planning	Identifies when dosing uncertainty requires drug-level follow-up	Scenarios showing when predicted exposure becomes unstable or clinically uncertain	Pharmacist and stewardship team	Reinforces that the model complements, rather than replaces, expert judgment and therapeutic drug monitoring
Human oversight and documentation	Records the reasoning behind accepted or rejected recommendations	Summary of factual recommendation, counterfactual drivers, and clinician decision	Prescriber, pharmacist, stewardship team	Supports accountable decision-making and preserves clinician authority over final dosing decisions

Evaluation Strategy

Predictive Accuracy and Clinical Acceptability

The predictive component should be evaluated retrospectively and prospectively for whether its recommendations are clinically acceptable, pharmacometrically plausible, and aligned with expert review. Studies of machine-learning vancomycin dosing and concentration prediction demonstrate the kinds of model-development pathways that could inform validation, but this conceptual framework should avoid claiming effectiveness without direct evaluation [24-26, 28]. Evaluation should include agreement with expert dosing decisions, consistency with exposure targets, and review of cases where the model recommendation conflicts with clinical expectations. Clinical acceptability should be judged not only by prediction error, but also by whether pharmacists can understand and safely act on the recommendation.

Counterfactual Quality and Plausibility

Counterfactual quality should be evaluated using criteria such as plausibility, sparsity, diversity, actionability, and clinical safety. General counterfactual methods emphasize proximity and diversity, but antimicrobial dosing requires additional review by pharmacology and infectious diseases experts to ensure that feature changes are physiologically and therapeutically credible [9]. A counterfactual that changes MIC, renal function, and target selection simultaneously may be mathematically valid but clinically confusing, so explanation quality must be judged in relation to how clinicians actually reason about dose adjustment. The evaluation should therefore combine formal counterfactual criteria with structured expert assessment.

Prospective User Study

A prospective user study should examine whether counterfactual support improves clinician understanding, confidence, and appropriateness of antimicrobial dosing decisions in simulated or controlled clinical settings. Healthcare XAI literature cautions that explanation tools should be tested for their effect on real decision-making behavior, because explanations can create misplaced trust as well as useful transparency [7, 8]. Participants could compare cases with standard dosing support against cases with counterfactual explanations, focusing on whether the explanation helps them identify renal-function

sensitivity, MIC uncertainty, and PK/PD trade-offs. Such evaluation should remain conceptual and hypothesis-driven until formal implementation studies establish clinical value.

Table 3 provides an evaluation and governance framework for determining whether counterfactual antimicrobial dosing AI is predictive, plausible, clinically useful, safe, and responsibly integrated into stewardship workflows.

Table 3. Evaluation and Governance Framework for Counterfactual Antimicrobial Dosing AI

Evaluation domain	Core question	Suggested assessment method	Failure mode addressed	Governance implication
Predictive dose accuracy	Does the factual dose recommendation align with accepted pharmacometric and expert dosing expectations?	Retrospective validation against expert-reviewed dosing cases, concentration-based exposure estimates, and target-attainment logic	Accurate-looking recommendation that does not reflect antimicrobial PK/PD principles	Model cannot proceed to clinical pilot without expert pharmacology review
Clinical acceptability	Would pharmacists and prescribers consider the recommendation usable in real prescribing contexts?	Structured clinician review of simulated cases, Likert acceptability ratings, and disagreement analysis	Technically valid but clinically unusable dose output	Interface and recommendation format must be redesigned before deployment
Counterfactual plausibility	Are the generated what-if scenarios physiologically and microbiologically credible?	Expert review of renal-function changes, MIC assumptions, target changes, and dose feasibility	Mathematically valid but impossible or unsafe scenario	Counterfactual generator must include hard clinical constraints
Counterfactual sparsity	Does each scenario change only the minimum necessary clinical assumption?	Quantify number of changed variables and review whether each change is clinically justified	Confusing explanations that change too many features simultaneously	Scenario ranking should favor simple, interpretable contrasts
Counterfactual diversity	Does the system show distinct types of useful uncertainty rather than repeated variations of the same scenario?	Review scenario coverage across renal function, MIC, PK/PD target, and toxicity constraints	Redundant what-if outputs that add information burden	Output panel should limit scenarios to clinically distinct alternatives
Actionability	Does the counterfactual help the clinician decide whether to adjust dose, monitor, reassess MIC, or consult stewardship?	Simulated prescribing tasks comparing standard support versus counterfactual support	Explanation that is transparent but does not improve decision making	Counterfactuals should be filtered by decision relevance
Safety boundary adherence	Does the system avoid recommending unsafe exposures, infeasible orders, or unsupported antimicrobial strategies?	Automated rule checks plus pharmacist review of high-risk cases	Dose recommendation beyond therapeutic, formulary, or monitoring limits	Safety constraints must override model-generated alternatives
Explanation calibration	Does the explanation communicate uncertainty without creating excessive trust?	User study assessing confidence, comprehension, and willingness to override	Misplaced trust caused by persuasive but uncertain explanations	Interface language should use cautious clinical phrasing
Workflow integration	Can the output be used during prescribing, pharmacy verification, or stewardship rounds without disrupting care?	Observational workflow testing and time-to-interpretation measurement	Decision support that is too slow, complex, or poorly timed	Deployment should occur at clinically meaningful decision points
Override governance	Are clinician modifications captured and interpreted responsibly?	Structured override taxonomy distinguishing toxicity concern, infection severity, workflow constraint, and data uncertainty	Treating every override as a correct training label	Override data should support audit and model improvement, not automatic retraining
Prospective impact	Does counterfactual support improve dosing decisions or monitoring plans under controlled conditions?	Prospective simulation study or pilot implementation with predefined safety endpoints	Premature claims of clinical benefit	Clinical use should remain bounded until prospective evidence is available
Equity and generalizability	Does model behavior remain reliable across patient groups, care units, pathogens, and antimicrobial classes?	Subgroup validation and site-level performance comparison	Uneven model performance across populations or institutions	Local validation and monitoring are required before broader implementation

