Pharmacophore

ISSN-2229-5402

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

COUNTERFACTUAL DOSING WITH CAUSAL DIGITAL TWINS UNDER PATIENT HETEROGENEITY, DISEASE EVOLUTION, AND TIME-VARYING TREATMENT RESPONSE

Grace Collins[1]*, Oliver Evans[2], Noah Williams[1], Ethan Brooks[3]

1. ***Department of Counterfactual Dosing and Causal Digital Twins, Faculty of Pharmacy, University of Sydney, Sydney, Australia.***
2. ***Department of Patient Heterogeneity and Disease Evolution, Faculty of Pharmacy, University of Queensland, Brisbane, Australia.***
3. ***Department of Time-Varying Treatment Response Modeling, Faculty of Pharmacy, University of Melbourne, Melbourne, Australia.***

ARTICLE INFO

Received:
15 April 2026
Received in revised form:
15 August 2026
Accepted:
16 August 2026
Available online:
28 August 2026

Keywords: Causal digital twin, Individualized dosing, Counterfactual treatment, Time-varying confounding, Pharmacometrics, Disease evolution

ABSTRACT

Individualized dosing is commonly framed as the selection of a dose that best predicts exposure, biomarker response, efficacy, or toxicity for a particular patient. That framing is incomplete when treatment decisions alter subsequent physiology, disease state, adherence, monitoring, and future treatment allocation. In such settings, the clinically relevant question is not merely what outcome is likely under the observed dose, but how the same patient might evolve under alternative feasible dose histories. This article proposes a causal dose twin: a conceptual treatment-simulation model that combines longitudinal causal structure, treatment-history representation, mechanistic pharmacokinetic–pharmacodynamic and disease models, sequential patient-state updating, and safety-constrained counterfactual comparison. The proposed architecture distinguishes prediction from causal estimation, observed response from treatment effect, prescribed dose from realized exposure, and model confidence from decision-relevant uncertainty. It treats patient heterogeneity as dynamic rather than fixed and represents disease progression, adherence, organ-function change, co-interventions, and monitoring as components that can alter both future response and future treatment. Counterfactual dose strategies are therefore evaluated as trajectories rather than isolated actions. The article further argues that validation must be matched to context of use and must include causal identification, state reconstruction, pharmacological adequacy, uncertainty calibration, transportability, safety behavior, and governance. The causal dose twin is presented as an original methodological synthesis, not as an empirically validated dosing system. Its usefulness is conditional on defensible causal assumptions, adequate longitudinal data, drug- and disease-specific mechanistic knowledge, clinically meaningful constraints, and accountable human authorization. The framework provides a research structure for moving individualized dosing beyond descriptive personalization toward bounded comparison of alternative treatment histories.

To Cite This Article: Collins G, Evans O, Williams N, Brooks E. Counterfactual Dosing with Causal Digital Twins under Patient Heterogeneity, Disease Evolution, and Time-Varying Treatment Response. Pharmacophore. 2026;17(4):83-92. https://doi.org/10.51847/XWR5pHgaUT

Introduction

Medical digital twins are increasingly described as computational representations that remain linked to an individual through repeated observations and can be used to simulate possible future states. Their scientific value, however, depends on more than personalization of input variables. A patient-specific model becomes relevant to dosing only when it can represent treatment history, update its internal state as new information arrives, and examine how alternative interventions could alter subsequent trajectories. A static profile, even when highly detailed, does not constitute a treatment simulator unless it has an explicit temporal structure, intervention semantics, and a defensible relationship between modeled actions and outcomes [1]. This distinction is particularly important in pharmaceutical science, where dose decisions influence exposure, response, toxicity, monitoring intensity, adherence, rescue treatment, organ function, and the next dose decision.

Corresponding Author: Grace Collins; Department of Counterfactual Dosing and Causal Digital Twins, Faculty of Pharmacy, University of Sydney, Sydney, Australia. E-mail: grace.collins@sydney.edu.au

Patient heterogeneity also cannot be reduced to a set of baseline covariates appended to an otherwise uniform model. Biological systems are multiscale and dynamically coupled: molecular disposition, receptor signaling, tissue adaptation, disease progression, comorbidity, behavior, and clinical intervention can evolve at different rates and interact nonlinearly. A patient who initially resembles a reference population may later depart from it because of resistance, tolerance, declining organ function, intercurrent illness, treatment interruption, or changing disease burden. A useful causal twin must therefore represent heterogeneity as an evolving state rather than a permanent label and must permit uncertainty about which mechanisms dominate at a particular time [2]. The central challenge is not simply to increase model complexity, but to preserve clinically interpretable relationships between state, treatment, time, and outcome.

This challenge exposes a fundamental difference between personalized prediction and individualized treatment-effect estimation. A model may predict that a patient receiving a high dose will experience a particular outcome, yet that association does not reveal what would have happened to the same patient under a lower dose, a delayed dose, a treatment interruption, or a monitoring-contingent strategy. Estimating individualized treatment effects requires an explicit contrast between alternative interventions and depends on assumptions about treatment assignment, confounding, overlap, outcome definition, and the time horizon of interest [3]. Consequently, high predictive accuracy for observed outcomes cannot by itself establish that a model can compare dose strategies causally or that its preferred strategy would improve a clinically relevant outcome.

