

Article

Leveraging ChatGPT for Vancomycin Therapeutic Drug Monitoring: Simulation Using Bayesian Estimation and Hyperparameter Optimization

Akira Kageyama ^{1,2}, Takahiko Aoyama ^{2,*} , Rikuya Maehara ³, Dai Harada ¹, Takashi Kawakubo ¹ and Yasuhiro Tsuji ² 

¹ Department of Pharmacy, The Jikei University Hospital, 3-19-18 Nishi Shimbashi, Tokyo 105-8471, Japan; kageyama.akira@jikei.ac.jp (A.K.); d-harada@jikei.ac.jp (D.H.); kawakubo@jikei.ac.jp (T.K.)

² Laboratory of Clinical Pharmacometrics, School of Pharmacy, Nihon University, Funabashi 274-8555, Japan; tsuji.yasuhiro@nihon-u.ac.jp

³ Department of Pharmacy, The Jikei University Kashiwa Hospital, 163-1, Kashiwashita, Kashiwa 277-8567, Japan; y141167@jikei.ac.jp

* Correspondence: aoyama.takahiko@nihon-u.ac.jp

Abstract

The usefulness of ChatGPT, a large language model, has recently been explored in medical research. However, no studies have examined its reproducibility or applicability to therapeutic drug monitoring (TDM), a core task of clinical pharmacists. In this simulation study, we evaluated the feasibility of using ChatGPT for vancomycin (VCM) TDM based on Bayesian estimation. A total of 1000 virtual patients were generated by Monte Carlo simulations using a population pharmacokinetic model of VCM. Bayesian-estimated pharmacokinetic parameters and predicted concentrations were input into ChatGPT, and dosage regimens were compared among the three conditions, using temperature as a hyperparameter ($T = 0.1, 0.5$, and 1.0). Reproducibility was evaluated using the mode percentage in repeated runs. The reproducibility of the ChatGPT output was higher at $T = 0.1$ than at $T = 0.5$ and $T = 1.0$. When ChatGPT simulated the mode-recommended regimen ($T = 0.1$), the target attainment rate of the area under the serum concentration (AUC) ($400\text{--}600\text{ mg}\cdot\text{h/L}$) improved from 25.5% (pre-optimization AUC (fixed-dose regimen)) to 71.5% (post-optimization AUC (ChatGPT-guided regimen)). These findings demonstrate that ChatGPT-based TDM using Bayesian estimation can enhance dose optimization. Adjusting the hyperparameter temperature to 0.1 improved reproducibility, suggesting that a reliable ChatGPT-assisted TDM support system may be clinically useful.



Academic Editor: Osvaldo Andrade Santos-Filho

Received: 8 March 2026

Revised: 26 April 2026

Accepted: 27 April 2026

Published: 29 April 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the Österreichische Pharmazeutische Gesellschaft. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

Keywords: ChatGPT; therapeutic drug monitoring; vancomycin; simulation study; temperature; Bayesian estimation

1. Introduction

In recent years, artificial intelligence (AI) has demonstrated significant potential in medical research, including accelerating drug discovery and predicting disease outcomes [1]. A notable advancement in AI technology is the release of ChatGPT, a large language model (LLM) developed by OpenAI (San Francisco, CA, USA) [2]. ChatGPT enables human-like conversations and has demonstrated promising performance in tasks such as text classification and answering questions [3]. A recent scoping review focusing on ChatGPT and its role in pharmacy practice revealed not only its usefulness for medication

advice, drug dosage calculation, and drug–drug interactions but also highlighted limitations such as variability in accuracy, lack of reproducibility, and challenges in applying it to personalized treatment [4]. A recent study comparing three LLMs—ChatGPT, Gemini, and Copilot—on common pharmacokinetic problems showed the superior performance of ChatGPT; it achieved the highest number of correct answers [5]. This finding suggests that LLMs may serve as an educational strategy for teaching pharmacokinetics to clinical pharmacy students and that ChatGPT holds promise as a supportive tool for therapeutic drug monitoring (TDM) operations. Despite these advances, research integrating LLM with pharmacokinetics remains limited, and no studies have reported the use of LLM in clinical TDM.

Vancomycin (VCM) is widely used to treat serious infections such as bacteremia, endocarditis, pneumonia, and meningitis and remains the first-line therapy for methicillin-resistant *Staphylococcus aureus* infections [6–8]. The efficacy of VCM is closely associated with the ratio of the area under the serum concentration (AUC)–time curve to the minimum inhibitory concentration; however, its dose-dependent nephrotoxicity remains a major concern [9,10]. Therefore, TDM is essential to optimize efficacy while minimizing toxicity. Current guidelines recommend AUC-based dosing, assuming a minimum inhibitory concentration (MIC) of 1 mg/L, whereby the AUC/MIC target is simplified to AUC alone [11,12]. Furthermore, recent observational studies and meta-analyses have demonstrated that AUC-based dosing reduces nephrotoxicity compared with trough concentration ($C_{\min}$)-based dosing targeting 15–20 mg/L [13–16]. Within the framework of model-informed precision dosing, Bayesian estimation using population pharmacokinetic (PopPK) models enables accurate AUC prediction and supports individualized therapy [17]. Presently, in clinical practice, TDM platforms typically require manual intervention by healthcare professionals, involving the interpretation of Bayesian-estimated pharmacokinetic parameters and the iterative design of dosing plans to achieve target AUC values. By contrast, integrating ChatGPT with Bayesian estimation could potentially automate the dosing plan design process by directly translating pharmacokinetic parameters into clinically interpretable dosing recommendations. This approach could complement the existing TDM platforms by maintaining individualized dosing strategies while reducing the workload associated with TDM. Therefore, combining Bayesian estimation based on PopPK models that incorporate individual serum concentration data with ChatGPT could be a useful tool for TDM in VCM.

However, ChatGPT includes a hyperparameter, temperature (T), which controls the balance between the randomness and determinism of its output. Higher temperatures generate more diverse and creative responses, whereas lower temperatures yield more deterministic and conservative outputs [18,19]. Accuracy and reproducibility are critical for clinical VCM TDM when designing regimens based on Bayesian estimations. Thus, temperature settings may influence the accuracy and reproducibility of ChatGPT-generated dosing recommendations. Therefore, optimizing this hyperparameter is essential for maximizing the utility of ChatGPT for VCM dosing support.

While the broader question of whether AI can be used for dose adjustment remains an important area of investigation, this study was designed with the purpose of serving as a methodological and exploratory analysis focusing on a fundamental aspect of LLM-based decision support, namely, the impact of hyperparameter settings on reproducibility and consistency. We conducted Monte Carlo simulations of virtual patients using a PopPK model of the VCM and evaluated how ChatGPT hyperparameters influenced the reproducibility and output behavior of dosing calculations derived from Bayesian-estimated pharmacokinetic parameters. This study is a methodological and exploratory investigation

to assess the feasibility and characteristics of an LLM-based support under controlled simulation conditions.