Limitations

Assumption of a Stable PK/PD Model

A central limitation is that the counterfactual engine depends on the validity of the underlying PK/PD assumptions for the population in which it is used. Critically ill patients may experience altered volume of distribution, augmented renal clearance, organ-support effects, and rapidly changing physiology, all of which can weaken simple relationships between renal function, dose, and exposure [2, 3]. Even hybrid pharmacokinetic and machine-learning approaches may fail when deployed beyond the covariate patterns represented during development [10]. Therefore, counterfactual explanations should be displayed as model-based reasoning aids rather than definitive predictions of individual drug exposure.

Dependence on Accurate MIC and Renal Function Data

The model's recommendations and counterfactual explanations are only as reliable as the renal-function and microbiology inputs from which they are generated. MIC values may be affected by testing variability and interpretive limitations, while estimated renal function may lag behind true clearance in unstable patients [14, 15]. If these inputs are uncertain, the interface should display that uncertainty and avoid presenting a single counterfactual trajectory as if it were clinically guaranteed. This limitation is especially important for stewardship use, where susceptibility assumptions and renal trends often evolve as new laboratory and clinical information becomes available.

Conclusion

A counterfactual AI model for antimicrobial dose optimization could transform dosing support from a single unexplained recommendation into a structured set of clinically interpretable alternatives. By integrating renal function, MIC values, and PK/PD targets, such a model could show why a dose is recommended and how that recommendation would be expected to change under plausible alternative patient states.

The main strength of this approach is that it aligns AI-assisted prescribing with the reasoning style of clinical pharmacokinetics and antimicrobial stewardship. Counterfactual explanations can make renal sensitivity, susceptibility uncertainty, and exposure-target assumptions explicit, allowing clinicians to evaluate the recommendation rather than merely accept or reject it.

Important challenges remain before such a system could be safely implemented. Real-time data quality, inter-institutional variability, validation across pathogens and antimicrobials, and regulatory expectations for AI-driven dosing support all require careful attention.

Implementation pilots within hospital antimicrobial stewardship programs would provide a practical pathway for studying this approach in controlled clinical workflows. Inter-institutional collaboration could also help refine dosing models, share counterfactual design principles, and accelerate responsible development of explainable precision pharmacotherapy.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Rybak MJ, Le J, Lodise TP, Levine DP, Bradley JS, Liu C, et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant Staphylococcus aureus infections: a revised consensus guideline and review by the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the Society of Infectious Diseases Pharmacists. Am J Health Syst Pharm. 2020;77(11):835-64.
2. Guilhaumou R, Benaboud S, Bennis Y, Dahyot-Fizelier C, Dailly E, Gandia P, et al. Optimization of the treatment with beta-lactam antibiotics in critically ill patients—guidelines from the French Society of Pharmacology and Therapeutics and the French Society of Anaesthesia and Intensive Care Medicine. Crit Care. 2019;23(1):104.
3. Abdul-Aziz MH, Alffenaar JW, Bassetti M, Bracht H, Dimopoulos G, Marriott D, et al. Antimicrobial therapeutic drug monitoring in critically ill adult patients: a position paper. Intensive Care Med. 2020;46(6):1127-53.
4. Nigo M, Tran HT, Xie Z, Feng H, Mao B, Rasmy L, et al. PK-RNN-V E: A deep learning model approach to vancomycin therapeutic drug monitoring using electronic health record data. J Biomed Inform. 2022;133:104166.
5. Imai S, Takekuma Y, Miyai T, Sugawara M. A new algorithm optimized for initial dose settings of vancomycin using machine learning. Biol Pharm Bull. 2020;43(1):188-93.
6. Miyai T, Imai S, Yoshimura E, Kashiwagi H, Sato Y, Ueno H, et al. Machine learning-based model for estimating vancomycin maintenance dose to target the area under the concentration curve of 400–600 mg·h/L in Japanese patients. Biol Pharm Bull. 2022;45(9):1332-9.
7. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.

8. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11):e745-50.
9. Mothilal RK, Sharma A, Tan C. Explaining machine learning classifiers through diverse counterfactual explanations. In: Proc Conf Fairness Accountability Transparency; 2020. p. 607-17.
10. Hughes JH, Keizer RJ. A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors. CPT Pharmacometrics Syst Pharmacol. 2021;10(10):1150-60.
11. Li QY, Tang BH, Wu YE, Yao BF, Zhang W, Zheng Y, et al. Machine learning: a new approach for dose individualization. Clin Pharmacol Ther. 2024;115(4):727-44.
12. Ribba B, Dudal S, Lavé T, Peck RW. Model-informed artificial intelligence: reinforcement learning for precision dosing. Clin Pharmacol Ther. 2020;107(4):853-7.
13. Jager NG, Chai MG, van Hest RM, Lipman J, Roberts JA, Cotta MO. Precision dosing software to optimize antimicrobial dosing: a systematic search and follow-up survey of available programs. Clin Microbiol Infect. 2022;28(9):1211-24.
14. Aljutayli A, Marsot A, Nekka F. An update on population pharmacokinetic analyses of vancomycin, part I: in adults. Clin Pharmacokinet. 2020;59(6):671-98.
15. Mouton JW, Muller AE, Canton R, Giske CG, Kahlmeter G, Turnidge J. MIC-based dose adjustment: facts and fables. J Antimicrob Chemother. 2018;73(3):564-8.
16. Wang Z, Ong CL, Fu Z. AI models to assist vancomycin dosage titration. Front Pharmacol. 2022;13:801928.
17. Barredo AA, Del Ser J, Gil-Lopez S, Díaz-Rodríguez N, Bennetot A, Chatila R, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion. 2020;58:82-115.
18. Matsuzaki T, Kato Y, Mizoguchi H, Yamada K. A machine learning model that emulates experts' decision making in vancomycin initial dose planning. J Pharmacol Sci. 2022;148(4):358-63.
19. Ho WH, Huang TH, Chen YJ, Zeng LY, Liao FF, Liou YC. Prediction of vancomycin initial dosage using artificial intelligence models applying ensemble strategy. BMC Bioinformatics. 2021;22(Suppl 5):637.
20. Kim D, Choi HS, Lee D, Kim M, Kim Y, Han SS, et al. A deep learning-based approach for prediction of vancomycin treatment drug monitoring: retrospective data of critically ill patients. JMIR Form Res. 2024;8:e45202.
21. Tootooni MS, Barreto EF, Wutthisirisart P, Kashani KB, Pasupathy KS. Determining steady-state trough range in vancomycin drug dosing using machine learning. J Crit Care. 2024;82:154784.
22. Chen YW, Lin XK, Huang C, Wu W, Lin WW, Chen S, et al. Vancomycin trough concentration in adult patients with periprosthetic joint infection: a machine learning-based covariate model. Br J Clin Pharmacol. 2024;90(9):2188-99.
23. Ma P, Ma H, Liu R, Wen H, Li H, Huang Y, et al. Prediction of vancomycin plasma concentration in elderly patients based on multi-algorithm mining combined with population pharmacokinetics. Sci Rep. 2024;14(1):27165.
24. Lee YW, Kim JH, Park JJ, Park H, Seo H, Kim YK. Development and external validation of a machine learning model to predict the initial dose of vancomycin for targeting an area under the concentration-time curve of 400–600 mg·h/L. Int J Med Inform. 2025;196:105817.
25. Lee H, Kim YJ, Kim JH, Kim SK, Jeong TD. Optimizing initial vancomycin dosing in hospitalized patients using machine learning approach for enhanced therapeutic outcomes: algorithm development and validation study. J Med Internet Res. 2025;27:e63983.
26. Hu T, Ding X, Han F, An Z. Machine learning approach for personalized vancomycin steady-state trough concentration prediction: a superior approach over Bayesian population pharmacokinetic model. Front Pharmacol. 2025;16:1549500.
27. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56-67.
28. van Os W, O'Jeanson A, Troisi C, Liu C, Brooks JT, Hughes JH, et al. Machine learning-based model selection and averaging outperform single-model approaches for a priori vancomycin precision dosing. CPT Pharmacometrics Syst Pharmacol. 2025;14(10):1650-60.