Calibrated within-population patient individualisation of antimicrobial exposure: a pre-registered simulation and external-evaluation study

Youssef Ghaly

yg@gatmedi.org

GATMEDI https://orcid.org/0009-0002-9554-3691

Research Article

Keywords: Model-informed precision dosing (MIPD), Population pharmacokinetics, Antimicrobial dosing, Individualised dosing, Monte Carlo simulation, Probability of target attainment, Pharmacometrics

Posted Date: July 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10299120/v1

Additional Declarations: The authors declare potential competing interests as follows: The author is the founder of GATMEDI, the company developing the individualisation engine evaluated in this study.

Calibrated within-population patient individualisation of antimicrobial exposure: a pre-registered simulation and external-evaluation study

Youssef Ghaly
GATMEDI
Correspondence: yg@gatmedi.org

Preprint — version 1 July 2026.

Abstract

Model-informed precision dosing updates a population pharmacokinetic prior with a patient's own measured drug concentrations to individualise the dosing of narrow-therapeutic-index drugs. The approach is established and available in several commercial systems; what must be demonstrated for any particular implementation is that the individual estimates it returns are unbiased, adequately precise, and well calibrated, the last being the property on which safe dose selection depends. We report a pre-registered validation of within-population individualisation in an engine whose population priors are drawn from a compiled medical-knowledge substrate. The protocol, pass/fail criteria, sample sizes, and analysis were fixed and cryptographically committed before any result was generated, and the engine was content-hashed and produced deterministic output. Individualisation was evaluated within population in two complementary ways: simulation-based recovery of individual clearance for eight published priors spanning adult and paediatric critical-care populations across three antimicrobials, with 500 simulated patients per prior; and external evaluation in which data were simulated from one adult vancomycin model and forecast using a different adult vancomycin prior, comparing the population-typical forecast with the individualised forecast on held-out concentrations. Posterior credible intervals were calibrated for all eight priors at the 50, 80, 90, and 95% levels. Recovery was near-unbiased for seven of eight priors and met a stringent imprecision criterion for five; the exceedances were pre-specified consequences of the estimator and the sparse-sampling design. Individualisation reduced external prediction error by 32 to 52% without increasing bias. Within-population individualisation was calibrated and improved prediction at the scale tested.

1. Introduction

The dosing of narrow-therapeutic-index antimicrobials is difficult precisely because the between-patient variability in drug handling is large relative to the exposure window that separates efficacy from toxicity. For vancomycin and the beta-lactams in critically ill patients, clearance can differ several-fold between individuals of the same weight and renal function, driven by augmented renal clearance, organ support, fluid shifts, and changes in protein binding. A fixed, label-based dose therefore under-exposes some patients, with risk of treatment failure and the selection of resistance, while over-exposing others, with risk of toxicity. Exposure targets for these agents are well established (an AUC_{24}/MIC of 400 to 600 for vancomycin; a defined fraction of the dosing interval with free concentration above the MIC for beta-lactams), but attaining them reliably requires the dose to be adapted to the individual patient rather than to the population average.

The clinical consequences are measurable. In the DALI study — a prospective point-prevalence study of eight beta-lactams in 361 evaluable critically ill patients across 68 hospitals — 16% of those treated for infection did not attain even the most conservative pharmacodynamic target (free drug above the MIC for 50% of the dosing interval, 50% fT>MIC), and attainment fell at stricter targets: among the principal intensive-care agents, 67% of patients on piperacillin and 70% on meropenem reached 100% fT>MIC, and fewer than half of all patients reached the preferred 100% fT>4×MIC target. In a cost analysis of the EXPAT study, 50 of 79 critically ill patients (63%) attained the 100% fT>MIC target. A substantial minority of patients therefore fail to reach beta-lactam exposure targets that fixed, label-based dosing does not resolve. Individualisation is the established remedy, so the open question for any implementation is not whether to individualise but whether the individual estimates it returns can be trusted. This is where medical AI most often fails: contemporary systems are frequently overconfident in proportion to, not in spite of, their errors — returning a wrong answer with the same assurance as a right one. A dosing engine that cannot represent its own uncertainty honestly is, in that case, not safer than fixed dosing.

That remedy is model-informed precision dosing, also termed Bayesian forecasting. A population pharmacokinetic model provides a prior over the patient's pharmacokinetic parameters; the patient's own measured concentration is then used to update that prior to a patient-specific posterior, from which the next dose is chosen. The method is mature, is implemented in several commercial dose-optimisation systems, and is not claimed as novel here. The object of validation in this study is whether a particular engine performs within-population individualisation correctly, transparently, and with calibrated uncertainty, using priors drawn from a compiled medical-knowledge substrate. The differentiator under development is the substrate from which the priors and the surrounding clinical facts are obtained, not the dosing arithmetic, which is deliberately standard so that it can be checked against the public literature. The validation here uses antimicrobials, but the object under test is the individualisation method rather than any single drug: the engine and its uncertainty and abstention behaviour are drug-agnostic, and antimicrobials provide the first instance with published population models and a directly measurable target (Section 6).

