Available online **www.ijpras.com**

International Journal of Pharmaceutical Research & Allied Sciences, 2026, 15(2):61-72
https://doi.org/10.51847/Ppht0M5a9H

Review Article

ISSN : 2277-3657
CODEN(USA) : IJPRPM

Artificial Intelligence in Clinical Pharmacokinetics and Pharmacodynamics: A Realist Review of Contexts, Mechanisms, Outcomes, and Failure Modes

Carlos Ramirez[1]*, Elena Torres[2], Pablo Ortega[1], Sofia Mendes[3]

[1]*Department of AI in Clinical PK/PD and Realist Review, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain.*
[2]*Department of Contexts and Mechanisms in AI Modeling, Faculty of Pharmaceutical Sciences, University of Lisbon, Lisbon, Portugal.*
[3]*Department of Outcomes and Failure Modes, Faculty of Pharmacy, University of Porto, Porto, Portugal.*

***Email:** carlos.ramirez@ub.edu*

ABSTRACT

Artificial intelligence is increasingly used to predict pharmacokinetic parameters, characterize exposure–response relationships, select population models, interpret therapeutic drug-monitoring data, and support individualized dosing. However, analytical performance alone does not explain why apparently similar systems succeed in one setting, fail in another, or create unintended clinical consequences. This realist review examines how data, model, patient, organizational, and decision contexts influence the mechanisms and outcomes of artificial-intelligence-enabled pharmacokinetic and pharmacodynamic applications. The initial programme theory proposes that these applications are most likely to support useful decisions when a clearly defined clinical task is matched to representative longitudinal data, appropriate pharmacological structure, calibrated uncertainty, explicit therapeutic targets, and workflows that allow clinicians to interrogate or reject recommendations. Evidence was identified through purposive and iterative searching, appraised for relevance to programme-theory development and rigour of the supporting inference, and synthesized through context–mechanism–outcome configurations, negative cases, demi-regularities, and rival explanations. The reviewed evidence suggests that performance may improve when machine learning captures residual heterogeneity, integrates complex covariates, or reduces dependence on a single misspecified population model. Conversely, incomplete sampling, biased labels, inappropriate priors, weak validation, miscalibration, distribution shift, and automation-related reliance may amplify errors as predictions are converted into doses. The refined programme theory therefore treats artificial intelligence not as an autonomous dosing intervention but as a context-dependent resource embedded within pharmacological models, measurement systems, clinical reasoning, governance, and monitoring. Its clinical value remains conditional on alignment across these components and cannot be inferred from computational performance alone.

Key words: *Clinical pharmacokinetics, Pharmacodynamics, PK/PD, Dose–exposure–response, Population modelling, Precision dosing*

Received: 19 February 2026; **Revised:** 12 May 2026; **Accepted:** 18 May 2026

INTRODUCTION

Artificial intelligence is entering clinical pharmacology through heterogeneous applications that include covariate identification, pharmacokinetic parameter estimation, longitudinal exposure prediction, treatment-response modelling, therapeutic drug-monitoring interpretation, and dose-decision support. These applications differ substantially in their data requirements, proximity to patient care, dependence on pharmacological assumptions,

and consequences of error. Treating them collectively as a single technological intervention can therefore obscure the difference between an algorithm that assists model development and a system that directly influences an individual dose [1].

Prevailing discussions frequently emphasize predictive accuracy, computational flexibility, or comparisons with conventional statistical models. Such framing is insufficient for clinical pharmacokinetics and pharmacodynamics because an apparently accurate output becomes meaningful only in relation to a defined patient population, observation process, exposure or response target, therapeutic window, and clinical action. A useful system must consequently be evaluated not only by what it predicts, but also by how its intended problem was formulated, whether its validation reflects the intended use, and how its output enters clinical reasoning [2].

Model-informed precision dosing provides a particularly important decision context. It combines prior pharmacokinetic or pharmacodynamic knowledge with patient characteristics, concentration measurements, biomarkers, or treatment responses to estimate an individual state and support dose adjustment. Although artificial intelligence may extend this process through nonlinear representation learning, model selection, residual correction, or pattern recognition, the broader precision-dosing workflow still depends on qualified models, reliable measurements, appropriate therapeutic targets, usable software, and institutional implementation capacity [3].

The central question is therefore not whether artificial intelligence "works" in clinical PK/PD in the abstract, but what configurations may work, for whom, under which circumstances, through which mechanisms, and with which intended or unintended outcomes. A realist-review design is appropriate because it treats outcomes as contingent products of interactions among context, intervention resources, actor or system responses, and implementation processes rather than as uniform effects of a technology [4]. This review develops and refines an explanatory programme theory while maintaining clear boundaries between analytical performance, clinical utility, causal interpretation, implementation activity, and patient benefit.

Realist-review questions and initial programme theory

The review addressed six linked questions: what constitutes the initial programme theory for artificial-intelligence-enabled PK/PD; in which clinical, data, model, and organizational contexts these resources operate; which resources trigger relevant reasoning, behavioural, informational, or system responses; which proximal and distal outcomes follow; why apparently similar implementations generate different or unintended results; and how the programme theory should be refined and bounded. In realist terms, an algorithm is not itself a mechanism. It is a resource that may alter inference, confidence, attention, model choice, clinician behaviour, or workflow when introduced into a particular context.