2. Materials and Methods

2.1. ChatGPT Method

The default ChatGPT model (gpt-4o-mini) [20] was accessed using the OpenAI API and programmatically controlled using the statistical software R (version 4.5.1, R Foundation for Statistical Computing, Vienna, Austria).

2.2. Generation of a Virtual Patient Dataset

The virtual patient was defined as a male aged 60 years, weighing 70 kg, with a VCM dose of 1000 mg administered twice daily using a 1 h infusion over a 6-day treatment period. Serum creatinine (SCr) level was fixed at 1.0 mg/dL at the initiation of VCM therapy. Creatinine clearance was calculated using the Cockcroft–Gault equation [21] based on age, sex, weight, and SCr. The dosage and administration were based on the maintenance dose specified in the guidelines, considering renal function and weight [12]. To isolate the effect of the ChatGPT hyperparameters on reproducibility and output behavior, patient characteristics and dosing conditions were fixed. This approach was adopted to minimize the variability arising from covariates and to allow for a controlled evaluation of the impact of the ChatGPT temperature settings on model outputs.

2.3. Population Pharmacokinetic Model for Monte Carlo Simulations

The population pharmacokinetic parameters for the central compartment volume of distribution (V_1), peripheral compartment volume of distribution (V_2), distribution clearance (Q), and total clearance (CL_{VCM}) were obtained from healthy Japanese subjects as reported by Yamamoto et al. [22] (Table S1). The rate constant for transfer from the central to the peripheral compartment (k_{12}) was calculated by dividing Q by V_1 : $k_{12} = Q/V_1$. The rate constant for the transfer from the peripheral compartment to the central compartment (k_{21}) was calculated by dividing Q by V_2 , where $k_{21} = Q/V_2$. The elimination rate constant (k_e) was calculated by dividing CL_{VCM} by V_1 : $k_e = CL_{VCM}/V_1$. Serum VCM concentrations for 1000 cases were generated using individual values of population pharmacokinetic parameters to reflect inter-individual variability, and log-normally distributed residual errors were added to simulate intra-individual variability. These simulated concentrations were regarded as true values.

2.4. Setting of C_{min} Sampling Points

The C_{min} values at 48 h (C_{min48h}), 96 h (C_{min96h}), and 144 h ($C_{min144h}$), generated from the simulated virtual patient dataset, were extracted from the concentration–time profiles and used as input data for Bayesian estimation.

2.5. Bayesian Estimation Settings

Bayesian estimation was performed using the fixed covariates method, in which SCr was fixed throughout the entire VCM dosing period. The estimation incorporated multiple trough concentrations (C_{min48h} , C_{min96h} , and $C_{min144h}$) obtained at 48, 96, and 144 h after the initiation of VCM therapy, which were jointly used as input data for the Bayesian estimation. The estimated pharmacokinetic parameters were subsequently used to reconstruct the concentration–time profiles for AUC calculation (see Section 2.7 for details).

2.6. Population Pharmacokinetic Model for Bayesian Estimation of Individual Pharmacokinetic Parameters

For Bayesian estimation, the $C_{\min}$ of the virtual patient was used for the calculations. The population pharmacokinetic parameters for the steady-state volume of distribution (V_{ss}), k_{12} , k_{21} , and CL_{VCM} were obtained from Japanese data reported by Yasuhara et al. [23] (Table S2). V_1 was calculated using V_{ss} , k_{12} , and k_{21} as $V_1 = k_{21} \times V_{ss} / (k_{12} + k_{21})$. ke was calculated by dividing CL_{VCM} by V_1 : $ke = CL_{VCM} / V_1$.

2.7. Method for Calculating AUC

Blood sampling points corresponding to those used in a previous clinical trial [24] were applied to each dosing interval, using an eight-point sampling scheme within one dosing interval ($\tau = 12$ h), to calculate AUC. Because VCM was administered twice daily, the AUC_{0-24} was calculated as the sum of AUCs over two consecutive dosing intervals using the trapezoidal method. True AUC values were calculated using the trapezoidal method by extracting serum VCM concentrations at eight sampling points from the VCM concentration–time profile generated when the dataset was created. The predicted AUC values were calculated from the predicted VCM concentration–time profile based on the pharmacokinetic parameters estimated by Bayesian estimation, and the trapezoidal method was applied using the VCM concentrations at the same eight sampling points.

2.8. Prompt Syntax in ChatGPT

The data input to ChatGPT was obtained using R version 4.5.1 and RStudio (version 2026.01.1, Posit PBC, Boston, MA, USA). The full prompt template is presented in Table S3. The prompts were structured in a stepwise format. Serum VCM concentrations and pharmacokinetic parameters were estimated externally by Bayesian estimation prior to being provided to ChatGPT. CL_{VCM} and V_1 , obtained from the Bayesian estimation, were entered into the prompt as input variables. ChatGPT did not perform Bayesian estimation or independently estimate pharmacokinetic parameters. The dose calculation procedure and the AUC calculation formula were explicitly predefined in the prompt, and ChatGPT was instructed to execute these calculations in a rule-based manner. Within the ChatGPT-based procedure, AUC was not calculated using the trapezoidal method but was instead calculated analytically using the predefined formula $AUC_{0-24} = (Dose \times 2) / CL_{VCM}$. The ChatGPT output format was standardized by specifying a unified response structure within the prompt. The administration schedule, SCr values measured at each dosing time and at the final blood sampling point, $C_{\min}$ at the final blood sampling point, predicted $C_{\min}$ at the final blood sampling point, and estimated pharmacokinetic parameters were entered individually into ChatGPT in the API using R version 4.5.1. The administration design using ChatGPT was compiled after setting a specified response format.

2.9. Hyperparameter Settings

Temperature is a hyperparameter that controls the degree of randomness in model outputs, with lower values producing more deterministic responses. In this study, three temperature settings ($T = 0.1$, 0.5 , and 1.0) were used for the reproducibility analysis, whereas $T = 0.1$ was used to assess the usefulness of ChatGPT for dose design. All other ChatGPT hyperparameters were held constant as follows: maximum tokens = 4096, seed = 1; top $p = 1$; frequency penalty = 0; and presence penalty = 0.

2.10. Reproducibility and Evaluation of the Usefulness of ChatGPT for Dosage Design

The overall workflow of virtual patient generation via Monte Carlo simulation, Bayesian estimation, ChatGPT-based dosage design, and evaluation of the reproducibility and usefulness of the ChatGPT recommendations is illustrated in Figure 1.