Calibration is the safety-critical property. An individualised point estimate is of limited use unless the uncertainty reported around it can be relied upon at its stated rate; a confidently wrong dose is more dangerous than an acknowledged inability to estimate. The primary endpoint of this study is therefore the empirical coverage of the engine's posterior credible intervals, evaluated alongside the bias and precision of the individual estimates and the improvement in prediction that the measurement update provides. The scope of the study is within-population individualisation: the patient is drawn from, or matched to, the population in which the prior was developed. Generalisation across populations is a distinct validation question with a different design and is not addressed here; it is identified in the Limitations as future work.

The remainder of the paper sets out the pre-registration and reproducibility discipline, the population-prior battery, the metrics and their pre-specified bars, and the within-population results, including the controls, the demonstration patients, and the points at which the engine missed a pre-registered bar.

2. Methods

2.1 Scope of disclosure

The population pharmacokinetic priors used in this study are published models and are cited so that the validation is independently checkable. The method by which the substrate is compiled, and by which priors, exposure targets, and clinical facts are assembled and linked within it, is proprietary and is out of scope for this paper. Nothing in the sections that follow is required to reproduce that process, and the descriptions of the individualisation procedure are confined to standard, publicly documented pharmacometric methods.

2.2 Role of artificial intelligence

It is worth stating precisely where artificial intelligence does and does not act, because the two are easily conflated. The individualisation arithmetic — the population model, the Bayesian update, the Monte-Carlo propagation to an exposure metric — is deliberately classical, the same mathematics the cited literature uses, implemented so that every step can be checked against published worked examples, and it is. AI enters in two places, neither of which is the arithmetic. First, the population priors, exposure targets, and surrounding clinical facts are drawn from a compiled medical-knowledge substrate, the construction of which is proprietary and out of scope (Section 2.1). Second, the engine's runtime behaviour — returning a calibrated posterior with an explicit credible interval, and declining or flagging when inputs do not support an estimate — is evaluated directly here (Sections 3.1, 3.4, 3.6). The part that can be checked against the literature is not the AI; the AI contribution is the provenance of what goes in and the calibration and honesty of what comes out.

2.3 Pre-registration and reproducibility

The validation protocol, every numeric pass/fail bar, the sample sizes, the fixed denominators, the negative controls, and the analysis were written down and cryptographically committed before any result was generated. The commitment included the expected failure modes: in particular, a shrinkage-induced bias exceedance for the highest-variability prior under a sparse two-sample design was predicted in advance rather than explained after the fact. The engine was frozen and content-hashed prior to the run, and its deterministic outputs reproduce byte-for-byte at a fixed random seed. No bar was moved, no prior was dropped, no run was repeated to obtain a more favourable result, and no exceedance was reframed after results were seen; the decision to bank the run was taken after review under a pre-specified clause requiring every result to be reported as observed. Provenance identifiers (commit, snapshot, content hash, seed) are listed in the Appendix.

2.4 Validation battery and populations

The battery comprised eight published population pharmacokinetic priors for three antimicrobials, chosen to span the range of critical-care populations in which these drugs are dosed, including neurosurgical, continuous-infusion sepsis, extracorporeal membrane oxygenation, continuous kidney replacement therapy, augmented renal clearance, and paediatric critical care (Table 1). Each prior is characterised by its between-subject coefficient of variation for clearance, denoted omega and reported as ω_{CL}. The fixed denominators, set before the run, were 500 simulated patients per prior for the recovery analysis (4,000 in total), 500 patients per direction for the external evaluation (2,500 held-out concentration predictions per direction), ten clinical facts with two false-fact controls, and three demonstration patients.

Prior	Drug	Population	ω_{CL}
Wei et al. (2022)	Vancomycin	Neurosurgical, adult	0.215
He et al. (2021)	Vancomycin	Paediatric, augmented renal clearance	0.226
Roberts et al. (2011)	Vancomycin	Continuous infusion, adult sepsis	0.389
Hahn et al. (2021)	Piperacillin–tazobactam	Adult, ECMO	0.229
Selig et al. (2022)	Piperacillin–tazobactam	Adult, CKRT	0.616
Klastrup et al. (2020)	Piperacillin	Adult, continuous infusion	0.574
Martinez-Casanova et al. (2023)	Piperacillin	Adult, continuous infusion	0.436
Thy et al. (2022)	Piperacillin	Paediatric, CRRT	0.520

Table 1. Population pharmacokinetic priors comprising the validation battery. ω_{CL} is the between-subject coefficient of variation for clearance reported by the source model. ECMO, extracorporeal membrane oxygenation; CKRT/CRRT, continuous kidney/renal replacement therapy.

2.5 Analytic core

The deterministic exposure calculation on which individualisation depends was validated separately, under its own pre-registered and locked bars, as a precondition for the present study. In that validation the engine reproduced published exposure values to within 2% (worst case 1.77% across seven quantitative cases), the numerical integrator agreed with the closed-form analytic solution to within 0.1% (worst case 0.077%), and every case with missing or unusable inputs was refused rather than answered (ten of ten). The exposure arithmetic is therefore treated here as established, and the present study concerns the recovery of individual parameters and the calibration of the resulting estimates.