The initial programme theory proposes that artificial-intelligence-enabled PK/PD is most likely to support useful decisions when the clinical question is explicit, observations represent the intended population and time course, pharmacological structure constrains or complements data-driven learning, uncertainty is calibrated for the relevant decision range, and recommendations are presented in a form that clinicians can interrogate, contextualize, defer, or override. The same resource may fail when samples are sparse or mistimed, covariates are missing or stale, population-model priors are mismatched, targets are weak proxies for benefit, or interfaces encourage uncritical reliance. **Figure 1** presents the original conceptual synthesis for context–mechanism–outcome map for AI-enabled PK/PD, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

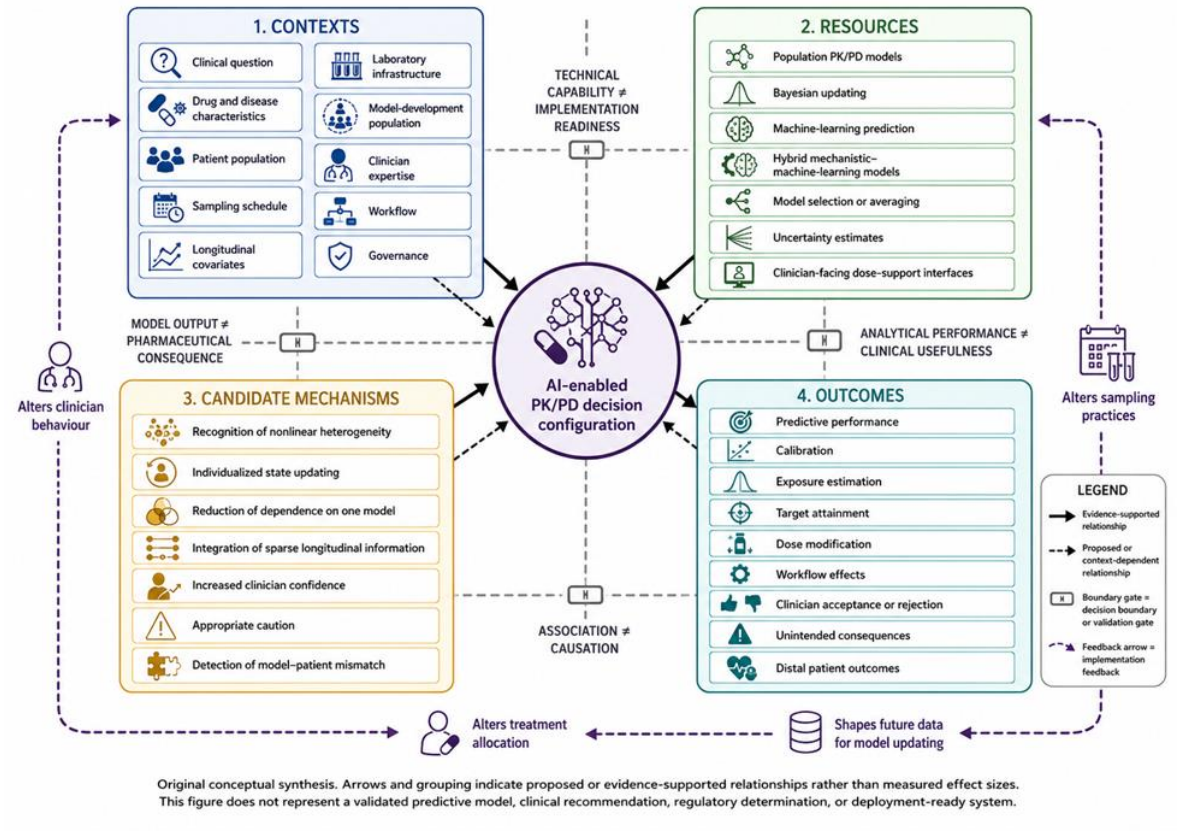

Figure 1. Context–Mechanism–Outcome Map for AI-Enabled PK/PD

The figure distinguishes intervention resources from candidate mechanisms because computational operations do not automatically explain outcomes. For example, a Bayesian forecast or machine-learning prediction supplies information, whereas the mechanism may involve a clinician revising confidence, recognizing a mismatch, requesting another sample, or changing a dose. Likewise, a technically strong model may remain inactive when its result arrives too late, cannot be integrated into workflow, or conflicts with an unexplained clinical observation. The initial theory is therefore a set of propositions for interrogation against the literature, not a validated causal model or universal description of AI-enabled dosing.

Search, selection, appraisal, and extraction methods

Searching followed a purposive, iterative realist strategy. Initial searches combined artificial intelligence, machine learning, clinical pharmacokinetics, pharmacodynamics, population modelling, PK/PD, therapeutic drug monitoring, model-informed precision dosing, dose–exposure–response relationships, and clinical decision support. Subsequent searches were progressively refined around emerging explanatory concepts, including hybrid modelling, model selection, model averaging, time-varying covariates, calibration, external validation, transportability, clinician interaction, automation-related failure, bias, uncertainty, digital representations of patient state, and implementation barriers. Citation chaining and targeted journal searching were used to locate primary studies, methodological analyses, and reviews capable of testing or refining the programme theory.