This article develops an original causal treatment-simulation model termed the **causal dose twin**. The contribution is conceptual and architectural rather than empirical. It organizes individualized dosing around five linked requirements: a clearly specified counterfactual dose estimand; a longitudinal causal structure that retains treatment history and time-varying confounding; a pharmacological and disease-state model that can be updated as the patient changes; a constrained simulator for comparing clinically feasible dose trajectories; and a validation and governance structure proportional to the consequence of use. The model is not proposed as an autonomous prescriber, a regulator-accepted methodology, or a universally applicable solution. Its purpose is to define what would have to be represented, tested, and governed before a patient-specific simulation could support credible comparison of alternative dose histories.

Why Individualized Dosing Requires Counterfactual Rather Than Descriptive Models

Model-informed precision dosing has established that population pharmacokinetic, pharmacodynamic, disease, and covariate information can be combined with patient observations to support dose individualization. Yet the existence of a model capable of producing a patient-specific estimate does not establish routine clinical value. Translation depends on data quality, workflow integration, model qualification, interpretability, stakeholder acceptance, and evidence that the resulting decision improves upon existing practice [4]. More fundamentally, many dosing tools remain descriptive in the sense that they estimate current parameters, forecast outcomes under a selected regimen, or identify a dose expected to reach a predefined exposure target. These functions can be useful, but they do not necessarily answer the counterfactual question that underlies treatment choice: how would this patient's future differ under each clinically admissible dose strategy?

At the bedside, individualized dosing commonly combines a population model with current patient measurements to obtain a posterior estimate of exposure or response. This Bayesian structure is valuable because it allows prior pharmacological knowledge to be revised by observed concentrations, biomarkers, organ function, or treatment response [5]. However, posterior individualization and counterfactual identification are different operations. A posterior parameter estimate characterizes the patient within a model; it does not automatically establish the causal effect of changing treatment. If the measured state has already been altered by previous doses, clinical interventions, adherence, or selective monitoring, then the patient's apparent response cannot be interpreted independently of the process that generated both the observations and the treatment history.

Digital-health measurements may make the patient record more temporally detailed, but greater data volume does not remove this inferential distinction. Repeated wearable measurements, laboratory values, electronic administration records, home monitoring, and patient-reported information can enable more responsive dosing, yet each source introduces questions of timing, measurement validity, missingness, interoperability, and clinical relevance [6]. A descriptive model may absorb these signals and improve short-horizon prediction while still failing to distinguish whether a change in outcome is attributable to treatment, disease evolution, adherence, co-intervention, or observation intensity. Data availability is therefore not equivalent to data suitability for counterfactual simulation.

Heterogeneous treatment-effect methods reinforce the need to specify treatment contrasts rather than infer individual benefit from overall outcome prediction. These methods estimate how effects vary across patient characteristics, but their conclusions remain conditional on causal identification, adequate treatment overlap, consistent treatment definitions, and appropriate outcome models [7]. For dosing, the challenge is more demanding because the intervention is often a sequence rather than a binary treatment, patient characteristics change over time, and prior dose decisions can alter later covariates. The proposed causal dose twin consequently defines individualization as the comparison of alternative, clinically feasible dose histories for a specified patient state, outcome, and horizon. **Table 1** organizes the evidence, constructs, and interpretation boundaries needed to develop why individualized dosing requires counterfactual rather than descriptive models within the article's central argument.

Table 1. Evidence domains, core questions, scientific requirements, and interpretation boundaries for Why individualized dosing requires counterfactual rather than descriptive models

Construct or evidence domain	Core question	Scientific basis	Role in the article	Failure or overclaiming risk	Boundary statement
Patient-specific prediction	What outcome is expected under an observed or prespecified regimen?	Pharmacometric and predictive models can condition forecasts on individual covariates and observations [5].	Provides the descriptive foundation from which individualized dosing begins.	Treating accurate prediction as evidence of treatment benefit.	Prediction under one regimen does not identify the outcome under an unobserved alternative.
Counterfactual dose estimand	Which alternative dose histories are being compared, for whom, over what horizon, and for which outcome?	Individualized treatment-effect estimation requires explicit treatment contrasts [3].	Defines the causal quantity that the proposed twin must simulate.	Allowing the treatment question to shift after model fitting.	The estimand must be specified before interpreting model output.
Dynamic patient heterogeneity	Which patient characteristics modify response, and how do they change after treatment begins?	Treatment effects may vary across measured patient characteristics [7].	Converts heterogeneity from a baseline subgroup label into an evolving model component.	Assuming that baseline effect modifiers remain stable throughout treatment.	Observed heterogeneity does not prove the mechanism producing it.
Model-informed precision dosing	How should prior pharmacological knowledge and patient observations be combined?	Bayesian updating can revise individual exposure or response estimates [5].	Supplies the patient-specific parameterization layer.	Assuming that an individualized posterior is a causal treatment effect.	Parameter individualization is necessary in many applications but is not sufficient for counterfactual identification.
Digital longitudinal evidence	Which repeated observations are timely, valid, and causally interpretable?	Digital-health data can support repeated dosing updates but require validation and workflow integration [6].	Provides the evidence stream for updating patient state.	Equating more frequent measurement with greater clinical truth.	Each observation must be evaluated for measurement error, timing, missingness, and relevance.
Clinical translation	What evidence is required before model output may influence care?	Routine use depends on validation, implementation, infrastructure, and stakeholder factors [4].	Establishes that model construction and clinical authorization are separate stages.	Presenting technical feasibility as deployment readiness.	The proposed model does not establish clinical utility or regulatory acceptability.
Association versus intervention	Does an observed exposure–outcome relationship represent the effect of changing dose?	Treatment assignment, prior response, monitoring, and co-interventions can generate noncausal associations.	Motivates the causal-history and confounding structure developed in the next section.	Recommending a dose on the basis of associations produced by prior clinical decisions.	Association may support hypothesis formation but cannot alone authorize a causal dose comparison.
Safety-bounded individualization	Which dose alternatives are clinically admissible before optimization begins?	Individualized treatment must operate within contraindications, toxicity limits, monitoring requirements, and available action support.	Prevents the causal twin from treating every mathematically representable action as clinically eligible.	Allowing an optimization objective to override nonnegotiable constraints.	Safety constraints are independent requirements, not optional reward terms.