2.6 Individualisation procedure and metrics

Individualisation updated the patient's clearance only; volume of distribution and inter-compartmental clearance were held at their population-typical values. This clearance-only scope is a deliberate restriction and is noted as such in the Limitations. The posterior was obtained by maximum-a-posteriori estimation with a local-Gaussian (Laplace) approximation to the posterior, and attainment probabilities were propagated by Monte-Carlo sampling over the posterior; these are standard pharmacometric procedures. Exposure was expressed as AUC_{24}/MIC for vancomycin and as the percentage of the dosing interval with free drug concentration above the MIC for the beta-lactams.

Three properties of the recovered individual estimate were assessed against pre-specified bars, each expressed relative to the prior's own clearance variability. Bias, the mean error of the recovered clearance on the log scale, was required to satisfy an absolute value no greater than $0.15 \cdot \omega_{CL}$. Imprecision, the root-mean-square error of the recovered clearance, was required to be no greater than $0.6 \cdot \omega_{CL}$; this corresponds to the two-sample update explaining approximately 64% of the between-subject clearance variance, and is a deliberately demanding criterion. Calibration was assessed as the empirical coverage of the posterior credible intervals at nominal levels of 50, 80, 90, and 95%, and was required to fall no more than 10 percentage points below nominal at every level. For the external evaluation, the individualised forecast was required to reduce the root-mean-square prediction error to no more than 0.80 times that of the population-typical forecast, and was required not to worsen the mean prediction error.

2.7 Controls and fact-checking

Ten clinical facts central to the demonstration scenarios were each checked against a cited external reference, with critical safety facts required to match at 100%, overall facts at no less than 90%, and no false statement asserted. Two deliberately false statements were presented as controls to confirm that they were withheld. Three further negative controls were specified: a drug with no available prior, which must be refused with no numeric dose; an uninformative measurement, which must produce a negligible

posterior update; and a prior whose residual-error structure could not be extracted from its source, which must be refused rather than completed with a fabricated error model.

3. Results

3.1 Calibration

Empirical coverage of the posterior credible intervals met the pre-registered bar at all four nominal levels for every population prior (Figure 1). Across the eight priors, empirical coverage ranged from 47.8 to 54.4% at the nominal 50% level and from 92.2 to 97.6% at the nominal 95% level, tracking the nominal level closely and never falling more than 10 percentage points below it. Calibrated coverage is a necessary condition for safe dose individualisation, because it indicates that the stated uncertainty can be relied upon at its nominal rate; the local-Gaussian posterior approximation was adequate at this design, and the under-coverage that would have required a more computationally intensive posterior was not observed.

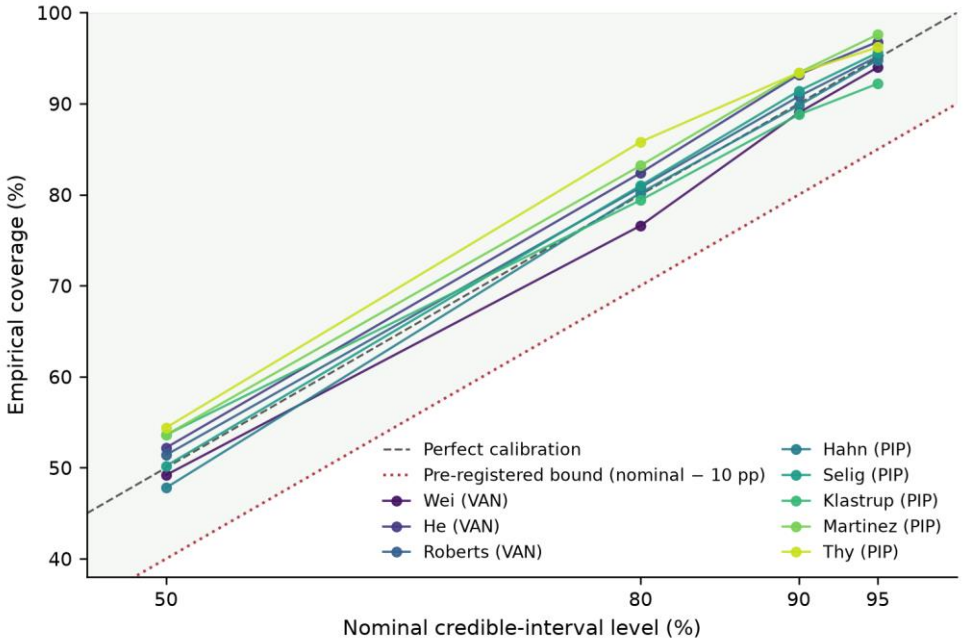

Figure 1. Posterior calibration. Empirical coverage of the credible intervals against the nominal level for each of the eight population priors. The dashed line denotes perfect calibration and the dotted line the pre-registered lower bound (nominal − 10 percentage points); all priors lie on or above the bound at every level.

