# ARTICLE

# Receiver Operating Characteristic Analysis and Clinical Trial Simulation to Inform Dose Titration Decisions

John David Clements[1],*, Juan Jose Perez Ruixo[2], John P. Gibbs[3], Sameer Doshi[1], Carlos Perez Ruixo[2] and Murad Melhem[4]

Optimal dose selection in clinical trials is problematic when efficacious and toxic concentrations are close. A novel quantitative approach follows for optimizing dose titration in clinical trials. A system of pharmacokinetics (PK), pharmacodynamics, efficacy, and toxicity was simulated for scenarios characterized by varying degrees of different types of variability. Receiver operating characteristic (ROC) and clinical trial simulation (CTS) were used to optimize drug titration by maximizing efficacy/safety. The scenarios included were a low-variability base scenario, and high residual (20%), interoccasion (20%), interindividual (40%), and residual plus interindividual variability scenarios, and finally a shallow toxicity slope scenario. The percentage of subjects having toxicity was reduced by 87.4% to 93.5%, and those having efficacy was increased by 52.7% to 243%. Interindividual PK variability may have less impact on optimal cutoff values than other sources of variability. ROC/CTS methods for optimizing dose titration offer an individualized approach that leverages exposure-response relationships.
*CPT Pharmacometrics Syst. Pharmacol.* (2018) **7**, 771–779; doi:10.1002/psp4.12354; published online on 15 October 2018.

## Study Highlights

**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**
☑ Utility indices, or other approaches, combined with simulations may be used to assess which dose regimens to include in future clinical studies.
**WHAT QUESTION DID THIS STUDY ADDRESS?**
☑ Provides quantitative methodologies for defining dosing decisions for the dose titration of narrow therapeutic index compounds.
**WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?**
☑ The ROC and CTS analyses described within define dosing algorithms that minimize toxicity and maximize

efficacy in clinical development under different variability scenarios.
**HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?**
☑ The approach combines safety and efficacy criteria with subject-level drug exposure to establish a quantitatively justified and individualized dose titration scheme. This potentially enables the exploration, at all stages of drug development, of dose levels that may otherwise be considered unfeasible due to safety concerns.

In drug development, optimal dosing regimens of new or potential therapies are designed to maximize their clinical benefit in patients while minimizing drug-related toxicities. Selecting an optimal dosing regimen for new molecular entities is pivotal in the demonstration of positive benefit-to-risk balance, and, in turn, to drug approval. For drugs having a wide therapeutic margin, a single available dose (flat dosing) may be safe and effective even without regard to dose adjustments in special populations despite pharmacokinetic (PK) variability resulting from intrinsic and extrinsic factors.[1–4] However, for drugs with narrow therapeutic indices, flat dosing generally does not provide the optimal benefit-to-risk ratio across subjects in the target population. In such cases, dose titration algorithms may be needed to achieve a desirable benefit-to-risk balance.

Dose titration algorithms aim to make rational individual dose adjustment decisions and are generally based on PK or pharmacodynamic (PD) measurements. For many years, therapeutic drug monitoring and approaches based on achieving a target exposure range have been widely used for PK-based dose individualization of drugs with narrow therapeutic indices.[5–12] These approaches mitigate the risk of toxicity that is associated with excessive PK concentrations or hyperpharmacology.

PD-guided dose titration algorithms are based on monitoring biomarker(s)[13–17] related to mechanism or clinical

772

signs and/or symptoms.[18] In theory, PD-guided approaches offer an alternative to PK-guided approaches if it is known that PD, and not PK, is a better predictor of efficacy and/or toxicity. It is drug-specific as to whether PD or a combination of PK and PD is the most efficient alternative to PK-alone based algorithms for dose individualization and titration regimens. It is also acknowledged that logistic and operational constraints in monitoring drug responses may limit the efficiency of PD-guided dose titration algorithm. Whether using PK, PD, or PK/PD-guided dose titration approaches, processes to optimize regimens are conceptually similar.

Dose titration algorithms, where the maintenance dose regimen is individualized based on PK end points, such as maximum concentration ($C_{max}$), area under the curve (AUC), or trough concentrations ($C_{trough}$), are often empirically determined in drug development, and may be implemented in clinical studies without prior (e.g., *in silico*) assessment of their ability to maximize efficacy and minimize toxicity. The lack of standard methods warrants the development of scientifically based quantitative approaches to defining dose titration algorithms.