Causal Structure, Treatment History, and Time-Varying Confounding

The causal dose twin begins with a time-indexed representation of how treatment, patient state, observation, and outcome are related. This structure must distinguish variables measured before the current decision from variables already affected by earlier treatment. In longitudinal care, renal function, disease markers, toxicity, symptoms, adherence, and monitoring intensity may influence the next dose while also carrying the effects of previous doses. Such variables are simultaneously prognostic, treatment-determining, and treatment-affected. Conventional adjustment can therefore fail because conditioning on them may block part of the treatment effect, introduce selection bias, or otherwise distort the contrast of interest. G-methods were

developed for precisely this form of treatment-confounder feedback, although their validity still requires well-defined interventions, adequate treatment support, correct temporal ordering, and sufficiently measured confounding [8].

A causal simulator must also distinguish causal estimation from mechanistic trajectory generation. Mechanistic models encode hypotheses about how dose produces exposure, how exposure changes biological state, and how that state evolves over time. Causal methods encode the conditions under which an intervention contrast can be interpreted from observed data. These functions are complementary but not interchangeable. A mechanistically detailed model may simulate plausible trajectories while preserving bias from treatment allocation or omitted causes; a causal estimator may identify an intervention effect without representing the biological mechanism in sufficient detail for extrapolation to new dosing patterns. Comparisons between agent-based simulation and the parametric g-formula illustrate that distinct model classes can represent interventions and time evolution differently, even when they address related longitudinal questions [9]. The proposed architecture therefore maintains separate but communicating causal-identification and pharmacological-simulation layers.

Treatment history must be represented as more than a list of prescribed doses. A clinically meaningful ledger should retain intended dose, prescribed dose, administration time, infusion or formulation details, interruptions, reductions, missed administrations, rescue treatment, monitoring-triggered changes, and uncertainty about actual intake. This ledger defines the intervention path that generated the current patient state and constrains the alternative strategies that can be compared. The parametric g-formula demonstrates how sustained or dynamic treatment strategies can be defined explicitly and used to simulate longitudinal outcomes under specified intervention rules [10]. In a causal dose twin, this logic is extended conceptually to pharmacological trajectories: each candidate strategy must specify how future doses respond to evolving measurements rather than merely naming an isolated dose amount.

Longitudinal dose comparison becomes more complex when outcomes include survival, competing events, treatment discontinuation, rescue therapy, or informative censoring. Different implementations of the parametric g-formula can rely on different models for outcome processes, hazards, and time-varying covariates, and no implementation eliminates sensitivity to model specification [11]. The causal dose twin must therefore expose rather than conceal these choices. It should identify the time origin, decision intervals, intervention versions, outcome horizon, competing events, and conditions under which simulation stops or abstains. It must also report when treatment overlap is weak, when a candidate strategy extends beyond observed clinical support, and when unmeasured confounding makes the comparison nonidentifiable. Causal structure is thus not an ornamental diagram added to a predictive model; it is the organizing constraint that determines which histories can be interpreted, which counterfactuals can be simulated, and which conclusions must remain withheld.

Architecture of the Causal Dose Twin

The proposed causal dose twin separates patient-state estimation from treatment-policy comparison. Its state layer represents uncertain current pharmacokinetic, pharmacodynamic, disease, toxicity, organ-function, and behavioral conditions, whereas its intervention layer compares alternative future dose histories. Bayesian data assimilation can revise latent patient states as observations arrive, while a separate policy component evaluates candidate actions against modeled future consequences [12]. This separation prevents a state estimate from being mistaken for a treatment recommendation and allows the twin to update what is believed about the patient without automatically changing what actions are clinically authorized.

A single pharmacometric structure may not remain adequate across all patients, disease phases, or data densities. The architecture therefore contains a model bank rather than one privileged model, with local applicability weights indicating which structural assumptions are most defensible for the present state. Model-combination methods can represent variation in local model suitability and expose disagreement that would be concealed by selecting one global model [13]. The resulting ensemble is not assumed to remove bias: models may share omitted mechanisms, causal errors, or unsuitable parameterizations. Material disagreement should instead widen uncertainty, trigger sensitivity analysis, or cause the system to abstain.

Counterfactual simulation occurs by creating parallel computational copies of the same updated patient state and exposing each copy to a different clinically feasible dose history. This concept is consistent with proposals to use individualized digital-twin copies for treatment simulation [14]. Each branch must preserve the same predecision evidence while differing only according to its defined intervention rule. Candidate branches may represent continuation, escalation, de-escalation, interval modification, interruption, or monitoring-contingent adaptation. Their outputs are comparative trajectories, not factual predictions of what will necessarily occur, and they remain conditional on the causal, pharmacological, and measurement assumptions embedded in the twin.