3.2 Within-population recovery: bias and imprecision

Recovery of the individual clearance was near-unbiased for seven of the eight priors (Table 2, Figure 2). Every bias was negative, consistent with the expected shrinkage of the maximum-a-posteriori estimate toward the prior mean under a sparse two-sample design. The single exceedance was the Thy paediatric prior, the model carrying the largest residual error in the battery, for which the bias was −0.080 against a bar of 0.078, a margin of 0.002. This is the high-variability shrinkage signature that was predicted in the pre-registration; the bar was not widened and the sampling design was not enriched to accommodate it.

Imprecision met the $0.6 \cdot \omega_{CL}$ bar for five of the eight priors. The three exceedances were the Wei and Hahn priors, which are the two tightest priors in the battery, and the Thy prior, which carries the largest residual error. For these three, the two-sample update explained 56, 63, and 64% of the clearance variance respectively, just short of the approximately 64% implied by the bar. This pattern follows from the structure of the criterion rather than from a deficiency of the update: the absolute estimation-error floor set by residual noise and a two-sample design does not shrink in proportion to ω, so priors with little variability to resolve, or with large measurement noise, are the most demanding cases for a fraction-of-omega bar.

Prior	ω_{CL}	Bias (bar 0.15·ω)	RMSE (bar 0.6·ω)	C50	C80	C90	C95
Wei (VAN)	0.215	−0.029 (0.032)	0.142 (0.129)	49.2	76.6	89.0	94.0
He (VAN)	0.226	−0.009 (0.034)	0.128 (0.136)	52.2	82.4	93.2	96.8
Roberts (VAN)	0.389	−0.004 (0.058)	0.144 (0.233)	51.4	80.8	90.8	95.2
Hahn (PIP)	0.229	−0.019 (0.034)	0.140 (0.137)	47.8	80.2	89.8	94.8
Selig (PIP)	0.616	−0.057 (0.093)	0.299 (0.370)	50.2	81.0	91.4	95.6
Klastrup (PIP)	0.574	−0.009 (0.086)	0.117 (0.344)	53.6	79.4	88.8	92.2
Martinez (PIP)	0.436	−0.029 (0.065)	0.181 (0.262)	53.6	83.2	93.4	97.6
Thy (PIP)	0.520	−0.080 (0.078)	0.313 (0.312)	54.4	85.8	93.4	96.2

Table 2. Within-population recovery (500 simulated patients per prior). Bias is the mean error and RMSE the root-mean-square error of the recovered log-clearance, each with its pre-registered bar in parentheses. C50–C95 are empirical coverage (%) at the nominal 50, 80, 90, and 95% levels; the coverage bars (nominal − 10 pp: 40, 70, 80, 85%) were met by all priors. Values exceeding a bar: bias for Thy; RMSE for Wei, Hahn, and Thy.

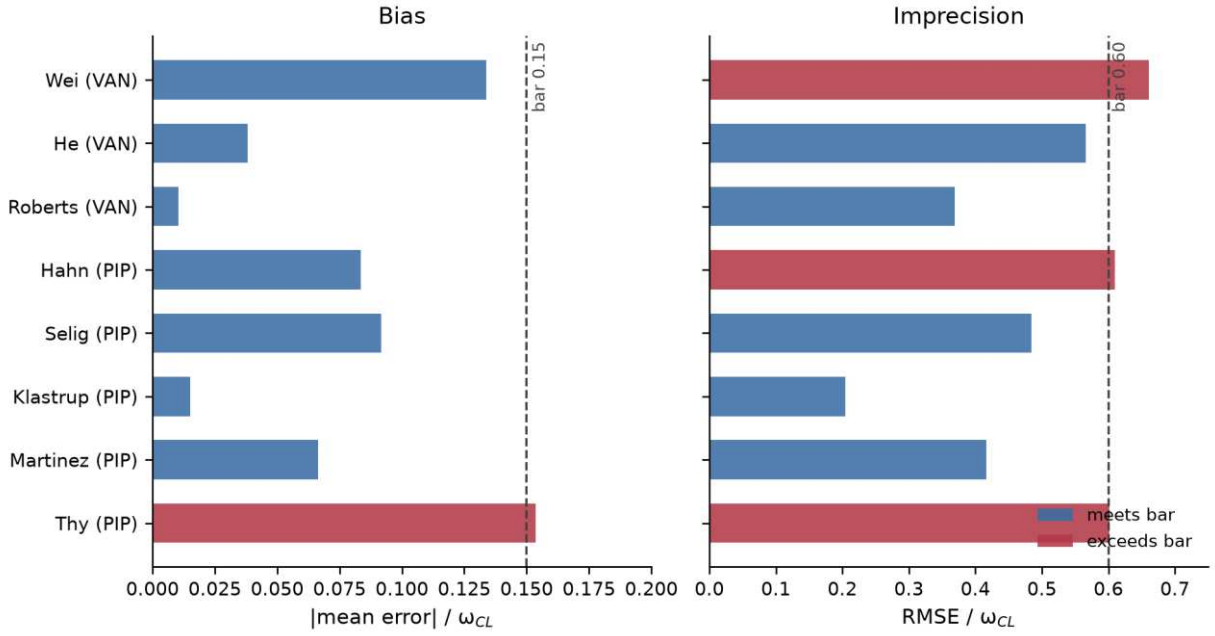

Figure 2. Within-population recovery against the pre-registered bars, expressed as a fraction of each prior's clearance variability ω_{CL}. Bias (left) exceeded the 0.15 bar for one prior (Thy); imprecision (right) exceeded the 0.60 bar for three priors (Wei, Hahn, Thy). Bars meeting the criterion are shown in blue and those exceeding it in red.