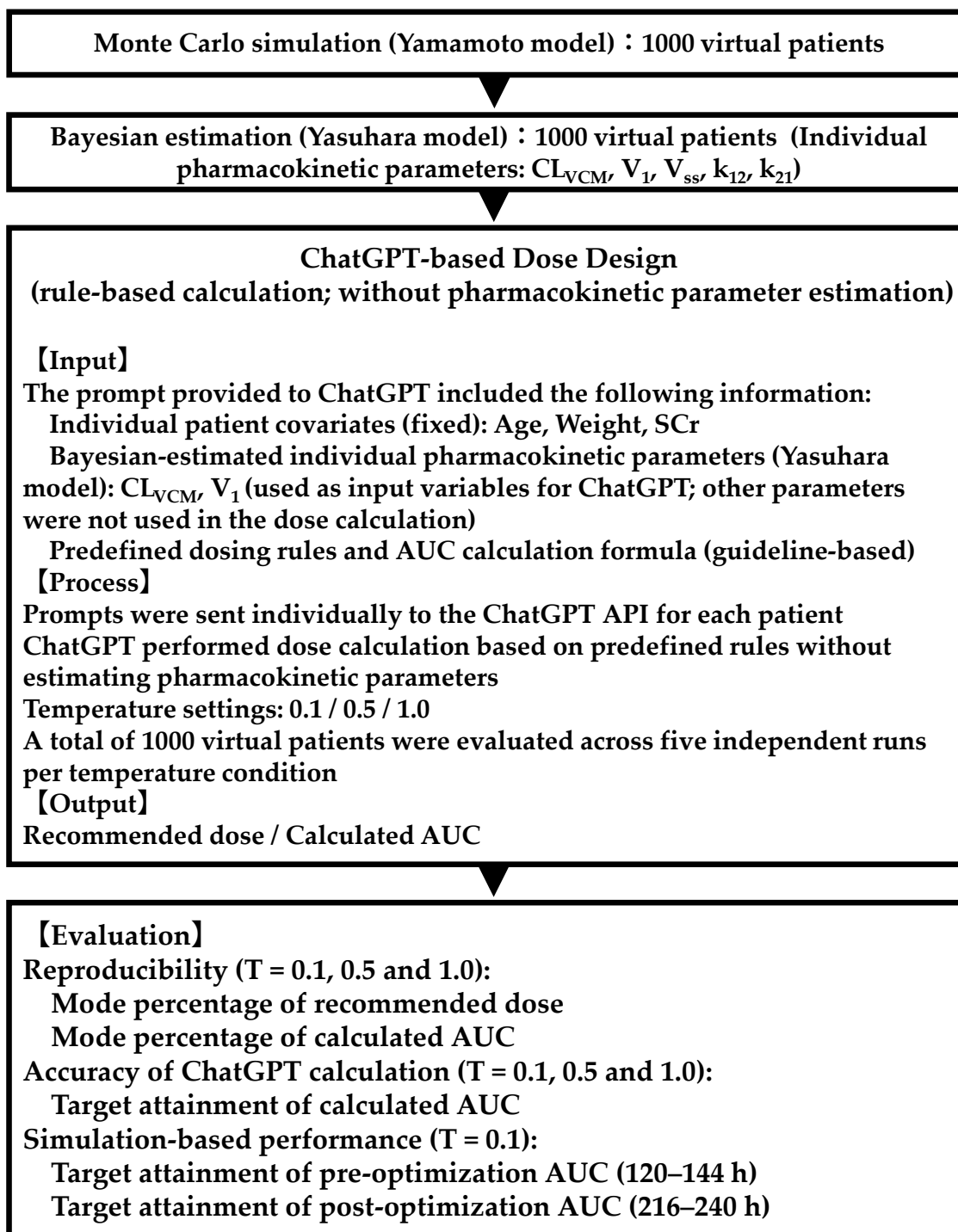


Figure 1. Research workflow.

A total of 1000 independent virtual patients were generated according to the procedures described in Sections 2.2–2.6. For the same 1000 patients, the recommended doses and calculated AUCs generated by ChatGPT were obtained five times (runs 1–5) under three temperature conditions (T = 0.1, 0.5, and 1.0), according to the procedures described in Sections 2.8 and 2.9.

Reproducibility at the individual-patient level across runs (runs 1–5) was evaluated using the mode percentage, and the results for T = 0.1, T = 0.5, and T = 1.0 were compared as paired data (identical patients). The mode percentage represents the proportion of identical outputs among the five repeated runs under each temperature condition, reflecting the

reproducibility of the ChatGPT-generated results rather than the continuous nature of the calculated variables. The mode percentage was calculated using Equations (1) and (2):

$$\text{Mode percentage of recommended dose (\%)} = \frac{\text{Number of mode recommended dose among 5 runs}}{5} \times 100 \quad (1)$$

$$\text{Mode percentage of calculated AUC (\%)} = \frac{\text{Number of mode calculated AUC among 5 runs}}{5} \times 100 \quad (2)$$

Because the reproducibility was assessed across five repeated runs, the possible values of the mode percentages were discrete and limited to 20%, 40%, 60%, 80%, and 100%.

The mode-calculated AUC from the five runs was adopted for each patient, and the target attainment rate of the AUC was calculated. The dosing regimens generated by ChatGPT were determined using prompts based on the calculated AUC. According to guidelines, an AUC of 400–600 mg·h/L is considered the target range [11,12]. Therefore, for each virtual patient, the target attainment rate of the ChatGPT-calculated AUC was determined using Equation (3).

$$\text{Target attainment rate of calculated AUC (\%)} = \frac{\text{Number of target attainment calculated AUC}}{1000} \times 100 \quad (3)$$

Here, the number of target attainment cases using the calculated AUC is the number of cases with $400 \leq \text{AUC} \leq 600 \text{ mg}\cdot\text{h/L}$.

The pre-optimization AUC (fixed-dose regimen) was defined as the AUC calculated from the simulated serum VCM concentrations 120–144 h after the initiation of VCM therapy under the initial fixed-dose regimen (1000 mg every 12 h). This pre-optimization AUC served as the baseline for comparison with the AUCs obtained using the ChatGPT-recommended regimens. To evaluate the clinical applicability of the dosing regimens generated by ChatGPT, the mode-recommended dosing schedule from the five runs at $T = 0.1$ was applied to the same virtual patient dataset using the pharmacokinetic parameters employed in the simulation based on the population pharmacokinetic model reported by Yamamoto et al. [22]. The time course of serum VCM concentrations from day 7 onward was calculated, and the model-simulated AUC was determined. The post-optimization AUC (ChatGPT-guided regimen) was evaluated at 216–240 h after dose adjustment at 144 h. The initial fixed-dose regimen was administered for up to 144 h, after which the dose was adjusted based on the ChatGPT recommendations and maintained thereafter. The AUC values used for the evaluation, including the pre-optimization AUC and post-optimization AUC, were calculated using the trapezoidal method, as described in Section 2.7. Target attainment rates were calculated using Equations (4) and (5):

$$\text{Target attainment rate of pre-optimization AUC (\%)} = \frac{\text{Number of target attainment pre-optimization AUC}}{1000} \times 100 \quad (4)$$

$$\text{Target attainment rate of post-optimization AUC (\%)} = \frac{\text{Number of target attainment post-optimization AUC}}{1000} \times 100 \quad (5)$$

Here, the number of target attainment cases using pre-optimization AUC and the number of target attainment cases using post-optimization AUC are the number of cases with $400 \leq \text{AUC} \leq 600 \text{ mg}\cdot\text{h/L}$.