The receiver operating characteristic (ROC) analysis is a well-accepted tool for diagnostic development[19,20] and has been used in predicting efficacy from a given drug concentration, which in turn may be advantageous for increasing the clinical benefit of patients or facilitating first-in-human study design.[21,22] In the past 15 years, the use of clinical trial simulation (CTS) in evaluating the selection of dosing regimens that maximize efficacy and minimize toxicity risk has been well documented in the literature.[18,23,24] However, there is scarcity of information on utilizing CTS to assess the efficiency of dose titration schemes.[25,26]

In this analysis, we propose the use of a combination of ROC and CTS techniques as an objective and quantitative approach to optimize dose titration algorithms. In addition, these approaches are a way of assessing potential benefits and risks prior to implementation in clinical trials. To the best of our knowledge, there are no literature reports exploring and demonstrating the use of ROC and CTS analyses for optimizing dosing of narrow therapeutic index compounds under development. The methods and results described provide supportive evidence of the generalizability of the proposed quantitative approach.

## METHODS
### Motivating example
The provided example uses virtual subjects having linked PK, PD, efficacy, and toxicity models that are representative of those found in drug development. A hypothetical molecule (compound A) at a dose of 5 mg b.i.d. had suboptimal efficacy, although the toxicity profile was acceptable. At a dose of 10 mg b.i.d., efficacy was favorable but the incidence of toxicity was unacceptable, thereby limiting the dose to 5 mg. If a 10 mg b.i.d. dose level was considered too risky, could a titration schema be devised to maximize efficacy and minimize toxicity? Descriptions of the models, methods, and scenarios for compound A are provided below, followed by a description of the model-based simulation techniques that were used to define and assess dose titration algorithms.

### Population PK/PD, efficacy, and toxicity models
The values of the population PK/PD models' parameters used for the different compound A simulations and scenarios are reported in **Table 1**. Compound A was simulated to exhibit biphasic disposition with linear elimination and first-order absorption. The PK model parameters were log-normally distributed, in which both interindividual variability (IIV) and interoccasion variability (IOV) were expressed using an exponential error structure. The residual unexplained error (RUV) model was proportional. Five PK-related scenarios were created using different degrees of variability from different sources. The first scenario was the base scenario (BaseScenario), which included low levels of RUV, IOV, and IIV. Each of the high RUV (HighRUV), high IOV (HighIOV), high IIV (HighIIV), and combined high RUV plus IIV (HighRUVIIV) scenarios had additional high variability, compared to BaseScenario, derived from high RUV, IOV, IIV, and RUV plus IIV, respectively. The PK/PD relationship was described using a sigmoidal maximum effect ($E_{max}$) model, in which IIV on half-maximal effective concentration was expressed using an exponential error structure. The residual error model was additive. To serve as an example, a control file for the PK/PD relationship of BaseScenario is available in the Supplementary Materials file (**Supplementary Materials S1**).

Based on the PD response ($PD_{Response}$), the probability of experiencing efficacy, p(Efficacy), was characterized using a logistic regression model. The logit of p(Efficacy) was linearly related to $PD_{Response}$ expressed as a percent of $E_{max}$, as follows:

$$\text{logit}\, p(Efficacy) = \alpha + \beta \times PD_{Response} \tag{1}$$

where $\alpha$ and $\beta$ are the intercept and the slope of the linear relationship that links the logit p(Efficacy) and $PD_{Response}$. **Figure S1a** characterizes the relationship, where there was an increasing probability from 0–100% of having efficacy as $PD_{Response}$ increases from 35–55%. A 50% probability of efficacy corresponded to 45% $PD_{Response}$.

The probability of experiencing toxicity, p(AE), was also characterized using a logistic regression model. The logit of p(AE) was linearly related to compound A plasma concentrations ($C_p$), as follows:

$$\text{logit}\, p(AE) = \alpha + \beta \times C_p \tag{2}$$

where $\alpha$ and $\beta$ are the intercept and the slope of the linear relationship that links the logit p(AE) and $C_p$. The steepness of the relationship was varied for different scenarios. The BaseScenario suggests a narrow concentration range (steep slope) within which the probability of an event increases from 0% to 100%. For the sixth scenario, a shallow slope was utilized for the probability of toxicity (AEshallow), where the probability of toxicity increased from 0% to 100% over a broader range of concentrations. The increase in the probability of toxicity for AEshallow occurred between 200 and 300 ng/mL, whereas for the other scenarios the increase in the probability occurred between 240 and 260 ng/mL. For both scenarios, the 50% probability of toxicity corresponded to a concentration of 250 ng/mL.