3.3 Within-population external evaluation

External evaluation tested whether a single measurement improves prediction when the prior is not the model that generated the data, which is the situation in routine practice. This analysis requires two priors for the same drug developed in the same population class; the adult vancomycin priors provided that

pairing. Data were simulated from one adult vancomycin model and forecast using the other, comparing the population-typical forecast with the forecast obtained after a single early trough was incorporated (Table 3, Figure 3). The individualised forecast reduced the root-mean-square prediction error by 52% in one direction (ratio 0.477) and by 32% in the other (ratio 0.683), both within the pre-registered bar of 0.80. In both directions the mean prediction error also fell, from 149.6 to 44.3% and from −29.0 to −0.7%, so the update improved accuracy without introducing bias.

Direction (truth → prior)	Prior MPE%	Prior RMSE%	Individualised MPE%	Individualised RMSE%	RMSE ratio
Wei → Roberts	149.6	813.5	44.3	388.3	0.477
Roberts → Wei	−29.0	115.9	−0.7	79.1	0.683

Table 3. Within-population external evaluation for the adult vancomycin priors (500 patients, 2,500 held-out concentrations per direction). MPE, mean prediction error; RMSE, root-mean-square prediction error. The RMSE ratio is the individualised value divided by the population-prior value; the pre-registered bar was 0.80.

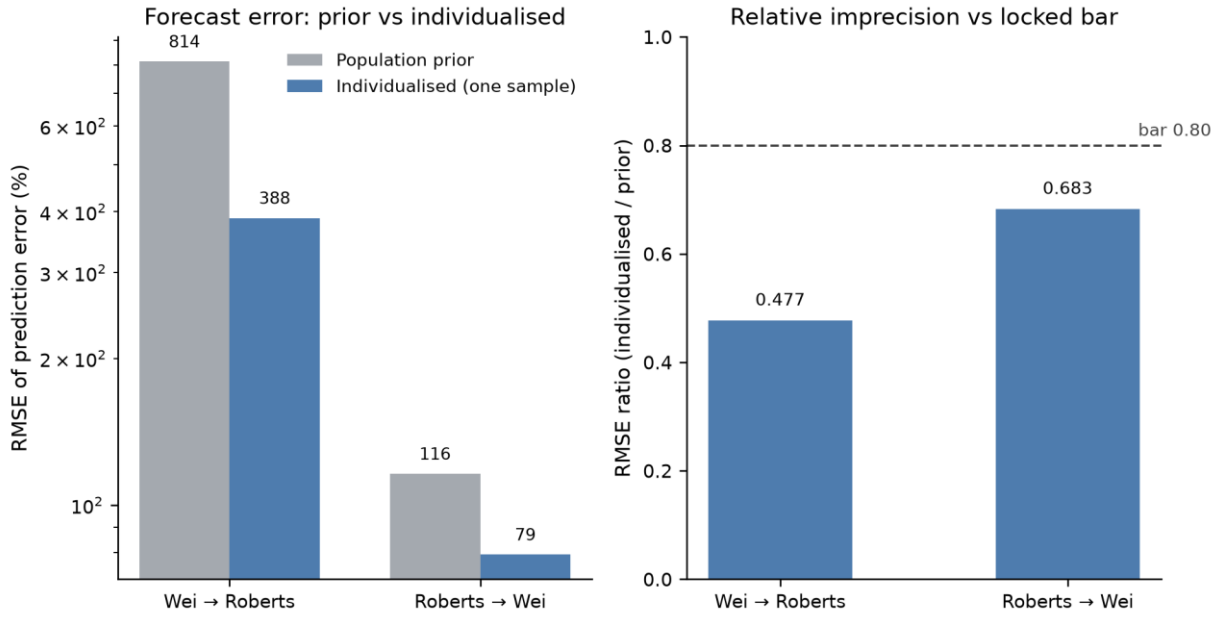

Figure 3. Within-population external evaluation. Left: prediction error of the population prior and of the individualised forecast, by direction (logarithmic axis). Right: the ratio of individualised to prior RMSE against the pre-registered bar of 0.80; both directions met the bar.

3.4 Facts and controls

The engine matched 100% of critical safety facts and 100% of facts overall, and asserted none of the two false-fact controls. The negative controls behaved as specified: a drug with no available prior (gentamicin) was refused with no numeric dose; an uninformative measurement produced a posterior update below the pre-specified threshold; and a prior whose residual-error structure could not be extracted from its source was refused rather than completed. One outcome is reported with its detail. The pre-registration listed amoxicillin among the no-prior refusal examples, but because a beta-lactam class anchor exists the engine instead returned a declared, explicitly uncertainty-bearing prediction labelled as not validated, rather than refusing. This is a governed prediction tier and not a fabricated definite dose; it is reported as observed, and the clean no-prior refusal control is therefore gentamicin.