The architecture also requires an evidence-return pathway because internal simulation cannot validate itself. Experimental, clinical, and longitudinal observations may refine computational assumptions, reveal structural failure, or invalidate previously credible parameterizations; proposed links between digital twins and empirical systems emphasize this iterative relationship [15]. A provenance ledger should therefore retain data lineage, treatment definitions, model versions, update history, validation status, and context of use. **Table 2** organizes the evidence, constructs, and interpretation boundaries needed to develop architecture of the causal dose twin within the article's central argument. **Figure 1** presents the causal dose-twin counterfactual simulator, showing how the article's key components and boundaries are connected within architecture of the causal dose twin.

Table 2. Components, relationships, uncertainties, and validation needs within Architecture of the causal dose twin

Component or layer	Inputs	Core function	Expected output	Uncertainty or limitation	Validation requirement
Counterfactual estimand layer	Target population, decision time, feasible strategies, outcome, horizon	Defines the treatment contrast to be simulated	Explicit comparison question	Ambiguous intervention versions or shifting outcomes	Independent review of estimand and clinical relevance
Longitudinal causal structure	Treatment history, covariates, monitoring, adherence, outcomes	Represents temporal ordering and treatment-confounder feedback	Admissible causal pathways and adjustment requirements	Unmeasured causes and incorrect temporal assumptions	Causal-diagram review, sensitivity analysis, and identification audit
Patient-state representation	Concentrations, biomarkers, disease status, organ function, symptoms	Maintains a probabilistic estimate of the current patient	Posterior distribution over clinically relevant states	Partial observability and measurement error	Posterior predictive checks and state-recovery assessment
Treatment-history ledger	Intended, prescribed, administered, interrupted, and inferred doses	Preserves path dependence and realized exposure uncertainty	Time-indexed treatment record	Missing administrations and inaccurate timestamps	Reconciliation against multiple clinical evidence streams
Mechanistic PK–PD–disease core	Drug properties, patient parameters, exposure, response, toxicity	Simulates pharmacological and disease-state transitions	Future exposure, response, toxicity, and disease trajectories	Structural misspecification and parameter nonidentifiability	External validation at exposure, biomarker, and outcome levels
Model bank and applicability layer	Alternative structural models, patient similarity, local evidence	Represents structural uncertainty and local relevance	Model-weighted or competing simulations	Shared omissions and unstable model weights	External comparison and disagreement analysis
Observation assimilator	New measurements, adherence evidence, interventions, timestamps	Updates patient and population knowledge sequentially	Revised state and uncertainty	Informative missingness and delayed observations	Prequential evaluation and controlled update testing
Candidate-strategy generator	Clinical action set, monitoring rules, treatment objectives	Constructs feasible dose histories	Defined fixed or adaptive intervention branches	Unsupported or clinically unrealistic actions	Action-space review by clinicians and pharmacologists
Safety-constraint shield	Contraindications, toxicity limits, monitoring requirements, hard rules	Excludes or truncates unacceptable strategies	Safety-filtered candidate set	Unknown interactions and incomplete rules	Edge-case testing, red-team scenarios, and failure injection
Uncertainty and applicability monitor	Parameter, model, causal, measurement, and transport uncertainty	Determines whether comparison is interpretable	Uncertainty decomposition, warnings, or abstention	Unrecognized model failure	Calibration, shift testing, and coverage assessment
Credibility and provenance ledger	Data lineage, model version, assumptions, validation records	Makes simulations reconstructable and reviewable	Auditable evidence record	Incomplete documentation or uncontrolled updates	Independent reproducibility and governance audit
Human authorization interface	Comparative trajectories, uncertainty, constraints, disagreement	Supports accountable interpretation and final decision	Bounded scenarios with reasons to accept, reject, or abstain	Automation bias and cognitive overload	Human-factors studies and prospective workflow evaluation