**Table 1** Population pharmacokinetic, pharmacodynamic, efficacy, and toxicity models' parameters used in the simulations for the various scenarios. HighRUVIIV combines the variability specifications of HighRUV and HighIIV

| | Parameter values | | |
| --- | --- | --- | --- |
| | **Typical value** | **IIV (%)** | **IOV (%)** |
| **Pharmacokinetic parameters** | | | |
| Systemic clearance: CL/F (L/h) | 6 | 20 (HighIIV: 40) | 0 (HighIOV: 20) |
| Central volume of distribution: V2/F (L) | 40 | 20 (HighIIV: 40) | 0 (HighIOV: 20) |
| Intercompartmental clearance: Q/F (L/h) | 30 | – | – |
| Peripheral volume: V3/F(L) | 120 | – | – |
| Absorption rate constant: $k_a$ ($h^{-1}$) | 0.6 | 20 (HighIIV: 40) | 0 (HighIOV: 20) |
| Residual variability (proportional) % | 10 (HighRUV: 20) | – | – |
| **Pharmacodynamic model parameters** | | | |
| Maximum drug effect: $E_{max}$ (%) | 100 | – | – |
| Potency: $EC_{50}$ (ng/mL) | 80 | 20 | – |
| Hill coefficient: $\gamma$ | 1.5 | – | – |
| Residual variability (additive): (%) | 8 | – | – |
| **Efficacy model parameters** | | | |
| Intercept: $\alpha$ | –22.5 | – | – |
| Slope: $\beta$ | 0.5 | – | – |
| **Safety model parameters** | | | |
| Intercept: $\alpha$ | –80 (AEshallow: –20) | – | – |
| Slope: $\beta$ | 0.32 (AEshallow: 0.08) | – | – |

$EC_{50}$, half-maximal effective concentration; $E_{max}$, maximum effect; IIV, interindividual variability; IOV, interoccasion variability; RUV, residual unexplained variability.

## Defining the dose titration algorithm: ROC analysis

It was assumed that compound A could be administered to subjects based on an upward dose titration scheme, which consisted of (i) initiating subjects on 5 mg b.i.d. for 10 consecutive days (time to reach steady-state concentrations), (ii) assessing $C_{trough}$ on day 10, and (iii) on the basis of the $C_{trough}$ value, deciding whether a 5 or 10 mg b.i.d. dose would be administered for the final, additional 20 days of the treatment period. A subject titrated from the initial 5 mg b.i.d. dose level to the 10 mg b.i.d. dose level only if their day 10 $C_{trough}$ value was below the optimal cutoff value.

Drug concentrations on day 10 were simulated for 10,000 virtual subjects in each scenario; all receiving 5 mg b.i.d. Then, each virtual subject received 5 mg or 10 mg b.i.d. dosing for an additional 20 days and, on day 30, toxicity and efficacy was assessed based on simulated PK/PD data. The sampling density on day 30 was such that $C_{max}$ would be captured, with samples at predose and at 0.5, 1, 1.5, 2, 2.5, 3, 4, 6, 8, and 12 hours postdose.

For each possible $C_{trough}$ cutoff value, true positive (TP) and false negative (FN) classification results were recorded when a virtual subject experienced toxicity on day 30 and they had a $C_{trough}$ that was above or below the $C_{trough}$ cutoff value on day 10. The FP and true-negative (TN) classification results were recorded when a virtual subject did not experience toxicity on day 30 and they had a $C_{trough}$ that was above or below the $C_{trough}$ cutoff value on day 10. For each $C_{trough}$ evaluated, sensitivity and 1-specificity was calculated and plotted to create an ROC curve: on the *y*-axis is sensitivity, which is calculated as follows:

$$\text{Sensitivity} = \text{TP/(TP+FN)}, \qquad (3)$$

and on the *x*-axis is 1-specificity; where specificity is calculated as follows:

$$\text{Specificity} = \text{TN/(FP+TN)} \qquad (4)$$

Predictive performance for the use of $C_{trough}$ on day 10 as a predictor of day 30 drug-related toxicity was assessed using the area under the receiver operating characteristic (AUCROC) and 95% confidence interval (CI) as calculated by the Wilcoxon *U*-Statistic, where an AUCROC value of 0.5 indicates no apparent accuracy (i.e., random chance) as a value of 1.0 represents 100% accuracy.[27] The 95% CI values for sensitivity and specificity were calculated using the Wilson exact approach. The AUCROC was calculated by the linear trapezoidal rule. The positive predictive value (PPV) was calculated as follows:

$$\text{PPV} = \text{TP/(TP + FP)} \qquad (5)$$

The negative predictive value (NPV) was calculated as follows:

$$\text{NPV} = \text{TN/(TN + FN)} \qquad (6)$$

An efficiency criterion was used to define the optimal threshold (cutoff value) for dose escalation in an individual subject. Efficiency for each potential cutoff value was calculated as follows[28]:

$$\text{Efficiency} = \text{Sensitivity-Cost} * (100 - \text{Specificity}) \qquad (7)$$

where, sensitivity and specificity were expressed as percentages, and the overall cost of misclassification was calculated as indicated in Eq. 8 below:

$$Cost = (FPC)/(FNC) \times (1 - Inc)/Inc \quad (8)$$

where *Inc* was the proportion of subject having drug-related toxicity on day 30, and the ratio of false-positive cost (FPC), to false-negative cost (FNC), represents the relative cost of misclassification (i.e., false-positive (FP) and FN). Here, the FNC was arbitrarily set to 10 to reflect a higher relative importance of wrongly predicting that a subject would not have toxicity when in fact they would. The FPC was fixed at 1. The $C_{trough}$ measurement on day 10 associated with the maximal efficiency can be interpreted as the optimal cutoff concentration, where the proportion of FP and FN classifications are minimized based on their relative cost. The cost in Eq. 3 becomes a slope function within what is known as the Youden's J statistic, a tool used to assess the performance of a diagnostic test for a given cutoff value as defined in Eq. 9 below:[29]

$$J = Sensitivity + Specificity - 100 \quad (9)$$

When *J* equals 100, the maximum possible value for the statistic, perfect classification has been achieved, and when *J* equals 0, the lowest possible value for the statistic, there is no predictive value of the selected cutoff point.[29]

A further examination to determine the effects of varying the cost of FN and FP classifications will be made so as to determine the effect on the optimal cutoff value and PPVs and NPVs.

### Assessing the utility of the dose titration algorithm: clinical trial simulation analysis

To further assess the selection of potential cutoff values, CTS were conducted for various cutoff values. The incidence of toxicity, efficacy, and those on the higher dose level were summarized. For comparison to the ROC approach, the utility analysis of each scenario, relative toxicity, and efficacy were calculated as follows:

$$ToxicityUtility = 100 - \left( \frac{Inc_{AE,Cutoff} - Inc_{AE,LowDose}}{Inc_{AE,HighDose} - Inc_{AE,LowDose}} \right) \times 100 \quad (10)$$

$$EfficacyUtility = \frac{Inc_{Efficacy,Cutoff} - Inc_{Efficacy,LowDose}}{Inc_{Efficacy,HighDose} - Inc_{Efficacy,LowDose}} \times 100 \quad (11)$$

$$UtilityIndex = ToxicityUtility + EfficacyUtility \quad (12)$$

where, $Inc_{AE,Cutoff}$ is the incidence of toxicity associated with each specific cutoff, $Inc_{AE,LowDose}$ and $Inc_{AE,HighDose}$ are the incidence of toxicity associated with the 5 and 10 mg doses, $Inc_{Efficacy,Cutoff}$ is the incidence of efficacy associated with each specific cutoff, $Inc_{Efficacy,LowDose}$ and $Inc_{Efficacy,HighDose}$ are the incidence of subjects exceeding the PD target associated with the low and high doses. UtilityIndex was then calculated as the sum of the toxicity and efficacy utility values.

### Software

The PK/PD model-based simulations were conducted using mixed-effects methods as implemented in NONMEM version 7.2 software (ICON Development Solutions, Ellicott City, MD)[30,31] on a pseudo-cluster of Intel Xeon CPU X5660@ 2.80 GHz processors and an Intel Fortran compiler (12.1). Dataset preparation, efficacy and toxicity calculations, graphical representations, and statistical analyses, including ROC analyses, were performed using R version 3.4 or higher.[32]

## RESULTS

### Drug exposures

The scenarios' distributions of high-dose simulated subject-level minimum plasma concentration ($C_{min}$) and $C_{max}$ values, without a titration scheme applied, are shown in **Figure 1**, with subjects exceeding the upper threshold of 250 ng/mL residing above the red dashed line. The incidence of subjects going above the safety threshold on the high dose level, with no titration algorithm applied, is reported in the second row of **Table 2**.