3.5 Demonstration patients

Three representative complex intensive-care patients were processed end to end. For the first (vancomycin, acute kidney injury progressing to continuous renal replacement, obese), the engine produced an individualised posterior with an individual clearance of 3.42 L/h and an AUC_{24}/MIC posterior of 873 (10th to 90th percentile 694 to 1098), against a population-prior interval of 1207 (909 to 1600); the credible interval narrowed by 41% and sat above the 400 to 600 target, indicating a dose reduction (Figure 4). The selection record carried an explicit flag that the patient was receiving renal replacement while the selected prior did not cover that modality, surfacing a coverage gap rather than concealing it. For the second patient (piperacillin-tazobactam, augmented renal clearance, hypoalbuminaemic), the engine produced an individualised posterior consistent with high clearance; the selection step exposed a weakness, in that the locked selector placed this patient, who was not receiving renal replacement, on a continuous-kidney-replacement prior after a tie among the better-matched models, one of which lacked a required covariate. For the third patient (meropenem, continuous renal replacement, receiving valproate), the engine refused: the correct renal-replacement prior was selected, but its residual-error structure could not be extracted from the source, so the likelihood was under-specified, and the engine named the missing component rather than proceeding.

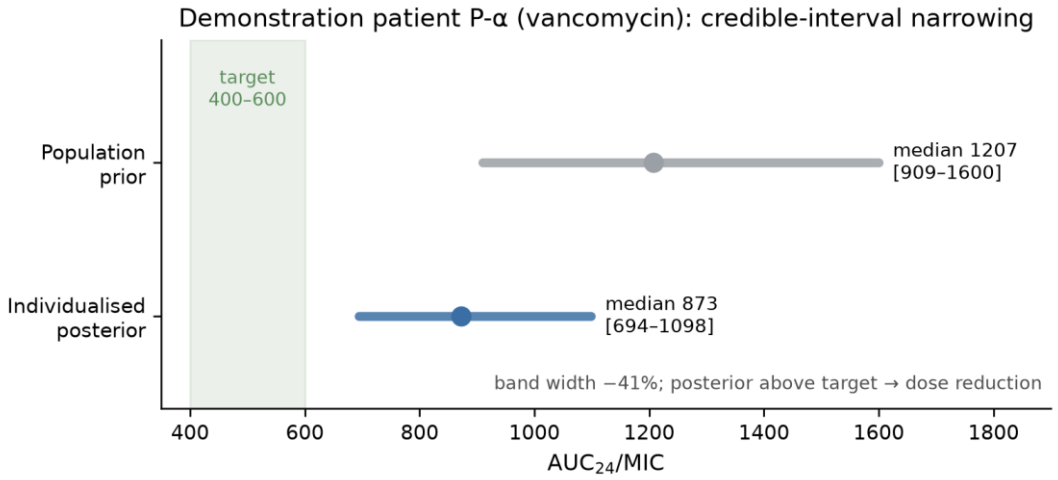

Figure 4. Demonstration patient P-α (vancomycin). The population-prior and individualised-posterior credible intervals for AUC_{24}/MIC, with the 400–600 target region shaded. Incorporating the patient's measurement narrowed the interval by 41%; the posterior lies above the target, indicating a dose reduction.

3.6 Comparison with an unaided language model

To isolate what the engine contributes beyond a capable language model, the identical task — same patient, same measured concentration, same target — was posed to a frontier language model with no pharmacokinetic engine attached. It cannot return a calibrated exposure posterior at all: it produces a point recommendation with no quantified uncertainty, no coverage guarantee, and no mechanism to decline when data are insufficient. The contrast is categorical, not a matter of degree. The calibrated, abstaining behaviour of Sections 3.1 and 3.4 is therefore attributable to the combination of substrate and runtime rather than to the underlying model — and it is exactly the property an unaided model cannot supply that determines whether such a system is safe near a dosing decision.

4. Discussion

Within the population in which each prior was developed, the engine produced calibrated individual estimates and improved prediction from a single measurement. Calibration held across all eight priors at every credible-interval level tested, which is the property most directly tied to safe dose selection. Recovery of the individual clearance was near-unbiased for all but the highest-residual prior, and the external evaluation showed that incorporating one measurement reduced prediction error by a third to a half relative to the population prior while not worsening bias. These findings concern the correctness, calibration, and transparency of the individualisation, not the originality of the method, which is established; the contribution under development is that the priors and the surrounding clinical facts are obtained from a compiled substrate, and the present results bound what can be claimed for that capability at the current scale of compilation.

The points at which the engine missed a pre-registered bar are informative and were retained in full. The single bias exceedance and the three imprecision exceedances are consequences of a sparse two-sample design interacting with priors that are either tightly constrained or noisy, and the bias exceedance in particular was anticipated in the pre-registration as the expected shrinkage behaviour of the estimator. The transparency behaviours are part of the result rather than incidental to it: the engine refused when a likelihood was under-specified, withheld false statements, returned an explicitly labelled prediction rather than a definite dose where only a class analogy was available, and surfaced a model-coverage gap and a model-selection weakness in the demonstration patients instead of masking them. For a system intended to support dosing decisions, the capacity to decline and to expose the limits of its own inputs is as important as the accuracy of the estimates it does return. Because these behaviours are properties of the engine rather than of the drugs studied, they are the part of the method that would carry to any further drug; what each new drug adds is its population model, not a new safety argument.