2.11. Evaluation of Prediction Accuracy for $C_{\min}$ and AUC and Estimation Accuracy of Pharmacokinetic Parameter

The group in which the post-optimization AUC fell within the range of 400–600 mg·h/L was defined as the within-target-range group (IN group), whereas the group in which the post-optimization AUC fell outside this range was defined as the out-of-target-range

group (OUT group). The predictability of $C_{\min}$ and AUC and the estimation accuracy of the pharmacokinetic parameters were evaluated using mean absolute prediction error (MAPE) and mean prediction error (MPE). The MAPE and MPE values served as indicators for assessing the prediction accuracy, estimation accuracy, prediction bias, and estimation bias and were calculated using Equations (6) and (7), respectively.

$$\text{MAPE (\%)} = \frac{1}{n} \sum_{j=1}^n \left| \frac{\text{Predict} - \text{True}}{\text{True}} \right| \times 100 \quad (6)$$

$$\text{MPE (\%)} = \frac{1}{n} \sum_{j=1}^n \left\{ \frac{\text{Predict} - \text{True}}{\text{True}} \right\} \times 100 \quad (7)$$

In this study, the true values refer to the ground-truth serum VCM concentrations, AUCs, and pharmacokinetic parameters generated using Monte Carlo simulations.

The prediction accuracy and bias of the Bayesian estimated serum concentrations and the estimation accuracy and bias of the Bayesian estimated pharmacokinetic parameters were compared and evaluated between the two groups.

2.12. Analytical Environment

The generation of individual pharmacokinetic parameters reflecting inter-individual variability, the generation of residuals reflecting intra-individual variability, the simulation of serum VCM concentrations, the Bayesian estimation of individual pharmacokinetic parameters, and AUC calculations were performed using R version 4.5.1. Individual pharmacokinetic parameters and residuals reflecting inter- and intra-individual variability were generated using the `rnorm` function with a fixed random seed. Serum VCM concentration simulations were performed using the `rxode2` package, and serum VCM concentrations were obtained using the `RxODE` function. Bayesian estimation was performed using the `nlmixr2` package, and pharmacokinetic parameters and predicted serum VCM concentrations were obtained using the `nlmixr` function. AUC calculations were performed using the `DescTools` package, and AUC values were obtained from serum VCM concentrations using the `AUC` function.

2.13. Statistical Analysis

Appropriate and inappropriate classifications within paired data were compared using McNemar's test. The Mann–Whitney U test was used to evaluate the accuracy and variability of the Bayesian estimation between the IN and OUT groups. All statistical data were analyzed using R version 4.5.1. A p value of <0.05 was considered statistically significant.

3. Results

3.1. Reproducibility of Recommended Dose and Calculated AUC

The mode percentage values of the recommended doses and calculated AUCs at $T = 0.1$, $T = 0.5$, and $T = 1.0$ are shown in Figure 2. As shown in Figure 2, a temperature-dependent trend was observed, with higher mode percentage values of the recommended doses and calculated AUCs at $T = 0.1$, followed by $T = 0.5$ and $T = 1.0$. The target attainment rates for the calculated AUC were 98.8%, 96.9%, and 98.5% at $T = 0.1$, 0.5, and $T = 1.0$, respectively.

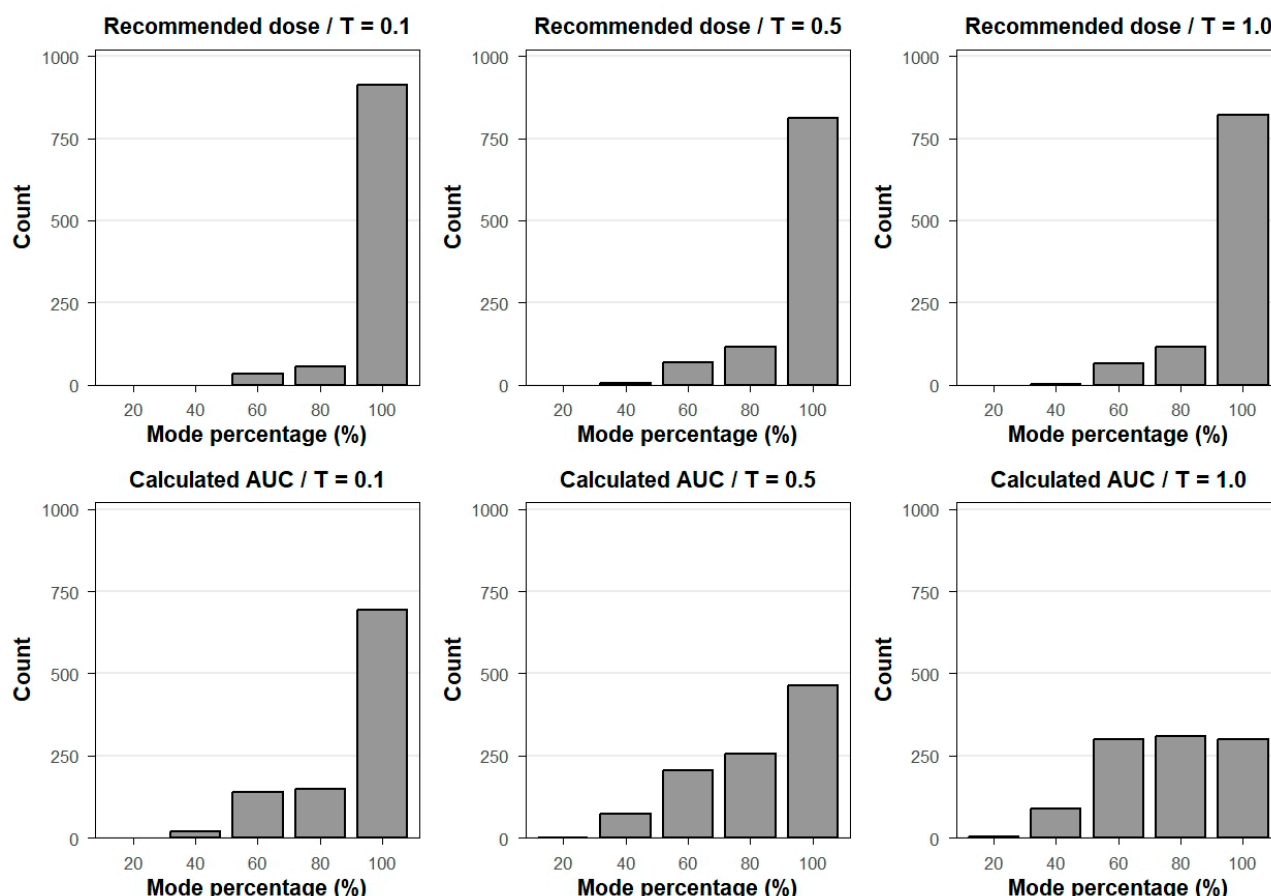


Figure 2. Mode percentage values of the recommended doses and calculated AUCs at $T = 0.1$, $T = 0.5$, and $T = 1.0$. The mode percentage represents the proportion of identical outputs under each temperature condition.