Evidence was eligible when it contributed directly to at least one component of a context–mechanism–outcome explanation or supplied evidence about modelling configuration, validation, implementation, clinical interpretation, failure, or unintended consequences. Selection was guided by theoretical relevance rather than by the expectation that every source would estimate the same intervention effect. Primary studies were prioritized for specific claims about model behaviour, validation, or implementation, whereas reviews and methodological articles were used to frame field-level configurations and appraisal boundaries. Sources whose conclusions could not be connected to a defined clinical PK/PD task, context, mechanism, outcome, or rival explanation were not used in the synthesis.

Appraisal separated relevance from rigour. Relevance concerned whether a source could inform, challenge, or refine a programme-theory proposition. Rigour concerned whether the methods and data supporting the particular inference were sufficiently credible for that purpose; it was not reduced to a single conventional quality score [4].

Extraction captured the clinical setting, patient population, drug or intervention, PK/PD task, observation structure, covariates, model architecture, comparator, validation design, implementation conditions, reported outcomes, negative findings, explanatory statements, and plausible rival explanations. Synthesis proceeded through comparison of positive, mixed, and negative cases; identification of recurring but nonuniversal patterns; reconciliation of apparently conflicting findings; and progressive refinement of the programme theory.

AI configurations across PK/PD tasks

Artificial intelligence in pharmacometrics encompasses several configurations rather than one model class. Applications include covariate screening, parameter estimation, surrogate modelling, exposure prediction, longitudinal-response modelling, structural-model support, population-model selection, individual Bayesian forecasting, and dose-decision assistance. The potential value of each configuration depends on its position in the analytical and clinical pathway. A model used to explore covariates during development presents different evidentiary and safety requirements from a model that generates a patient-specific dose recommendation [5].

Data-driven PK applications also span distinct levels of translation. Some models predict absorption, distribution, metabolism, clearance, or concentration profiles from molecular, preclinical, or patient-level variables. Others use longitudinal clinical observations to estimate exposure or response. Flexible learning may capture nonlinear associations and high-order interactions that are difficult to prespecify, but such flexibility does not establish pharmacological plausibility, causal relevance, or transportability beyond the development data [6]. A strong retrospective fit may therefore represent useful pattern recognition, overfitting, or exploitation of context-specific correlations.

Deep compartment models illustrate a structure-aware configuration in which neural time-series learning is combined with compartment-like representations of pharmacokinetic behaviour. This architecture may provide greater flexibility than a fixed parametric model while retaining an interpretable temporal organization of drug movement. Its immediate outcome is improved or more stable prediction of concentration–time data under the evaluated conditions, not proof that the learned representation corresponds to a unique biological mechanism or that its use improves dosing decisions [7].

Hybrid machine-learning and pharmacokinetic approaches provide another configuration. Rather than replacing the population model, machine learning may modify the influence of its priors, learn residual patterns, or identify circumstances in which a conventional model becomes unreliable. Selectively flattening an unsuitable prior can improve individual prediction when the structural model remains informative but its population assumptions are locally mismatched [8]. The candidate mechanism is not simply “use of machine learning”; it is the reduction of harmful prior dominance while preserving useful pharmacological structure.

These configurations can be organized by four dimensions: the source and temporal structure of data; the amount of embedded pharmacological knowledge; the point at which the model enters the decision pathway; and the consequence of an incorrect output. Purely data-driven, mechanistically constrained, and hybrid systems should not be ranked as universally superior because each may succeed under different conditions. The reviewed evidence instead supports a configuration-specific interpretation: architecture becomes relevant through its interaction with data quality, model–population fit, validation design, therapeutic target, and intended decision use [5]. Computational performance therefore remains an intermediate outcome whose pharmaceutical and clinical significance must be established separately.

Data and implementation contexts

Model-informed precision dosing depends on more than an executable model. Its operating context includes qualified pharmacokinetic or pharmacodynamic knowledge, timely patient observations, appropriate targets, usable software, data governance, and a workflow capable of acting on the output. When one component is absent, technically credible predictions may remain unused or may be applied outside their intended conditions [9].

Implementation is also shaped by clinician knowledge, trust, workload, interoperability, organizational leadership, and local resources. These factors can activate mechanisms such as confidence, attention, and coordinated action, or opposing mechanisms such as resistance, workarounds, and delayed use. Bedside adoption therefore cannot be inferred from model performance or software availability alone [10].

Therapeutic drug monitoring becomes model-informed when measured concentrations are interpreted through an appropriate population model, patient-specific information, and an explicit exposure target. Sample timing, assay reliability, turnaround time, and access to dosing expertise determine whether the measurement reduces uncertainty or merely adds another data point to the record [11].

Primary-care levothyroxine modelling illustrates a different context in which routinely available variables and a relatively simple dosing workflow may support machine-learning prediction. Nevertheless, development, validation, and clinical simulation do not independently establish prospective patient benefit, particularly when adherence, formulation changes, comorbidity, and time-varying thyroid status may affect the observed dose requirement [12]. **Table 1** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for initial programme theory, contexts, mechanisms, and outcome categories for artificial intelligence in clinical pharmacokinetics and pharmacodynamics.