Model-informed precision dosing is available in established commercial systems (for example DoseMeRx, InsightRX Nova, PrecisePK, BestDose), and this work does not claim the dosing method as novel. What those systems do not provide, and what is at issue here, is a different set of properties: priors and clinical facts carrying explicit provenance from a single compiled substrate rather than a hand-curated model library; uncertainty represented as a first-class, type-enforced property of every estimate rather than an optional output; an engine that declines or flags rather than guessing when inputs are insufficient; and validation under pre-registration with cryptographic locking and byte-for-byte determinism. The contribution is the discipline and the provenance, not the arithmetic.

5. Limitations

1. The evidence is from simulation and from cross-model external evaluation, not from raw patient-level data; prospective validation against measured concentrations in real patients is the primary outstanding step and is not provided here.
2. Individualisation updated clearance only, with volume of distribution and inter-compartmental clearance held at population-typical values; parameters other than clearance were not individualised.
3. One prior (Thy) exceeded the pre-registered bias bar, by 0.002; this is a real exceedance and is the anticipated shrinkage behaviour of maximum-a-posteriori estimation under sparse sampling, not an artefact.

4. Three priors (Wei, Hahn, Thy) exceeded the imprecision bar, for which a two-sample design explained 56 to 64% of clearance variance; meeting that bar for tightly constrained or noisy priors would require richer sampling.
5. Two model-selection weaknesses were surfaced but not corrected, because the engine was locked for the run: a vancomycin prior that did not cover renal replacement was used for a patient receiving it, with the gap flagged; and an augmented-renal-clearance patient was placed on a renal-replacement prior after a tie among candidates.
6. Target attainment for the beta-lactams rests on an assumed unbound fraction, which is a modelling assumption surfaced per case rather than a per-patient measurement.
7. The battery covers three antimicrobials and eight priors; the wider narrow-therapeutic-index drug class is structurally supported but was not validated here, and no claim is made for it.
8. The posterior was obtained with a local-Gaussian (Laplace) approximation, which was adequately calibrated at the sampling design used here but should not be assumed to remain so for arbitrary sampling schedules.
9. The fact-checking used a curated set of ten facts with two controls, which establishes a floor rather than a comprehensive safety evaluation.
10. Generalisation across populations, for example between paediatric and adult patients, is a separate validation question that this study did not assess; it requires a different design and is identified as future work.

6. Programme context and significance

This section places the study in the context of the wider programme. It is separated deliberately: the claims validated above are bounded strictly to what was demonstrated; the paragraphs here describe direction, and should be read as context, not results.

This study is the first in a planned series of pre-registered verification proof points for a single, compiled, provenance-bearing medical-knowledge substrate. Each has the same aim: to demonstrate, under pre-registration and reproducible locking, that the system performs a defined clinical task correctly, with calibrated uncertainty, and with honest abstention — properties built in from the start rather than retrofitted.

The capability under validation is a method rather than a result specific to one drug. The individualisation engine and its honesty guardrails — the calibrated posterior, the refusal when inputs are insufficient, and the explicitly labelled prediction — are not specific to antimicrobials but are drug-agnostic; the dosed drug enters only through its population model. Antimicrobials are the first instance because they offered both published population models and a directly measurable concentration target.

The validation was performed at an early stage of compilation, well below the intended scale of the substrate; the claims above are bounded accordingly. The significance is not in the magnitude of what has been compiled but in the method: that calibrated, honestly-abstaining behaviour on a safety-critical task is achievable this early is evidence about the approach, not about its current reach.

What this study does not yet show is the distinctive contribution of the substrate's content: the priors used here are published models, chosen precisely so the arithmetic could be checked against the public literature. Demonstrating that substrate-sourced priors improve on conventional model libraries, and that the behaviour generalises across populations and on raw patient data, are the explicit next steps.

7. Conclusion

Under a pre-registered and reproducible protocol, an engine using priors drawn from a compiled medical-knowledge substrate produced calibrated within-population individual exposure estimates for eight antimicrobial population models and reduced single-sample prediction error by 32 to 52% relative to the population prior, without worsening bias. The exceedances observed were pre-specified consequences of the estimator and the sparse-sampling design, and the engine declined when its inputs were insufficient. Within-population individualisation is thereby demonstrated at the current scale of the substrate.

Competing interests

The author is the founder of GATMEDI, the company developing the individualisation engine evaluated in this study.

Funding

This work was funded internally by GATMEDI. It received no external funding.

Data and code availability

The individualisation engine, its source code, and the underlying compiled substrate are proprietary and are not available. The study used simulated and cross-model data generated from the published population-pharmacokinetic models cited in the References; those population priors are therefore publicly available, but the engine and the evaluation harness are not.

Ethics

This study used only simulation and cross-model evaluation. It involved no human subjects and no patient data, and did not require ethics approval.