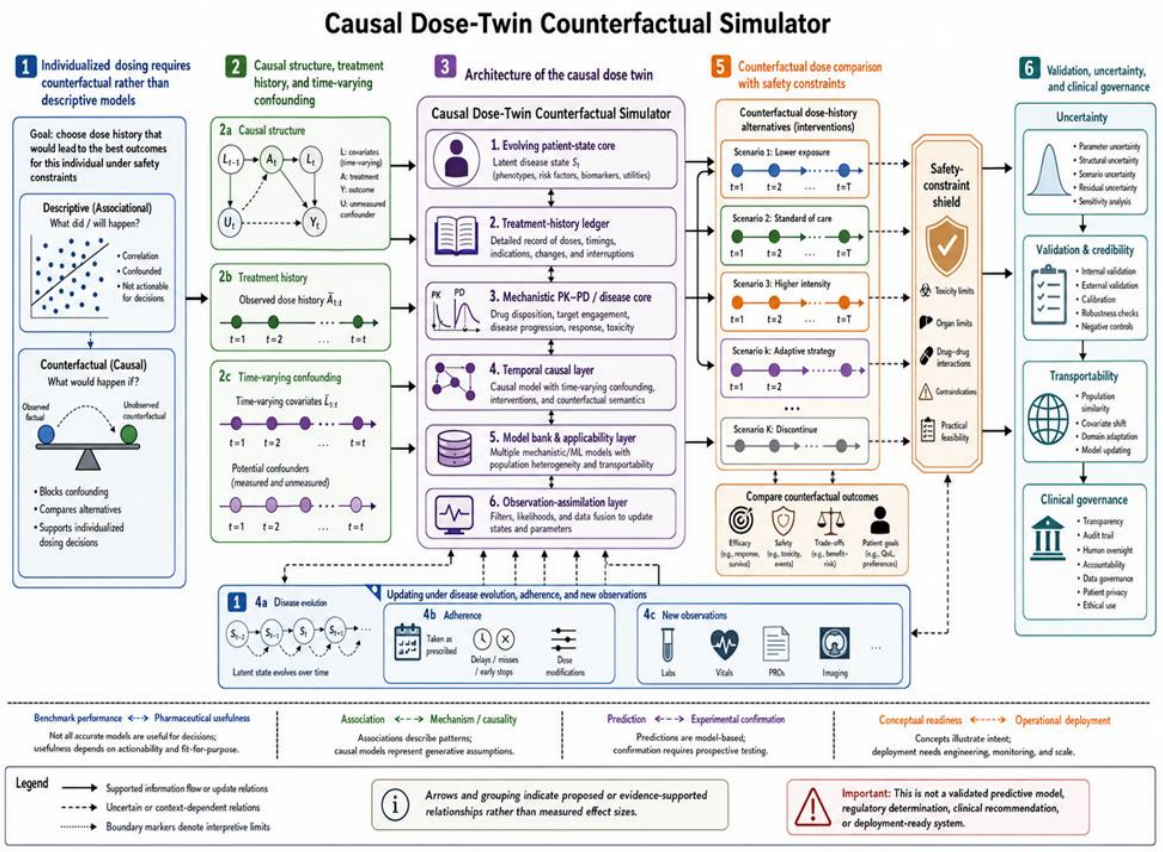

Figure 1. Causal Dose-Twin Counterfactual Simulator.

The figure is an original conceptual synthesis that organizes the article's central contribution across individualized dosing requires counterfactual rather than descriptive models, causal structure, treatment history, and time-varying confounding, causal structure, treatment history, time-varying confounding, architecture of the causal dose twin. Arrows and grouping indicate proposed or evidence-supported relationships rather than measured effect sizes. The figure does not represent a validated predictive model, regulatory determination, clinical recommendation, or deployment-ready system.

Updating Under Disease Evolution, Adherence, and New Observations

A causal dose twin must treat each observation according to what it can legitimately update. In antimicrobial precision dosing, serial concentrations, organ-function measurements, susceptibility information, biomarkers, and clinical context may materially revise exposure and response estimates [16]. A new concentration can inform clearance or distribution assumptions, but it should not automatically alter the disease model, causal graph, or adherence estimate unless the observation contains evidence relevant to those components. Updating should therefore be modular, preserving the distinction between measured evidence, inferred latent state, and structural assumptions.

Patient-level and population-level learning should also occur on different timescales. Continued-learning approaches to model-informed dosing distinguish within-patient updating from revision of population priors across accumulated cases [17]. The proposed twin adopts this distinction to avoid allowing one unusual observation to redefine a population model or allowing a slowly changing population prior to suppress clinically important patient-specific change. Updates to shared models should be version controlled, independently evaluated, and prevented from influencing active decisions until their validity has been reassessed for the intended context.

Disease evolution introduces nonstationarity into treatment response. Progression, remission, acquired resistance, tolerance, immune adaptation, cumulative toxicity, and changing baseline risk can alter the relationship between dose and outcome even when administration remains unchanged. Longitudinal disease-trajectory models may help characterize these changes and reveal heterogeneous response patterns [18]. They do not, however, identify causal treatment effects automatically. Within the twin, disease evolution should be represented as an uncertain transition process that can be challenged by incoming evidence and compared with alternative explanations such as measurement error, altered exposure, co-intervention, or adherence change.

Adherence complicates updating because prescribed, dispensed, administered, and ingested treatment may diverge. Evidence from adherence research shows that imperfect medication use can materially influence apparent response and remains difficult to measure accurately [19]. The twin should therefore maintain a probability distribution over plausible exposure histories rather than converting a weak adherence proxy into a definitive behavioral classification. When outcome deterioration can be explained by several competing histories, simulations should propagate those alternatives separately. Escalation should be

withheld when the system cannot distinguish inadequate pharmacological effect from uncertain treatment realization, unless independent clinical considerations justify action.

Counterfactual Dose Comparison with Safety Constraints

Counterfactual dose comparison should begin only after the candidate action space has been clinically bounded. Healthcare reinforcement-learning guidance identifies confounding, reward misspecification, inadequate treatment support, unsafe exploration, and unreliable off-policy evaluation as central threats [20]. These risks are especially consequential in dosing because an apparently favorable strategy may exploit documentation artifacts, selective monitoring, or actions rarely used in comparable patients. The proposed twin therefore excludes unsupported actions, reports weak overlap, and treats retrospective policy value as provisional rather than as evidence of clinical benefit.

Model-informed reinforcement learning can organize adaptive dosing by combining pharmacological structure with sequential action selection [21]. Within the proposed architecture, such methods generate candidate strategies but do not receive independent prescribing authority. The objective must represent multiple clinically relevant consequences, including efficacy, toxicity, treatment burden, monitoring requirements, and delayed effects. Nonnegotiable contraindications and safety limits should be implemented as an independent shield rather than absorbed entirely into an optimization reward that may be misspecified or traded against another objective.