Table 1. Initial programme theory, contexts, mechanisms, and outcome categories for Artificial Intelligence in Clinical Pharmacokinetics and Pharmacodynamics

Evidence domain or review construct	Review question	Eligible evidence type	Key extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Clinical task	What decision is being supported?	Clinical-pharmacology and dosing studies	Drug, population, prediction target, decision point	Direct connection between output and intended use	Defines the programme boundary	Tasks labelled "AI" may be clinically dissimilar	Analytical success cannot be generalized across tasks [1].
Model-informed dosing context	Which resources are required?	MIPD reviews and implementation studies	Prior model, observations, target, updating process	Evidence that the components form an actionable workflow	Identifies enabling resources	Components may be present without effective use	MIPD infrastructure does not prove AI-specific benefit [9].
Data and measurement context	When do observations reduce uncertainty?	TDM and longitudinal PK studies	Sampling time, assay, covariates, missingness, turnaround	Relevance to the patient state and decision time	Explains informational mechanisms	Dense data may still be biased or mistimed	Measurement availability is not equivalent to information quality [11].
Organizational context	What activates or blocks use?	Implementation reviews and qualitative evidence	Training, trust, workload, interoperability, leadership	Credible connection to adoption or sustained use	Explains behavioural and system responses	Reported barriers may vary across institutions	Implementation activity is not patient benefit [10].
Model configuration	How is pharmacological knowledge combined with learning?	PopPK, hybrid-model and ML studies	Architecture, priors, features, comparator, validation	Fit between configuration and claimed mechanism	Distinguishes alternative resources	Architecture may be described more fully than context	Flexibility does not establish mechanistic truth [5].
Candidate mechanism—proposed synthesis	What response does the resource trigger?	Explanatory fragments across evidence types	Updating, confidence, caution, model weighting, workflow response	Plausibility and consistency across positive and negative cases	Connects context to outcomes	Mechanisms are often not measured directly	An algorithmic operation is not itself a realist mechanism
Outcome hierarchy—proposed synthesis	Which outcomes follow?	Prediction, dosing, workflow and clinical studies	Calibration, exposure, dose change, workflow, safety, benefit	Clear separation of proximal and distal outcomes	Prevents outcome conflation	Distal outcomes are less frequently evaluated	Prediction, dosing utility and patient benefit must remain distinct
Boundary conditions—proposed synthesis	Where should inference stop?	Validation and methodological evidence	Population shift, calibration, uncertainty, interface, monitoring	Evidence relevant to intended use	Defines limits of programme theory	Boundaries may change after implementation	No configuration is presumed validated, universally applicable, or deployment-ready

Mechanisms of performance improvement

Hybrid models may improve performance when pharmacometric structure and data-driven learning contribute complementary information. In iohexol clearance estimation, population pharmacokinetics supplied an interpretable structural foundation while machine learning captured additional relationships among patient characteristics. The proposed mechanism is improved representation of residual heterogeneity rather than replacement of pharmacological knowledge [13].

A comparable configuration was applied to isavuconazole exposure prediction. The hybrid approach suggests that mechanistic priors may stabilize estimation while machine learning captures nonlinear variation not adequately represented in the base model. The evidence supports prediction within the examined setting, but it does not establish that the learned relationships are causal or transportable to substantially different populations [14].

Model selection offers a second performance mechanism. When several population models are available, a machine-learning procedure may select or weight the model that better matches the current patient or dataset. In vancomycin dosing, this approach reduced dependence on an arbitrarily selected single model and improved a priori prediction under the evaluated conditions [15].

Model averaging addresses related uncertainty by combining candidate predictions rather than committing to one structural model. Its benefit is most plausible when models make complementary errors and none is consistently dominant. Improved predictive performance remains a proximal outcome, however, and cannot alone demonstrate safer dose selection, exposure-target attainment, or patient benefit [16].

Mechanisms of error amplification and workflow failure

The same flexibility that supports prediction can amplify error when data provenance, intended use, and implementation assumptions are poorly aligned. Clinical-pharmacology AI may inherit biased observations, inappropriate labels, model misspecification, and weak comparators before adding further uncertainty through opaque transformations or unsupported extrapolation [17].

Bias can enter through patient selection, measurement practices, underrepresentation, outcome definition, and deployment. A model may therefore perform differently across groups even when its overall summary metric appears acceptable. Once connected to dosing, unequal error may become unequal exposure, delayed correction, or differential clinician reliance [18].

Methodological weaknesses can create a misleading impression of readiness. Small or unrepresentative samples, inappropriate splitting, incomplete handling of missing data, overfitting, and inadequate external evaluation are common sources of optimistic performance in supervised clinical prediction research [19]. Such limitations are especially consequential when the model output is converted into a treatment action.

Calibration is a separate requirement from discrimination or average error. A model that ranks patients adequately may still systematically overestimate or underestimate exposure, response, or risk at the point where a dose decision is made [20]. **Figure 2** presents the original conceptual synthesis for failure pathways from data context to unsafe dose recommendation, showing how the article's principal components, evidence relationships, uncertainties, and decision boundaries are connected.