3.2. Target Attainment of Post-Optimization AUC Using Recommended Doses

The target attainment rates of the pre-optimization AUC and post-optimization AUC were 25.5% and 71.5%, respectively. Table 1 shows the target attainment of the pre-optimization AUC and post-optimization AUC. The target attainment rate of the post-optimization AUC at $T = 0.1$ was significantly higher than that of the pre-optimization AUC ($p < 0.01$). Figure S1 presents the histograms of the pre-optimization AUC and post-optimization AUC. The results indicated that the excessively high AUCs observed with the fixed-dose regimen were corrected to values within the target range when the recommended doses were applied at $T = 0.1$.

Table 1. Paired comparison of target attainment and non-attainment between pre-optimization AUC and post-optimization AUC.

		Post-Optimization AUC		p-Value
		Attainment	Non-Attainment	
Pre-optimization AUC	Attainment	226	29	<0.01 **
	Non-attainment	489	256	

** $p < 0.01$; McNemar's test.

3.3. Relationship Between Post-Optimization AUC and Bayesian Estimation

The IN and OUT groups comprised 715 and 285 patients, respectively. Within the OUT group, 5 patients had AUC values below 400 mg·h/L, whereas 280 patients had

AUC values above 600 mg·h/L. The MAPE and MPE of the Bayesian estimation for $C_{\min}$, AUC, and CL_{VCM} are summarized in Figure 3, and the relationship between the true and estimated CL_{VCM} is shown in Figure S2. The IN group showed significantly lower MAPE values for $C_{\min}$, AUC, and CL_{VCM} than the OUT group ($p < 0.01$, $p < 0.01$, and $p < 0.01$, respectively), and the variability in the MPE for $C_{\min}$, AUC, and CL_{VCM} was significantly lower in the IN group than in the OUT group ($p < 0.01$, $p < 0.01$, and $p < 0.01$, respectively).

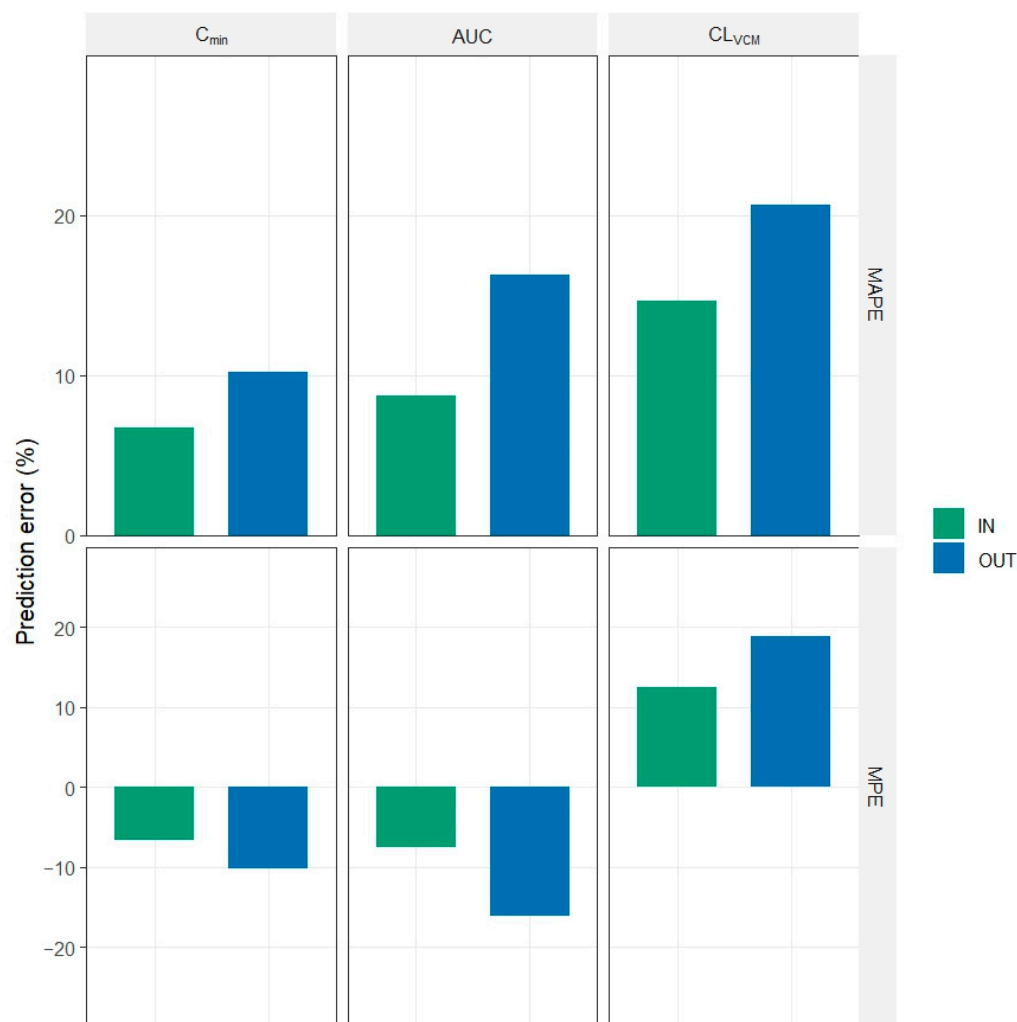


Figure 3. MAPE and MPE of Bayesian estimation of $C_{\min}$, AUC, and CL_{VCM} in IN and OUT groups. IN group, the group in which the post-optimization AUC fell within the target range of 400–600 mg·h/L; OUT group, the group in which the post-optimization AUC fell outside this range.

4. Discussion

In this study, we demonstrate the feasibility of TDM using ChatGPT based on Bayesian estimation results in a VCM simulation study. We found that setting the hyperparameter temperature to 0.1 improved the reproducibility of dosage design and that administration design using ChatGPT based on Bayesian estimation results could convert AUCs that failed to achieve target attainment into AUCs within the target range. These findings suggest that the temperature adjustment of ChatGPT is essential for developing ChatGPT-based TDM support tools that optimize personalized treatments. In addition, the use of Bayesian estimation results in a more effective administrative design.