References

1. Wei S, Zhang D, Zhao Z, Mei S. Population pharmacokinetic model of vancomycin in postoperative neurosurgical patients. Front Pharmacol. 2022;13:1005791. doi:10.3389/fphar.2022.1005791.
2. He CY, Ye PP, Liu B, Song L, van den Anker J, Zhao W. Population pharmacokinetics and dosing optimization of vancomycin in infants, children, and adolescents with augmented renal clearance. Antimicrob Agents Chemother. 2021;65(10):e00897-21. doi:10.1128/AAC.00897-21.
3. Roberts JA, Taccone FS, Udy AA, Vincent JL, Jacobs F, Lipman J. Vancomycin dosing in critically ill patients: robust methods for improved continuous-infusion regimens. Antimicrob Agents Chemother. 2011;55(6):2704-2709. doi:10.1128/AAC.01708-10.
4. Hahn J, Min KL, Kang S, Yang S, Park MS, Wi J, Chang MJ. Population pharmacokinetics and dosing optimization of piperacillin-tazobactam in critically ill patients on extracorporeal membrane oxygenation and the influence of concomitant renal replacement therapy. Microbiol Spectr. 2021;9(3):e00633-21. doi:10.1128/Spectrum.00633-21.
5. Selig DJ, DeLuca JP, Chung KK, et al. Pharmacokinetics of piperacillin and tazobactam in critically ill patients treated with continuous kidney replacement therapy: a mini-review and population pharmacokinetic analysis. J Clin Pharm Ther. 2022;47(8):1091-1102. doi:10.1111/jcpt.13657.

6. Klastrup V, Thorsted A, Storgaard M, Christensen S, Friberg LE, Öbrink-Hansen K. Population pharmacokinetics of piperacillin following continuous infusion in critically ill patients and impact of renal function on target attainment. Antimicrob Agents Chemother. 2020;64(7):e02556-19. doi:10.1128/AAC.02556-19.

7. Martínez-Casanova J, et al. Predictive factors of piperacillin exposure and the impact on target attainment after continuous infusion administration to critically ill patients. Antibiotics (Basel). 2023;12(3):531. doi:10.3390/antibiotics12030531.

8. Thy M, Urien S, Foissac F, et al. Piperacillin population pharmacokinetics and dosing regimen optimization in critically ill children receiving continuous renal replacement therapy. Antimicrob Agents Chemother. 2022;66(12):e0113522. doi:10.1128/aac.01135-22.

9. Peng Y, Liu Y, Cheng Z, Zhang Q, Xie F, Zhu S, Li S. Population pharmacokinetics of prolonged infusion for meropenem: tailoring dosing recommendations for Chinese critically ill patients on continuous renal replacement therapy with consideration for renal function. Drug Des Devel Ther. 2025;19:1105-1117. doi:10.2147/DDDT.S489603.

10. Roberts JA, Paul SK, Akova M, et al. DALI: defining antibiotic levels in intensive care unit patients: are current β-lactam antibiotic doses sufficient for critically ill patients? Clin Infect Dis. 2014;58(8):1072-1083. doi:10.1093/cid/ciu027.

11. Ewoldt TMJ, Abdulla A, Hunfeld NGM, et al. Health care costs of target attainment for beta-lactam antibiotics in critically ill patients: a retrospective analysis of the EXPAT study. Ther Drug Monit. 2022;44(1):224-229. doi:10.1097/FTD.0000000000000891.

12. Rybak MJ, Le J, Lodise TP, et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant Staphylococcus aureus infections: a revised consensus guideline and review by ASHP, IDSA, PIDS, and SIDP. Am J Health-Syst Pharm. 2020;77(11):835-864. doi:10.1093/ajhp/zxaa036.

13. Joint Formulary Committee. British National Formulary (online). London: BMJ Group and Pharmaceutical Press. Available from: https://www.medicinescomplete.com

14. Preston CL, editor. Stockley's Drug Interactions. 12th ed. London: Pharmaceutical Press; 2019.

15. European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters, version 16, 2026. https://www.eucast.org/clinical_breakpoints.

16. Picard M, Robitaille G, Karam F, et al. Cross-reactivity to cephalosporins and carbapenems in penicillin-allergic patients: two systematic reviews and meta-analyses. J Allergy Clin Immunol Pract. 2019;7(8):2722-2738.e5. doi:10.1016/j.jaip.2019.05.038.

Appendix. Provenance and reproducibility

The run reported here was banked from a frozen, content-hashed engine under a pre-registered lock. The deterministic results reproduce byte-for-byte on re-running the locked engine and harness at the fixed seed.

Item	Value
Lock authority (commit)	ae0dc61
Firewalled snapshot (commit)	f423c51
Results content hash (SHA-256)	193f66b1…
Snapshot report / results hashes	8212fa36… / 74a1fed4…
Analytic core (SHA-256)	6d9ed7fd…
Individualisation layer (SHA-256)	c1f2378f…
Random seed	20260627
Sample size	500 per prior (recovery) and per direction (external evaluation)
Analytic-core validation run hash	49ee1f1c…
Wall-clock runtime	207 s