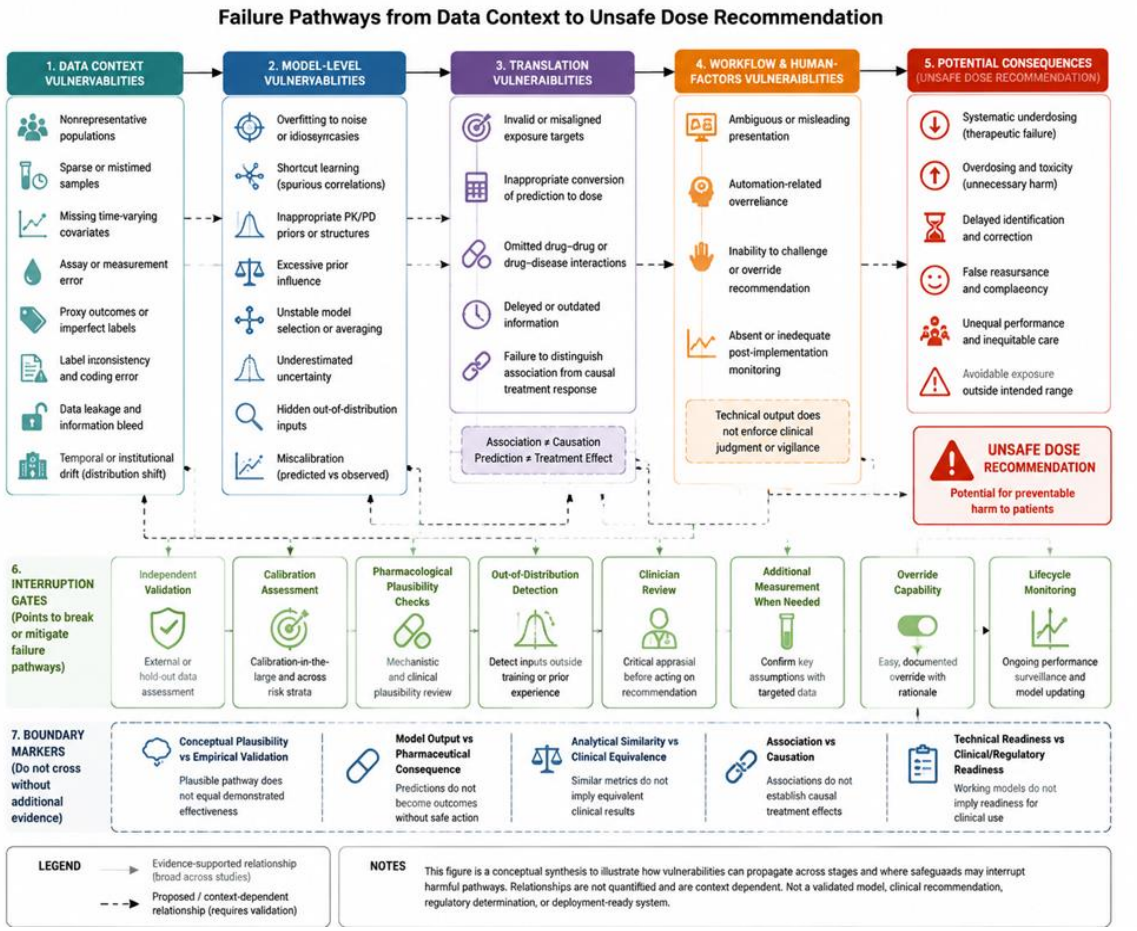

Figure 2. Failure Pathways from Data Context to Unsafe Dose Recommendation

Context–mechanism–outcome configurations

In critically ill patients receiving vancomycin, rapidly changing physiology, intermittent measurements, and urgent dose decisions create a context in which a locally usable Bayesian tool may provide actionable information. When clinicians receive timely concentration-informed forecasts and recognize them as relevant, the resource may trigger dose reconsideration. Prospective validation supports feasibility within the examined context but does not establish universal transportability [21].

External comparison of vancomycin forecasting models provides a contrasting configuration. Models intended for a similar task performed differently when evaluated in a common dataset, suggesting that development population, parameterization, residual-error structure, and sampling design condition the usefulness of the prior model. Nominal task similarity therefore does not ensure equivalent performance [22].

For oral targeted oncology therapies, therapeutic drug monitoring can reveal underexposure or overexposure only when a credible target, feasible sampling process, and actionable adjustment pathway exist. The mechanism is not measurement alone but clinician interpretation of a result that can be linked to an appropriate intervention. Prospective multicentre evidence indicates feasibility while also showing that applicability remains drug- and context-dependent [23].

In infliximab dosing, multi-model averaging offers a response to disagreement among candidate pharmacokinetic models. Where no single model transports reliably across all patients, combining predictions may reduce sensitivity to structural uncertainty. The resulting improvement remains tied to the evaluated population, target, and dosing framework [24].

Across intensive-care vancomycin dosing, comparative Bayesian forecasting, and oncology therapeutic drug monitoring, useful precision dosing depended on alignment among the target population, model, sampling process, exposure target, and actionable clinical pathway [21-23]. These demi-regularities support conditional rather than universal CMO propositions. **Table 2** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for context–mechanism–outcome configurations supported by the selected evidence.

Table 2. Context–mechanism–outcome configurations supported by the selected evidence