Reproducibility is often a major concern in AI-related research, and reproducibility issues have been reported in medical testing and model-based evaluation [25]. In addition, it has been reported that reproducibility in the field of pharmacokinetic research using

ChatGPT remains a challenge in the construction of PopPK models [26]. When the Bayesian estimation results were incorporated into the prompts, the reproducibility of ChatGPT's recommended dosage and the calculated AUC, assessed using the mode percentage, varied depending on the temperature setting, with higher mode percentage values observed at $T = 0.1$ than at $T = 0.5$ and 1.0 , as shown in Figure 2. Previous medical AI research has often used binary agreement metrics to assess response reproducibility, categorizing repeated outputs as identical or non-identical [27,28]. In this study, the recommended dose was a discretized continuous variable ranging from 100 to 1500 mg in 100 mg increments. Additionally, the calculated AUC is a continuous variable with an integer value. To the best of our knowledge, the reproducibility of quantitative outputs such as dose and AUC across temperature settings has not been systematically evaluated in previous LLM studies. Therefore, we evaluated the reproducibility across temperature settings using the mode percentage, which captures the frequency of identical outputs and allows for a quantitative comparison of temperature-dependent variations. The differences in the calculation results observed across the temperature settings in this study are likely attributable to the internal calculation errors generated by ChatGPT. Lower temperature settings are known to promote more deterministic responses [18,19]. Consistent with this, ChatGPT exhibited more deterministic calculation behavior at $T = 0.1$ than at $T = 0.5$ and $T = 1.0$, resulting in higher reproducibility, as reflected by increased mode percentages for the recommended dosage and calculated AUC. Despite these differences in reproducibility, the target attainment rates of the calculated AUCs were comparable across temperature settings, and the AUC values remained within the target range of 400–600 mg·h/L. Although statistical significance was not formally tested, these differences are minimal and unlikely to be clinically meaningful. In contrast, reproducibility was clearly improved at lower temperature settings, which may be clinically relevant for ensuring consistent outputs in TDM decision support. Taken together, these findings indicate that setting the temperature to 0.1 is necessary to ensure reproducibility in the development of clinical TDM support tools.

To further interpret the modeling framework used in this study, we intentionally fixed patient covariates such as sex, age, body weight, and renal function to reduce the variability arising from known covariates, thereby enabling a clearer evaluation of the effect of ChatGPT temperature settings on the output behavior. Interindividual variability was preserved using variability parameters derived from the original PopPK model, reflecting unexplained variability beyond the included covariates. This approach enables the construction of a controlled simulation environment while maintaining a realistic level of pharmacokinetic variability. This controlled design is consistent with methodological studies aimed at evaluating model behavior under simplified conditions [29]. In addition, different population pharmacokinetic models were used for the simulation and Bayesian estimation to better reflect real-world clinical situations, where the true underlying pharmacokinetic model is unknown, and model misspecification is unavoidable [30]. To further enhance generalizability, future studies incorporating interindividual variability in patient covariates will be warranted.

We further simulated the virtual patient by administering the ChatGPT-recommended dose at $T = 0.1$ and assessed whether the post-optimization AUC values were within the target range compared with the pre-optimization AUC values. As patient characteristics and renal function were fixed in the virtual patient setting, we also simulated a guideline-based regimen of 1000 mg every 12 h. The low target achievement rate before optimization of 25.5% is consistent with previous reports that empirical or non-interventional vancomycin administration strategies achieve an AUC target of only approximately 30%, reflecting substantial inter-individual variability and the limitations of fixed dosing without TDM [31,32]. Therefore, we believe that the values obtained in this simulation are consistent with the

real-world evidence that the target achievement rate before individual dosing optimization does not reach the optimal value. Although the pre-optimization AUC value had a target attainment rate of only 25.5%, the post-optimization AUC was 71.5%, indicating that ChatGPT-guided TDM based on Bayesian estimation achieved better dose adjustment than the fixed guideline-based dose. In contrast, previous studies using Bayesian estimation software reported a target attainment rate of 62.0% [32]. Thus, the results of the present study may be considered consistent with real-world clinical performance. In other words, it was suggested that setting $T = 0.1$ in ChatGPT-based TDM using Bayesian estimation results would improve the target attainment of AUC compared to guideline-based regimens while maintaining the reproducibility of the administration design shown in Table 1. Furthermore, we investigated the factors causing the post-optimization AUC to fall outside the target range. The calculated AUC based on ChatGPT, which utilizes Bayesian-estimated clearance, achieved a target attainment rate exceeding 90%. However, when the recommended dose was applied to the post-optimization forward pharmacokinetic simulation and evaluated using the simulated AUC over 216–240 h, the target attainment rate decreased to 71.5% ($T = 0.1$). Importantly, the two AUC metrics represent fundamentally different concepts. The calculated AUC based on ChatGPT reflects a theoretical value derived directly from Bayesian-estimated pharmacokinetic parameters and deterministic calculations, whereas the post-optimization AUC reflects the realized exposure after incorporating variability and model uncertainty through simulation. Importantly, this study intentionally employed different PopPK models for the simulation and Bayesian estimation, resulting in a model mismatch by design. This model mismatch is considered the primary source of the discrepancy observed in this study, leading to a certain degree of bias in the Bayesian estimation. The MAPE and MPE of $C_{\min}$, AUC, and CL_{VCM} between the two groups revealed that this discrepancy was due to the accuracy and variability of Bayesian estimation. If the accuracy of the pharmacokinetic parameters estimated using Bayesian estimation is reduced, even if ChatGPT calculates the AUC using these parameters and determines that it is within an appropriate range, the accuracy of the calculated AUC will inevitably be reduced. This discrepancy reflects error propagation from the Bayesian parameter estimation to the forward concentration–time simulation rather than a limitation of the arithmetic execution of ChatGPT. Notably, the majority of patients in the OUT group were beyond the upper target limit, indicating a tendency toward AUC overestimation (Figure 3). This can be attributed to the overestimation of clearance by Bayesian estimation in the OUT group, which led ChatGPT to recommend higher doses to achieve an AUC within the target range. Consequently, the resulting post-optimization AUC values were likely overestimated. The accuracy of serum concentration predictions and pharmacokinetic parameter estimations using Bayesian estimation is also important when performing ChatGPT-based TDM. Thus, the observed deviation in the post-optimization AUC can be attributed to the accuracy of the Bayesian parameter estimation rather than to the computational performance of ChatGPT itself.

This study has some limitations. First, data were virtually generated under controlled conditions. This study aimed to isolate and evaluate the influence of temperature parameters on ChatGPT behavior and reproducibility, intentionally excluding variations in patient characteristics such as age, sex, dosing regimen, renal function, and volume of distribution. Although this controlled design enhances internal validity, it may limit its applicability to complex patient backgrounds and the dynamic clinical conditions encountered in real-world practice. More importantly, this study was designed as a methodological and exploratory investigation to evaluate the behavior and reproducibility of ChatGPT under controlled simulation conditions. One of the key strengths of this simulation-based approach is that the true pharmacokinetic parameters and AUC are known, allowing a

direct comparison between the true values and those estimated by Bayesian methods and ChatGPT-guided dose optimization, which is not feasible in real-world clinical settings. However, as this study was conducted under controlled simulation conditions, the findings should be interpreted with caution in terms of real-world generalizability. Future studies incorporating variability in patient characteristics, such as age, body weight, and renal function, are warranted to better reflect clinical practice. Second, because the reproducibility was assessed based on five repeated runs, the resolution of the mode-based metrics was limited. This choice reflects computational feasibility and is considered an exploratory study. Future research should incorporate more repeated runs to enable a more accurate assessment. Third, because the characteristics and renal functions of the virtual patients were fixed, the data complexity was relatively low, potentially favoring the $T = 0.1$ setting. Validation using more complex real-world clinical data, such as data from patients undergoing hemodialysis, is necessary. Dose optimization was performed at a fixed time point (day 6) after three concentration measurements were obtained, which may not reflect the variability in clinical decision timing in real-world practice. In addition, no direct comparison with the established Bayesian TDM software platform was performed. Therefore, the relative performance of this approach compared to standard-of-care methods remains unclear. Future studies should include direct comparisons with existing Bayesian TDM tools using real-world clinical data to further evaluate the behavior and feasibility of LLM-based approaches, as well as the effect of temperature.