Theoretical model-enhanced dosing studies illustrate how pharmacological knowledge can support policy learning when observations are limited [22]. Their value lies in testing architecture, sensitivity, and failure modes under explicit assumptions, not in establishing bedside effectiveness. A strategy that performs favorably in a simulated environment may inherit the environment's structural simplifications, omit rare harms, or exploit unrealistic state information. The causal dose twin should therefore compare results across plausible models, perturb uncertain assumptions, and downgrade conclusions when rankings change materially across credible alternatives.

Drug-specific adaptive-dosing simulations further show that policy behavior is conditional on the modeled compound, disease, endpoint, action space, and safety structure [23]. This context dependence precludes a universal dose-policy layer. Each application requires its own feasible interventions, monitoring schedule, outcome hierarchy, stopping rules, and contraindications. The final output should not be a single optimized dose, but a bounded comparison showing expected trajectories, uncertainty, model disagreement, safety exclusions, and reasons for abstention. Human reviewers must be able to determine whether the comparison addresses the actual clinical decision and whether omitted considerations invalidate its use.

Validation, Uncertainty, and Clinical Governance

Validation must be proportional to the consequence of the proposed use. Credibility frameworks for model-informed drug development link model risk and context of use to the required depth of verification and validation [24]. A twin used for hypothesis generation may require less evidence than one influencing a bedside decision, but neither should be described simply as "validated" without specifying the target population, intervention class, endpoint, time horizon, and permitted influence. Validation evidence should be modular so that changes to data, model structure, or action rules trigger reassessment of the affected claims.

Digital twins require lifecycle verification, validation, and uncertainty quantification because their internal state and supporting evidence can change after initial construction [25]. Evaluation should include software verification, causal-assumption review, pharmacological adequacy, state reconstruction, counterfactual discrimination, calibration, external transportability, safety behavior, and update stability. Uncertainty should be decomposed into parameter, structural, measurement, causal-identification, and transport components. A narrow predictive interval cannot compensate for an omitted mechanism or an unidentified intervention contrast.

Clinical governance extends beyond technical performance. Ethical analysis of health digital twins identifies concerns involving consent, privacy, transparency, autonomy, data use, and responsibility across the information lifecycle [26]. The proposed governance layer therefore requires data provenance, access control, model-version records, update authorization, incident reporting, challenge pathways, and accountable human ownership. Patients and clinicians should be informed when outputs depend on uncertain adherence inference, incomplete histories, external data transfer, or models not validated for the local population.

Generalisability is not an intrinsic property that transfers automatically from one cohort or institution to another. Clinical models may behave differently across populations, workflows, measurement systems, treatment policies, and disease prevalence [27]. The causal dose twin must consequently evaluate transportability for each intended setting and monitor subgroup-specific failure after implementation. Aggregate performance should not conceal poor validity in underrepresented groups. **Table 3** organizes the evidence, constructs, and interpretation boundaries needed to develop validation, uncertainty, and clinical governance within the article's central argument.

Table 3. Evaluation, implementation, and research priorities arising from Counterfactual Dosing with Causal Digital Twins under Patient Heterogeneity, Disease Evolution, and Time-Varying Treatment Response

Evaluation dimension	What must be tested	Suitable evidence or assessment	Failure signal	Interpretive limitation	Decision relevance

Counterfactual estimand validity	Whether the simulated contrast matches the intended clinical question	Protocol-level specification and independent estimand review	Ambiguous strategies, time zero, outcome, or horizon	A valid estimate can still answer the wrong question	Determines whether simulation is clinically interpretable
Longitudinal causal identification	Whether treatment-confounder feedback and selection processes are addressed	Causal diagrams, positivity assessment, sensitivity analyses, negative controls	Strong dependence on unsupported assumptions	No method removes unmeasured confounding automatically	Governs whether treatment contrasts may be interpreted causally
Pharmacological adequacy	Whether dose, exposure, response, toxicity, and disease evolution are represented credibly	External PK, PD, biomarker, and outcome checks	Good outcome fit with implausible internal behavior	Predictive agreement does not prove mechanism	Supports or limits extrapolation to alternative dosing
State reconstruction	Whether the latent patient state tracks clinically meaningful change	Posterior predictive checks and held-out longitudinal observations	Persistent residual patterns or delayed response to change	Observation prediction does not prove treatment-effect validity	Determines whether counterfactual branches start from a credible state
Sequential updating	Whether updates improve inference without destabilizing the system	Prequential scoring, drift tests, version comparisons	Catastrophic drift, forgetting, or failure to respond	More frequent updating is not necessarily better	Controls whether new observations may alter active simulations
Counterfactual discrimination	Whether strategy rankings correspond to genuine differences	Randomized evidence, target-trial emulation, perturbation studies	Unstable rankings across reasonable specifications	Retrospective agreement does not establish clinical utility	Determines whether alternatives can be meaningfully distinguished
Safety-constraint behavior	Whether prohibited or unsupported actions are blocked	Edge-case tests, red-team scenarios, failure injection	Constraint violations or unsafe extrapolation	Scripted cases cannot cover every interaction	Sets the maximum permissible influence of the model
Uncertainty calibration	Whether reported uncertainty reflects observed error	External and temporally shifted validation	Overconfidence under shift or model disagreement	Predictive calibration is not causal certainty	Supports abstention and communication of decision risk
Transportability and subgroup validity	Whether conclusions persist across settings and patient groups	Multi-site external validation and subgroup stress testing	Aggregate success with subgroup failure	Statistical parity does not prove equitable benefit	Limits where the model may be used
Human interpretability	Whether users understand assumptions, uncertainty, and abstention	Human-factors studies and realistic case simulations	Automation bias or inability to identify failure	User satisfaction does not establish faithfulness	Determines whether human oversight is meaningful
Governance and provenance	Whether every simulation and update can be reconstructed	Version, data-lineage, authorization, and incident audits	Undocumented model or data changes	Auditability alone does not establish safety	Enables accountability, rollback, and independent review
Clinical utility	Whether use improves decisions or patient-relevant outcomes	Staged prospective comparative evaluation	Technical accuracy without improved care	Process improvement may not equal patient benefit	Required before consequential clinical reliance