Method, platform, or evidence category	Representative application	Reported strength	Validation context	Generalization concern	Contradictory evidence	Evidence gap
Bayesian MIPD	Vancomycin in intensive care	Prospective clinical implementation and individualized forecasting	Local ICU workflow and concentration sampling	Rapid physiology and institutional practice may differ elsewhere	Similar tools may perform differently with other models or populations	Comparative patient outcomes across settings
External model comparison	Vancomycin Bayesian forecasting	Direct comparison of multiple published models	Common external dataset	One evaluation population cannot represent all future settings	Candidate models showed unequal predictive suitability	Prospective consequences of choosing the wrong model
TDM-based precision dosing	Oral targeted oncology therapies	Prospective multicentre feasibility	Drug-specific targets and follow-up pathways	Target validity and adjustability vary by therapy	Not every measured deviation permits a meaningful dose response	Long-term clinical and safety consequences
Multi-model averaging	Infliximab dosing	Reduced dependence on one candidate model	Inflammatory bowel disease datasets	Weights may shift across populations and assays	Individual models remain variably applicable	Prospective benefit and monitoring after updating
Hybrid PopPK–ML	Iohexol clearance	Complementary structural and nonlinear representation	Application-specific clinical data	Learned residual relationships may be local	Improvement may reflect feature availability rather than a general mechanism	Independent transportability and dosing consequence
Hybrid exposure prediction	Isavuconazole	Flexible prediction under variable observations	Evaluated clinical dataset	Drug interactions and population differences may shift relationships	No direct evidence of universal superiority	Prospective dose and outcome evaluation
Model selection or averaging	Vancomycin precision dosing	Reduced single-model dependence	Comparative pharmacometric evaluation	Candidate library may omit the best model for a new context	Selection can become unstable under distribution shift	Monitoring and recalibration rules

Calibration, transportability, and clinician interaction

External validation is not a binary certificate. Its meaning depends on how the validation population, institution, time period, measurement process, and intended decision differ from development conditions. A model may validate in one related setting yet remain inappropriate for a population with different covariate distributions, therapeutic targets, or clinical workflows [25].

Assessment of healthcare prediction models should therefore extend beyond headline accuracy. Data provenance, participant selection, predictor availability, missingness, comparator choice, calibration, external evaluation, usability, transparency, and monitoring all affect whether an output can support its intended decision [26].

Uncertainty must also be characterized rather than hidden behind a point prediction. Where inputs are unfamiliar, measurements are sparse, or candidate models disagree, a useful system may need to signal uncertainty, defer a recommendation, or request additional information. Such behaviour may support caution but does not compensate for weak development or validation [27].

Clinician interaction is consequently part of the intervention rather than a final presentation step. Decision support should preserve the ability to inspect assumptions, compare recommendations with the patient's observed state, and reject or override an output. Explanations may assist reasoning, but they cannot substitute for calibration, transportability, or pharmacological validity [28].

Negative cases and unintended consequences

Machine learning may alter care even when no conventional software malfunction occurs. Repeated reliance can change attention, expertise, ordering behaviour, documentation, and the future data available for learning. False

reassurance, selective automation, and loss of clinical vigilance are therefore plausible outcomes of implementation rather than purely technical errors [29].

Claims that artificial intelligence matches or exceeds clinicians are often difficult to interpret when studies use retrospective datasets, weak comparators, selective reporting, or nonrepresentative clinical tasks. These designs can create apparent superiority without demonstrating that clinicians and models were compared under equivalent information, time, and responsibility [30].

Negative cases from medical AI also show how leakage, nonindependent data splitting, confounding, and narrow external testing can produce misleading performance. Although such examples do not prove identical failures in PK/PD, the underlying methodological vulnerabilities transfer directly to longitudinal concentration, covariate, and outcome modelling [31].

Responsible deployment therefore requires lifecycle attention to problem formulation, data generation, evaluation, workflow, accountability, updating, and monitoring [32]. Across medical AI, apparent performance can outpace methodological and clinical validity when data provenance, comparator design, independence of evaluation, and external testing are weak [29-31]. **Table 3** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for rival explanations, boundary conditions, negative cases, and refined programme theory for artificial intelligence in clinical pharmacokinetics and pharmacodynamics.

Table 3. Rival explanations, boundary conditions, negative cases, and refined programme theory for Artificial Intelligence in Clinical Pharmacokinetics and Pharmacodynamics

Cross-study pattern or configuration	Context or boundary condition	Mechanism or explanatory factor	Observed or reported outcome	Rival explanation	Certainty or rigour concern	Implication
High apparent performance	Retrospective development and internal testing	Flexible fitting captures dataset-specific structure	Strong reported prediction	Leakage, overfitting, weak comparator	Inadequate independent validation	Performance should not be equated with transportability
Acceptable ranking but inaccurate magnitude	Decision depends on predicted exposure or risk level	Miscalibration shifts the decision-relevant estimate	Potentially inappropriate threshold crossing	Population shift or incorrect outcome model	Calibration may be incompletely reported	Calibration is essential for dose consequence
Successful external validation	Validation context remains close to development	Similar data generation preserves performance	Stable validation result	Limited contextual challenge	"External" does not imply broad transportability	Applicability must be tied to intended use
Clinician acceptance	Interface provides a clear recommendation	Confidence and cognitive off-loading	Increased use	Automation-related overreliance	Behavioural mechanisms rarely measured	Adoption may coexist with reduced scrutiny
AI appears superior to clinicians	Retrospective or unequal comparison	Algorithm uses standardized data unavailable during routine care	Better reported metric	Comparator asymmetry or selective task definition	Reporting and design limitations	Superiority claims require equivalent conditions
Nontransportable success	Narrow or confounded dataset	Shortcut learning	Performance collapse under shift	Site-specific artefact rather than clinical signal	External testing may be absent or weak	Provenance and independence are central
Responsible lifecycle—proposed refinement	High-consequence dose decisions	Monitoring, escalation and accountability interrupt error propagation	Earlier detection or containment of failure	Organizational controls may exist only formally	Effectiveness of safeguards remains uncertain	Governance must remain active after implementation

Refined programme theory and implementation principles

Transparent reporting is a precondition for judging whether an AI-enabled prediction model is reproducible and applicable. Reports should describe the intended use, participants, data, predictors, outcomes, modelling procedures, evaluation, limitations, and role of human interaction sufficiently clearly for readers to assess the claimed decision function [33].