### Optimal $C_{trough}$ threshold for dose escalation

The results from the ROC analyses of different scenarios are found in **Figure 2** and **Figures S2–S6**. With respect to BaseScenario, **Figure 2a** demonstrates the overlap in $C_{trough}$ distributions between TP and TN populations. This highlights the difficulty in separating the groups. In **Figure 2b**, the $C_{trough}$ values associated with the half-maximum cumulative probabilities of TP and TN indicate what the median $C_{trough}$ is for each group. **Figure 2c** depicts the ROC curve. The calculated AUCROC was 92.9% (95% CI: 91.8–94.1). The maximum value of the efficiency analysis (**Figure 2d**) indicated that a day 10 $C_{trough}$ cutoff value of 62.1 ng/mL was the optimal balance for FP and FN classifications. **Figures S2–S6** for HighRUV, HighIOV, HighIIV, HighRUVIIV, and HighAEshallow follow similar logic.

Summary statistics and diagnostics for the six scenarios from CTS and ROC analyses are presented in **Table 2**. For BaseScenario, the predicted incidence of toxicity without any titration step was 9.41%. Predicted incidence was higher in the other scenarios due to higher PK variability or steepness in the toxicity response to compound A concentration. The PPV, NPV, sensitivity, and specificity were lower for HighRUV, HighIOV, and AEshallow when compared to HighIIV. The Youden's J and AUCROC statistics were 70% and 92.9% for BaseScenario and 74.7% and 96.1% for HighIIV. Compare this to the lower Youden's J and AUCROC statistics of 34.8% and 81.1% for HighRUV, 25.5% and 72.8% for HighIOV, and 57.3% and 88.1% for AEshallow. Both the AUCROC and the Youden J statistic values for BaseScenario, HighIIV, HighRUV (exception for Youden J statistic on HighRUV), and AEshallow indicate that prognostic properties for this "diagnostic" test were relatively good for determining an optimal safety threshold concentration. For HighIOV the relatively lower Youden's J statistic and the AUCROC value indicate relatively worse prognostic ability.

Compared to optimal cutoff values of 62.1, 62, and 57.5 ng/mL for BaseScenario, HighIIV, and AEshallow, both HighRUV and HighIOV had lower cutoff values of 45.8 and 44.2 ng/mL, respectively (**Table 2**). BaseScenario, HighIIV, and AEshallow had 71%, 60.9%, and 60.3% of subjects on the higher dose level vs. 33.1% and 30.8% for HighRUV and
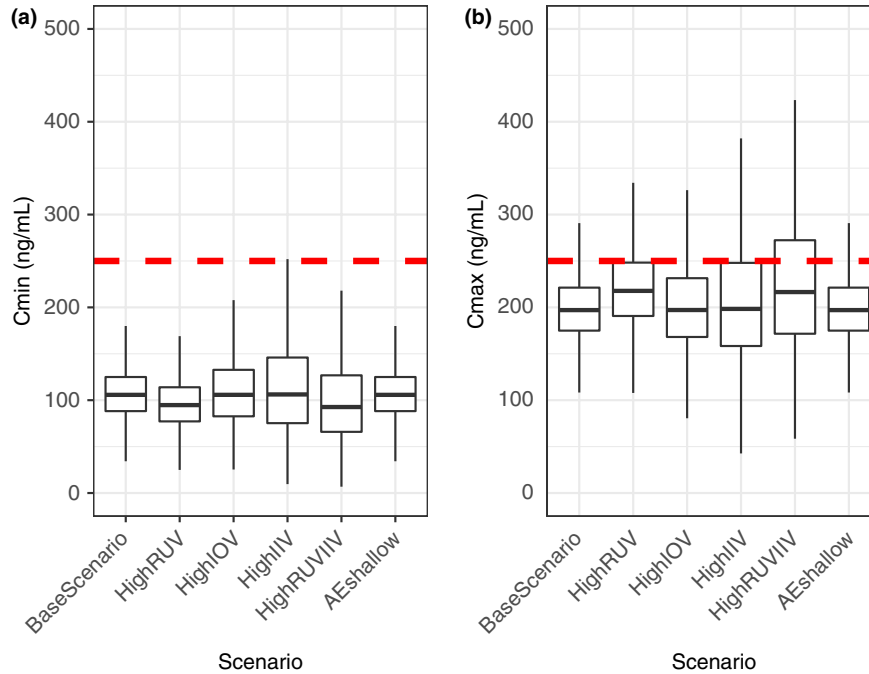
**Figure 1** High dose minimum plasma concentration ($C_{min}$) (**a**) and maximum plasma concentration ($C_{max}$) (**b**) distributions for the six scenarios without any titration scheme applied, showing in the high pharmacokinetic variability scenarios that $C_{min}$ is lower and $C_{max}$ is higher when compared with the other scenarios. (The red dashed line represents the 50% probability of toxicity threshold (250 ng/mL)). AE, adverse event; IIV, interindividual variability; IOV, interoccasion variability; RUV, residual unexplained variability.