5. Conclusions

In this study, we evaluated the behavior and reproducibility of ChatGPT in TDM based on Bayesian estimation results and demonstrated that temperature hyperparameters are crucial for maintaining reproducibility. Specifically, setting $T = 0.1$ improved reproducibility and produced stable results. This is likely due to the deterministic responses achieved by setting a lower temperature. Furthermore, ChatGPT-based dose planning based on the Bayesian estimation results improved the AUC within the target range. The results of this study highlight the importance of temperature control in ensuring reproducibility when applying ChatGPT to structured pharmacokinetic tasks. These findings should be interpreted as methodological and exploratory evaluations of LLM behavior when executing predefined pharmacokinetic calculations. Further validation using clinical data is required to determine the potential role of these approaches in clinical practice and decision-making.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/scipharm94020034/s1> Table S1: Yamamoto et al. population pharmacokinetic parameters in a healthy subject model used for pharmacokinetic analysis by Monte Carlo simulation; Table S2: Yasuhara et al. population pharmacokinetic parameters used for pharmacokinetic analysis by Bayesian estimation; Table S3: Prompt; Figure S1: Histograms of pre-optimization AUC and post-optimization AUC; Figure S2: Relationship between true and estimated CL_{VCM} .

Author Contributions: Conceptualization, A.K. and T.A.; Methodology, A.K.; Software, A.K. and T.A.; Validation, A.K. and T.A.; Formal analysis, A.K.; Investigation, A.K. and T.A.; Data curation, A.K.; Writing—original draft preparation, A.K. and T.A.; Writing—review and editing, A.K., T.A., R.M., D.H., T.K. and Y.T.; Supervision, Y.T.; Project administration, Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI (grant number JP24K09969).

Institutional Review Board Statement: Not applicable. The simulation study did not involve human participants or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: No real-world patient data were used in this study. All data were synthetically generated using a Monte Carlo simulation implemented in R version 4.5.1. The full prompt templates, model specifications, resulting dosing recommendations, and calculated AUC values used in the ChatGPT interactions are provided in the Supplementary Materials (Table S3), enabling independent replication of the AI-assisted workflow.

Acknowledgments: In preparing this manuscript and conducting this study, we used the OpenAI API (gpt-4o-mini) for programming and simulation-based dose and AUC calculations. The authors reviewed and edited all AI-generated outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
LLM	Large language model
TDM	Therapeutic drug monitoring
VCM	Vancomycin
AUC	Area under the serum concentration–time curve
$C_{\min}$	Trough concentration
PopPK	Population pharmacokinetic
SCr	Serum creatinine
V_1	Central compartment volume of distribution
V_2	Peripheral compartment volume of distribution
Q	Distribution clearance
CL_{VCM}	Total clearance
k_{12}	Rate constant for transfer from the central to the peripheral compartment
k_{21}	Rate constant for transfer from the peripheral to the central compartment
k_e	Elimination rate constant
$C_{\min 48h}$	Trough concentration at 48 h after the start of vancomycin administration
$C_{\min 96h}$	Trough concentration at 96 h after the start of vancomycin administration
$C_{\min 144h}$	Trough concentration at 144 h after the start of vancomycin administration
V_{ss}	Steady-state volume of distribution
T	Temperature
MAPE	Mean absolute prediction error
MIC	Minimum inhibitory concentration
MPE	Mean prediction error

References

- Bohr, A.; Memarzadeh, K. The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in Healthcare*; Bohr, A., Memarzadeh, K., Eds.; Elsevier: Cambridge, MA, USA, 2020; pp. 25–60. [CrossRef]
- OpenAI. ChatGPT: Optimizing Language Models for Dialogue. Available online: <https://openai.com/index/chatgpt/> (accessed on 2 March 2026).
- Alawida, M.; Mejri, S.; Mehmood, A.; Chikhaoui, B.; Isaac Abiodun, O. A comprehensive study of ChatGPT: Advancements, limitations, and ethical considerations in natural language processing and cybersecurity. *Information* **2023**, *14*, 462. [CrossRef]
- Lima, T.D.M.; Bonafé, M.; Baby, A.R.; Visacri, M.B. ChatGPT in pharmacy practice: Disruptive or destructive innovation? A scoping review. *Sci. Pharm.* **2024**, *92*, 58. [CrossRef]
- Carou-Senra, P.; Delgado-Taboada, I.; Alvarez-Lorenzo, C.; Diaz-Rodriguez, P.; Goyanes, A. Exploring the role of LLMs like ChatGPT in pharmacy education for supporting students' therapeutic decision-making. *Am. J. Pharm. Educ.* **2025**, *89*, 101462. [CrossRef] [PubMed]
- Liu, C.; Bayer, A.; Cosgrove, S.E.; Daum, R.S.; Fridkin, S.K.; Gorwitz, R.J.; Kaplan, S.L.; Karchmer, A.W.; Levine, D.P.; Murray, B.E.; et al. Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children: Executive summary. *Clin. Infect. Dis.* **2011**, *52*, 285–292. [CrossRef]