Structured appraisal must additionally examine bias and applicability across participants, predictors, outcomes, analysis, and AI-specific design choices. Such appraisal does not validate a model; it identifies whether available evidence supports confidence in the particular claim and target context [34].

Current PK/PD scholarship presents machine learning as a family of data-driven and hybrid configurations spanning prediction, model refinement, and decision support. The literature also continues to identify limitations in external validation, uncertainty characterization, interpretability, and prospective clinical evaluation [35].

The refined programme theory is therefore conditional. AI-enabled PK/PD may improve a proximal modelling or dosing outcome when a narrowly defined task is matched to representative longitudinal data, appropriate pharmacological structure, relevant validation, calibrated uncertainty, credible targets, and an actionable workflow. Candidate mechanisms include improved representation of residual heterogeneity, more appropriate weighting of candidate models, individualized updating, and clinician recognition of actionable information.

These mechanisms may remain inactive or reverse when observations are biased or delayed, priors are mismatched, evaluation is nonindependent, uncertainty is concealed, or a prediction is converted automatically into a dose. Implementation should consequently preserve pharmacological plausibility checking, clinician contestability, additional measurement where uncertainty is high, explicit escalation pathways, version control, monitoring for drift, and reassessment after material changes. These are proposed implementation principles requiring prospective evaluation, not a validated clinical protocol.

Review limitations

Realist synthesis depends on how completely source studies describe context, implementation, actor responses, and explanatory processes. Many reports emphasize model architecture and predictive results while providing limited evidence about clinician reasoning or workflow mechanisms, restricting the precision of CMO inference [4].

Evidence heterogeneity across drugs, populations, sampling designs, targets, comparators, and validation methods limits direct cross-study comparison. Methodological weaknesses reported in clinical prediction research also reduce confidence that all apparent improvements would persist under independent evaluation [19].

Transferability remains uncertain because external validation varies in contextual distance and intended use. A model evaluated outside its development dataset may still be insufficiently tested for a different institution, patient population, measurement process, or dosing consequence [25].

The review may also underrepresent unpublished failures and operational experience that did not produce journal articles. Rapid methodological development may change available architectures and guidance, while robustness and uncertainty practices remain heterogeneous [27]. The synthesis should therefore be interpreted as an evidence-bounded explanatory account rather than an exhaustive or permanent map of the field.

CONCLUSION

Artificial intelligence in clinical PK/PD cannot be understood as a uniform intervention whose value follows directly from predictive performance. Its effects arise from configurations linking patient and data context, pharmacological structure, model behaviour, validation, therapeutic targets, clinician reasoning, workflow, and governance. The reviewed evidence supports conditional mechanisms through which hybrid learning, individualized updating, and model selection or averaging may improve proximal predictions. It also identifies pathways through which biased data, mismatched priors, weak validation, miscalibration, hidden uncertainty, and inappropriate reliance may propagate toward unsafe recommendations. The refined programme theory therefore positions artificial intelligence as a decision resource whose usefulness depends on alignment, contestability, and continuing monitoring. Prospective evaluation must preserve the distinction between computational accuracy, dosing utility, implementation success, and patient benefit.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