Limitations and Boundary Conditions

The proposed causal dose twin is a theoretical synthesis assembled from causal inference, pharmacometrics, digital-twin modeling, sequential learning, and governance scholarship. The selected literature supports individual components, but it does not establish that their integration will produce valid individualized treatment effects. Important variables may remain unmeasured, pharmacological structures may be incomplete, treatment versions may be inconsistent, and longitudinal action support may be insufficient. Under these conditions, the system must communicate nonidentifiability or abstain rather than substitute mechanistic plausibility for causal evidence [8].

The framework is also context dependent. Drug classes differ in therapeutic windows, reversibility, delay between exposure and outcome, measurement availability, adherence observability, and consequences of error. A design appropriate for antibiotic concentration targeting may be unsuitable for oncology, immunotherapy, psychiatric treatment, or chronic disease

management. Model combination, adaptive policies, or uncertainty quantification cannot create information absent from the evidence base. External validation in one institution or subgroup does not establish transportability to another, particularly when clinical workflows and treatment policies differ [27].

The architecture cannot establish regulatory acceptance, clinical effectiveness, cost-effectiveness, or operational readiness. It does not replace randomized evidence when such evidence is feasible, nor does it remove the need for pharmacovigilance, patient preference, clinician judgment, and multidisciplinary review. Governance documentation can make decisions auditable, but auditability is not equivalent to safety. Similarly, a constrained interface can reduce misuse without eliminating automation bias. Any clinical influence would require staged prospective evaluation tied to a narrowly defined context of use and a predefined rollback pathway [24].

Research Agenda and Implementation Priorities

The first research priority is methodological integration. Future work should define dynamic dose estimands that align causal inference with PK–PD and disease models, including explicit treatment versions, decision intervals, adherence states, rescue interventions, competing events, and stopping rules. Comparative studies should examine when mechanistic simulation and longitudinal g-methods agree, when they diverge, and which assumptions determine the difference. Methods are also needed to propagate causal-identification uncertainty alongside parameter and structural uncertainty rather than reporting only conditional simulation intervals.

The second priority is staged evaluation. Development should progress from software verification and simulation-based state recovery to retrospective external validation, silent prospective evaluation, and controlled clinical testing. Update mechanisms should be assessed prequentially, with predefined criteria for drift, rollback, and revalidation. Safety shields require adversarial testing against unsupported actions, delayed toxicity, missing data, conflicting measurements, and model disagreement. Human-factors studies should assess whether clinicians recognize uncertainty, challenge inappropriate outputs, and use abstention information appropriately [25].

The third priority is infrastructure and governance. Implementation requires interoperable longitudinal records, timestamped administration data, model registries, version-controlled causal assumptions, secure provenance, and multidisciplinary oversight. Reporting standards should distinguish descriptive prediction, causal estimation, mechanistic simulation, policy evaluation, and clinical utility. Each output should identify the intended context, admissible action space, evidence support, uncertainty sources, and reasons for withholding a comparison. Such infrastructure would not make the causal dose twin deployment ready, but it would make future empirical evaluation more reproducible, transparent, and accountable [26].

Conclusion

The causal dose twin reframes individualized dosing as the bounded comparison of alternative treatment histories for an evolving patient rather than the optimization of a single predictive score. Its proposed architecture connects an explicit counterfactual estimand with longitudinal causal structure, treatment-history memory, pharmacological and disease-state simulation, adherence-aware updating, safety constraints, decomposed uncertainty, provenance, and human authorization. The contribution is not a validated dosing technology but a specification of the scientific relationships that credible treatment simulation would need to preserve. Its central boundary is equally important: neither personalization, mechanistic detail, predictive accuracy, nor internal optimization establishes an individual causal benefit. Progress therefore depends on drug-specific causal definitions, suitable longitudinal evidence, prospective validation, transparent uncertainty, and governance proportional to clinical consequence.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Laubenbacher R, Mehrad B, Shmulevich I, Trayanova N. Digital twins in medicine. Nat Comput Sci. 2024;4(3):184-91. doi:10.1038/s43588-024-00607-6
2. De Domenico M, Allegri L, Caldarelli G, d'Andrea V, Di Camillo B, Rocha LM, et al. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. NPJ Digit Med. 2025;8(1):37. doi:10.1038/s41746-024-01402-3
3. Curth A, Peck RW, McKinney E, Weatherall J, van der Schaar M. Using machine learning to individualize treatment effect estimation: Challenges and opportunities. Clin Pharmacol Ther. 2024;115(4):710-9. doi:10.1002/cpt.3159