7. Rybak, M.J.; Lomaestro, B.M.; Rotschafer, J.C.; Moellering, R.C.; Craig, W.A.; Billeter, M.; Dalovisio, J.R.; Levine, D.P. Vancomycin Therapeutic Guidelines: A summary of consensus recommendations from the Infectious Diseases Society of America, the American Society of Health System Pharmacists, and the Society of Infectious Diseases Pharmacists. *Clin. Infect. Dis.* **2009**, *49*, 325–327. [\[CrossRef\]](#)
8. Johannsson, B.; Beekmann, S.E.; Srinivasan, A.; Hersh, A.L.; Laxminarayan, R.; Polgreen, P.M. Improving antimicrobial stewardship: The evolution of programmatic strategies and barriers. *Infect. Control Hosp. Epidemiol.* **2011**, *32*, 367–374. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Giuliano, C.; Haase, K.K.; Hall, R. Use of vancomycin pharmacokinetic–pharmacodynamic properties in the treatment of MRSA infections. *Expert Rev. Anti-Infect. Ther.* **2010**, *8*, 95–106. [\[CrossRef\]](#)
10. Filippone, E.J.; Kraft, W.K.; Farber, J.L. The nephrotoxicity of vancomycin. *Clin. Pharmacol. Ther.* **2017**, *102*, 459–469. [\[CrossRef\]](#)
11. Rybak, M.J.; Le, J.; Lodise, T.P.; Levine, D.P.; Bradley, J.S.; Liu, C.; Mueller, B.A.; Pai, M.P.; Wong-Beringer, A.; Rotschafer, J.C.; et al. Therapeutic monitoring of vancomycin for serious methicillin-resistant *Staphylococcus aureus* infections: A revised consensus guideline and review by the American Society of Health System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the Society of Infectious Diseases Pharmacists. *Am. J. Health Syst. Pharm.* **2020**, *77*, 835–864. [\[CrossRef\]](#)
12. Matsumoto, K.; Oda, K.; Shoji, K.; Hanai, Y.; Takahashi, Y.; Fujii, S.; Hamada, Y.; Kimura, T.; Mayumi, T.; Ueda, T.; et al. Clinical practice guidelines for therapeutic drug monitoring of vancomycin in the framework of model-informed precision dosing: A consensus review by the Japanese society of chemotherapy and the Japanese society of therapeutic drug monitoring. *Pharmaceutics* **2022**, *14*, 489. [\[CrossRef\]](#)
13. Tsutsuura, M.; Moriyama, H.; Kojima, N.; Mizukami, Y.; Tashiro, S.; Osa, S.; Enoki, Y.; Taguchi, K.; Oda, K.; Fujii, S.; et al. The monitoring of vancomycin: A systematic review and meta-analyses of area under the concentration-time curve-guided dosing and trough-guided dosing. *BMC Infect. Dis.* **2021**, *21*, 153. [\[CrossRef\]](#)
14. Oda, K.; Jono, H.; Nosaka, K.; Saito, H. Reduced nephrotoxicity with vancomycin therapeutic drug monitoring guided by area under the concentration–time curve against a trough 15–20 µg/mL concentration. *Int. J. Antimicrob. Agents* **2020**, *56*, 106109. [\[CrossRef\]](#)
15. Aljefri, D.M.; Avedissian, S.N.; Rhodes, N.J.; Postelnick, M.J.; Nguyen, K.; Scheetz, M.H. Vancomycin area under the curve and acute kidney injury: A meta-analysis. *Clin. Infect. Dis.* **2019**, *69*, 1881–1887. [\[CrossRef\]](#)
16. Abdelmessih, E.; Patel, N.; Vekaria, J.; Crovetto, B.; SanFilippo, S.; Adams, C.; Brunetti, L. Vancomycin area under the curve versus Trough Only Guided Dosing and the Risk of Acute Kidney Injury: Systematic Review and Meta-analysis. *Pharmacotherapy* **2022**, *42*, 741–753. [\[CrossRef\]](#)
17. Neely, M.N.; Youn, G.; Jones, B.; Jelliffe, R.W.; Drusano, G.L.; Rodvold, K.A.; Lodise, T.P. Are vancomycin trough concentrations adequate for optimal dosing? *Antimicrob. Agents Chemother.* **2014**, *58*, 309–316. [\[CrossRef\]](#)
18. OpenAI. Best Practices for Prompt Engineering with OpenAI API. Available online: <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api> (accessed on 2 March 2026).
19. Davis, J.; Van Bulck, L.; Durieux, B.N.; Lindvall, C. The temperature feature of ChatGPT: Modifying creativity for clinical research. *JMIR Hum. Factors* **2024**, *11*, e53559. [\[CrossRef\]](#)
20. OpenAI. API Reference. Available online: <https://developers.openai.com/api/reference/overview> (accessed on 2 March 2026).
21. Cockcroft, D.W.; Gault, M.H. Prediction of creatinine clearance from serum creatinine. *Nephron* **1976**, *16*, 31–41. [\[CrossRef\]](#)
22. Yamamoto, M.; Kuzuya, T.; Baba, H.; Yamada, K.; Nabeshima, T. Population pharmacokinetic analysis of vancomycin in patients with Gram-positive infections and the influence of infectious disease type. *J. Clin. Pharm. Ther.* **2009**, *34*, 473–483. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Yasuhara, M.; Iga, T.; Zenda, H.; Okumura, K.; Oguma, T.; Yano, Y.; Hori, R. Population pharmacokinetics of vancomycin in Japanese adult patients. *Ther. Drug Monit.* **1998**, *20*, 139–148. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Nakashima, M.; Katagiri, K.; Oguma, T. Phase I studies on vancomycin hydrochloride for infection. *Chemotherapy* **1992**, *40*, 210–224. [\[CrossRef\]](#)
25. Beam, A.L.; Manrai, A.K.; Ghassemi, M. Challenges to the reproducibility of machine learning models in health care. *JAMA* **2020**, *323*, 305–306. [\[CrossRef\]](#)
26. Cloesmeijer, M.E.; Janssen, A.; Koopman, S.F.; Cnossen, M.H.; Mathôt, R.A.A.; SYMPHONY Consortium. ChatGPT in pharmacometrics? Potential opportunities and limitations. *Br. J. Clin. Pharmacol.* **2024**, *90*, 360–365. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Wang, L.; Chen, X.; Deng, X.; Wen, H.; You, M.; Liu, W.; Li, Q.; Li, J. Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *npj Digit. Med.* **2024**, *7*, 41. [\[CrossRef\]](#)
28. Haze, T.; Kawano, R.; Takase, H.; Suzuki, S.; Hirawa, N.; Tamura, K. Influence on the accuracy in ChatGPT: Differences in the amount of information per medical field. *Int. J. Med. Inform.* **2023**, *180*, 105283. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Augustin, D.; Lambert, B.; Robinson, M.; Wang, K.; Gavaghan, D. Simulating clinical trials for model-informed precision dosing: Using warfarin treatment as a use case. *Front. Pharmacol.* **2023**, *14*, 1270443. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Guhl, M.; Mercier, F.; Hofmann, C.; Sharan, S.; Donnelly, M.; Feng, K.; Sun, W.; Sun, G.; Grosser, S.; Zhao, L.; et al. Impact of model misspecification on model-based tests in pharmacokinetic studies with parallel design: Real case and simulation studies. *J. Pharmacokinet. Pharmacodyn.* **2022**, *49*, 557–577. [[CrossRef](#)]
31. Flannery, A.H.; Delozier, N.L.; Effoe, S.A.; Wallace, K.L.; Cook, A.M.; Burgess, D.S. First-dose vancomycin pharmacokinetics versus empiric dosing on area-under-the-curve target attainment in critically ill patients. *Pharmacotherapy* **2020**, *40*, 1210–1218. [[CrossRef](#)]
32. Okada, N.; Yoshii, T.; Tamura, M.; Baba, A.; Matsudomi, K.; Sakiyama, T.; Goto, R.; Kitahara, T. Clinical feasibility of nomogram-based vancomycin dosing strategy: A retrospective study. *Biol. Pharm. Bull.* **2026**, *49*, 422–428. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.