1. Ryan DK, Maclean RH, Balston A, Scourfield A, Shah AD, Ross J. Artificial intelligence and machine learning for clinical pharmacology. Br J Clin Pharmacol. 2024;90(3):629-39. doi:10.1111/bcp.15930
2. Shahin MH, Barth A, Podichetty JT, Liu Q, Goyal N, Jin JY, et al. Artificial intelligence: From buzzword to useful tool in clinical pharmacology. Clin Pharmacol Ther. 2024;115(4):698-709. doi:10.1002/cpt.3083
3. Darwich AS, Polasek TM, Aronson JK, Ogungbenro K, Wright DFB, Achour B, et al. Model-informed precision dosing: Background, requirements, validation, implementation, and forward trajectory of individualizing drug therapy. Annu Rev Pharmacol Toxicol. 2021;61:225-45. doi:10.1146/annurev-pharmtox-033020-113257
4. Duddy C, Wong G. Grand rounds in methodology: When are realist reviews useful, and what does a 'good' realist review look like? BMJ Qual Saf. 2023;32(3):173-80. doi:10.1136/bmjqs-2022-015236
5. McComb M, Bies R, Ramanathan M. Machine learning in pharmacometrics: Opportunities and challenges. Br J Clin Pharmacol. 2022;88(4):1482-99. doi:10.1111/bcp.14801
6. Ota R, Yamashita F. Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. J Control Release. 2022;352:961-9. doi:10.1016/j.jconrel.2022.11.014
7. Janssen A, Leebeek FWG, Cnossen MH, Mathôt RAA, OPTI-CLOT Study Group, SYMPHONY Consortium. Deep compartment models: A deep learning approach for the reliable prediction of time-series data in pharmacokinetic modeling. CPT Pharmacometrics Syst Pharmacol. 2022;11(7):934-45. doi:10.1002/psp4.12808
8. 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. doi:10.1002/psp4.12684
9. Minichmayr IK, Dreesen E, Centanni M, Wang Z, Hoffert Y, Friberg LE, et al. Model-informed precision dosing: State of the art and future perspectives. Adv Drug Deliv Rev. 2024;215:115421. doi:10.1016/j.addr.2024.115421
10. Dibbets AC, Koldeweij C, Osinga EP, Scheepers HCJ, de Wildt SN. Barriers and facilitators for bringing model-informed precision dosing to the patient's bedside: A systematic review. Clin Pharmacol Ther. 2025;117(3):633-45. doi:10.1002/cpt.3510
11. Wicha SG, Märtson AG, Nielsen EI, Koch BCP, Friberg LE, Alffenaar JWC, et al. From therapeutic drug monitoring to model-informed precision dosing for antibiotics. Clin Pharmacol Ther. 2021;109(4):928-41. doi:10.1002/cpt.2202
12. Janssen Daalen JM, Doesburg D, Hunik L, Kessel R, Herngreen T, Knol D, et al. Model-informed precision dosing using machine learning for levothyroxine in general practice: Development, validation and clinical simulation trial. Clin Pharmacol Ther. 2024;116(3):824-33. doi:10.1002/cpt.3293
13. Destere A, Marquet P, Gandonnière CS, Åsberg A, Loustaud-Ratti V, Carrier P, et al. A hybrid model associating population pharmacokinetics with machine learning: A case study with iohexol clearance estimation. Clin Pharmacokinet. 2022;61(8):1157-65. doi:10.1007/s40262-022-01138-x
14. Destere A, Marquet P, Labriffe M, Drici MD, Woillard JB. A hybrid algorithm combining population pharmacokinetic and machine learning for isavuconazole exposure prediction. Pharm Res. 2023;40(4):951-9. doi:10.1007/s11095-023-03507-y
15. 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. doi:10.1002/psp4.70084
16. Uster DW, Stocker SL, Carland JE, Brett J, Marriott DJE, Day RO, et al. A model averaging/selection approach improves the predictive performance of model-informed precision dosing: Vancomycin as a case study. Clin Pharmacol Ther. 2021;109(1):175-83. doi:10.1002/cpt.2065
17. Johnson M, Patel M, Phipps A, van der Schaar M, Boulton D, Gibbs M. The potential and pitfalls of artificial intelligence in clinical pharmacology. CPT Pharmacometrics Syst Pharmacol. 2023;12(3):279-84. doi:10.1002/psp4.12902
18. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7. doi:10.1136/bmjqs-2018-008370

19. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. BMJ. 2021;375. doi:10.1136/bmj.n2281
20. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating Diagnostic Tests and Prediction Models' of the STRATOS Initiative. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. doi:10.1186/s12916-019-1466-7
21. Ter Heine R, Keizer RJ, van Steeg K, Smolders EJ, van Luin M, Derijks HJ, et al. Prospective validation of a model-informed precision dosing tool for vancomycin in intensive care patients. Br J Clin Pharmacol. 2020;86(12):2497-506. doi:10.1111/bcp.14360
22. Broeker A, Nardecchia M, Klinker KP, Derendorf H, Day RO, Marriott DJ, et al. Towards precision dosing of vancomycin: A systematic evaluation of pharmacometric models for Bayesian forecasting. Clin Microbiol Infect. 2019;25(10):1286.e1-e7. doi:10.1016/j.cmi.2019.02.029
23. Groenland SL, van Eerden RAG, Westerdijk K, Meertens M, Koolen SLW, Moes DJAR, et al. Therapeutic drug monitoring-based precision dosing of oral targeted therapies in oncology: A prospective multicenter study. Ann Oncol. 2022;33(10):1071-82. doi:10.1016/j.annonc.2022.06.010
24. Kantasiripitak W, Outtier A, Wicha SG, Kensert A, Wang Z, Sabino J, et al. Multi-model averaging improves the performance of model-guided infliximab dosing in patients with inflammatory bowel diseases. CPT Pharmacometrics Syst Pharmacol. 2022;11(8):1045-59. doi:10.1002/psp4.12813
25. la Roi-Teeuw HM, van Royen FS, de Hond A, Zahra A, de Vries S, Bartels R, et al. Don't be misled: 3 misconceptions about external validation of clinical prediction models. J Clin Epidemiol. 2024;172:111387. doi:10.1016/j.jclinepi.2024.111387
26. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: A scoping review. NPJ Digit Med. 2022;5(1):2. doi:10.1038/s41746-021-00549-7
27. Marconi L, Cabitza F. Show and tell: A critical review on robustness and uncertainty for a more responsible medical AI. Int J Med Inform. 2025;202:105970. doi:10.1016/j.ijmedinf.2025.105970
28. Sokol K, Fackler J, Vogt JE. Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. NPJ Digit Med. 2025;8(1):345. doi:10.1038/s41746-025-01725-9
29. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-8. doi:10.1001/jama.2017.7797
30. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368. doi:10.1136/bmj.m689
31. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021;3(3):199-217. doi:10.1038/s42256-021-00307-0
32. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
33. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385. doi:10.1136/bmj-2023-078378
34. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro CL, Dhiman P, et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388. doi:10.1136/bmj-2024-082505
35. Psarellis YM, Pillai N, Dhakal S, Mavroudis PD. Machine-learning-enabled modeling of pharmacokinetics and pharmacodynamics. Drug Discov Today. 2026;31(3):104645. doi:10.1016/j.drudis.2026.104645