4. Darwich AS, Ogungbenro K, Vinks AA, Powell JR, Reny JL, Marsousi N, et al. Why has model-informed precision dosing not yet become common clinical reality? Lessons from the past and a roadmap for the future. Clin Pharmacol Ther. 2017;101(5):646-56. doi:10.1002/cpt.659
5. Keizer RJ, ter Heine R, Frymoyer A, Lesko LJ, Mangat R, Goswami S. Model-informed precision dosing at the bedside: Scientific challenges and opportunities. CPT Pharmacometrics Syst Pharmacol. 2018;7(12):785-7. doi:10.1002/psp4.12353
6. Kluwe F, Michelet R, Mueller-Schoell A, Maier C, Klopp-Schulze L, van Dyk M, et al. Perspectives on model-informed precision dosing in the digital health era: Challenges, opportunities, and recommendations. Clin Pharmacol Ther. 2021;109(1):29-36. doi:10.1002/cpt.2049
7. Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci U S A. 2019;116(10):4156-65. doi:10.1073/pnas.1804597116
8. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol. 2017;46(2):756-62. doi:10.1093/ije/dyw323
9. Murray EJ, Robins JM, Seage GR III, Freedberg KA, Hernán MA. A comparison of agent-based models and the parametric g-formula for causal inference. Am J Epidemiol. 2017;186(2):131-42. doi:10.1093/aje/kwx091
10. McGrath S, Lin V, Zhang Z, Petito LC, Logan RW, Hernán MA, et al. gfoRmula: An R package for estimating the effects of sustained treatment strategies via the parametric g-formula. Patterns (N Y). 2020;1(3):100008. doi:10.1016/j.patter.2020.100008
11. Wen L, Young JG, Robins JM, Hernán MA. Parametric g-formula implementations for causal survival analyses. Biometrics. 2021;77(2):740-53. doi:10.1111/biom.13321
12. Maier C, Hartung N, Kloft C, Huisinga W, de Wiljes J. Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. CPT Pharmacometrics Syst Pharmacol. 2021;10(3):241-54. doi:10.1002/psp4.12588
13. Chan A, Peck R, Gibbs M, van der Schaar M. Synthetic Model Combination: A new machine-learning method for pharmacometric model ensembling. CPT Pharmacometrics Syst Pharmacol. 2023;12(7):953-62. doi:10.1002/psp4.12965
14. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, et al. Digital twins to personalize medicine. Genome Med. 2020;12(1):4. doi:10.1186/s13073-019-0701-3
15. Akbarialiabad H, Seyyedi MS, Paydar S, Habibzadeh A, Haghighi A, Kvedar JC. Bridging silicon and carbon worlds with digital twins and on-chip systems in drug discovery. NPJ Syst Biol Appl. 2024;10(1):150. doi:10.1038/s41540-024-00476-9
16. 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
17. Maier C, de Wiljes J, Hartung N, Kloft C, Huisinga W. A continued learning approach for model-informed precision dosing: Updating models in clinical practice. CPT Pharmacometrics Syst Pharmacol. 2022;11(2):185-98. doi:10.1002/psp4.12745
18. Terranova N, Venkatakrishnan K. Machine learning in modeling disease trajectory and treatment outcomes: An emerging enabler for model-informed precision medicine. Clin Pharmacol Ther. 2024;115(4):720-6. doi:10.1002/cpt.3153
19. Burnier M, Egan BM. Adherence in hypertension. Circ Res. 2019;124(7):1124-40. doi:10.1161/CIRCRESAHA.118.313220
20. Gottesman O, Johansson F, Komorowski M, Faisal AA, Sontag D, Doshi-Velez F, et al. Guidelines for reinforcement learning in healthcare. Nat Med. 2019;25(1):16-8. doi:10.1038/s41591-018-0310-5
21. Tosca EM, De Carlo A, Ronchi D, Magni P. Model-informed reinforcement learning for enabling precision dosing via adaptive dosing. Clin Pharmacol Ther. 2024;116(3):619-36. doi:10.1002/cpt.3356
22. Ribba B, Bräm DS, Baverel PG, Peck RW. Model enhanced reinforcement learning to enable precision dosing: A theoretical case study with dosing of propofol. CPT Pharmacometrics Syst Pharmacol. 2022;11(11):1497-510. doi:10.1002/psp4.12858
23. De Carlo A, Tosca EM, Fantozzi M, Magni P. Reinforcement learning and PK-PD models integration to personalize the adaptive dosing protocol of erdafitinib in patients with metastatic urothelial carcinoma. Clin Pharmacol Ther. 2024;115(4):825-38. doi:10.1002/cpt.3176
24. Kuemmel C, Yang Y, Zhang X, Florian J, Zhu H, Tegenge M, et al. Consideration of a credibility assessment framework in model-informed drug development: Potential application to physiologically-based pharmacokinetic modeling and simulation. CPT Pharmacometrics Syst Pharmacol. 2020;9(1):21-8. doi:10.1002/psp4.12479
25. Sel K, Hawkins-Daarud A, Chaudhuri A, Osman D, Bahai A, Paydarfar D, et al. Survey and perspective on verification, validation, and uncertainty quantification of digital twins for precision medicine. NPJ Digit Med. 2025;8(1):40. doi:10.1038/s41746-025-01447-y
26. Huang PH, Kim KH, Schermer M. Ethical issues of digital twins for personalized health care service: Preliminary mapping study. J Med Internet Res. 2022;24(1). doi:10.2196/33081
27. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health. 2020;2(9). doi:10.1016/S2589-7500(20)30186-2