**Table 2** Summary of diagnostics for the optimal (predose) $C_{trough}$ cutoff value at steady state for a relative false negative to false positive classification cost ratio of 10

| | BaseScenario | HighRUV | HighIOV | HighIIV | HighRUVIIV | AEshallow |
|---|---|---|---|---|---|---|
| Simulations | 10,000 | 10,000 | 10,000 | 10,000 | 10,000 | 10,000 |
| Incidence (%) | 9.41 | 24.4 | 16.6 | 24.8 | 34.6 | 16.4 |
| Cutoff (ng/mL) | 62.1 | 45.8 | 44.2 | 62 | 44 | 57.5 |
| PPV (%) | 29.9 | 33.9 | 21.7 | 60.5 | 53.3 | 36.1 |
| NPV (%) | 99 | 95 | 94.8 | 98 | 96.3 | 96.6 |
| Sensitivity (%) | 92.6 (90.7–94.1) | 93.2 (92.1–94.1) | 90.4 (88.9–91.7) | 95.2 (94.3–96) | 96 (95.3–96.6) | 87.7 (86–89.2) |
| Specificity (%) | 77.4 (76.5–78.2) | 41.6 (40.5–42.7) | 35.1 (34.1–36.1) | 79.5 (78.6–80.4) | 55.4 (54.2–56.6) | 69.6 (68.6–70.6) |
| AUCROC (%) | 92.9 (91.8–94.1) | 81.1 (80–82.2) | 72.8 (71.3–74.2) | 96.1 (95.6–96.7) | 91 (90.3–91.6) | 88.1 (87–89.2) |
| Youden's J stat (%) | 70 | 34.8 | 25.5 | 74.7 | 51.4 | 57.3 |
| % With toxicity | 0.73 | 1.66 | 1.59 | 1.6 | 2.72 | 2.06 |
| % With efficacy | 75.2 | 45.1 | 44.5 | 66.3 | 47.5 | 67.1 |
| % Subjects on 10 mg | 71 | 33.1 | 30.8 | 60.9 | 37.7 | 60.3 |
| Toxicity utility | 92.2 | 93.2 | 90.5 | 95.1 | 95.8 | 87.5 |
| Efficacy utility | 79.4 | 34.5 | 31.7 | 75.2 | 35.6 | 67.3 |
| Utility index | 172 | 128 | 122 | 170 | 131 | 155 |

Using the ROC methodologies the percentage of subjects having toxicity is reduced and those having efficacy is increased. This is reflected in the percentage of subjects that achieve a 10 mg dose level, which ultimately increases the group-wise drug exposure. Scenarios having high RUV or high IOV had the lowest percentage of subjects being titrated to the 10 mg dose level.
AE, adverse event; AUCROC, area under the receiver operating characteristic; IIV, interindividual variability; IOV, interoccasion variability; NPV, negative predictive value; PPV, positive predictive value; RUV, residual unexplained variability.

HighIOV, respectively (**Table 2**). In addition, BaseScenario, HighIIV, and AEshallow had 75.2%, 66.3%, and 67.1% of subjects achieving efficacy vs. 45.1% and 44.5% for HighRUV and HighIOV, respectively. **Figure 3a** shows the percent of subjects achieving efficacy or toxicity, and **Figure 3b** shows the percentage change in subjects achieving efficacy or toxicity compared to a fixed low-dose strategy and the percentage change in subjects having toxicity
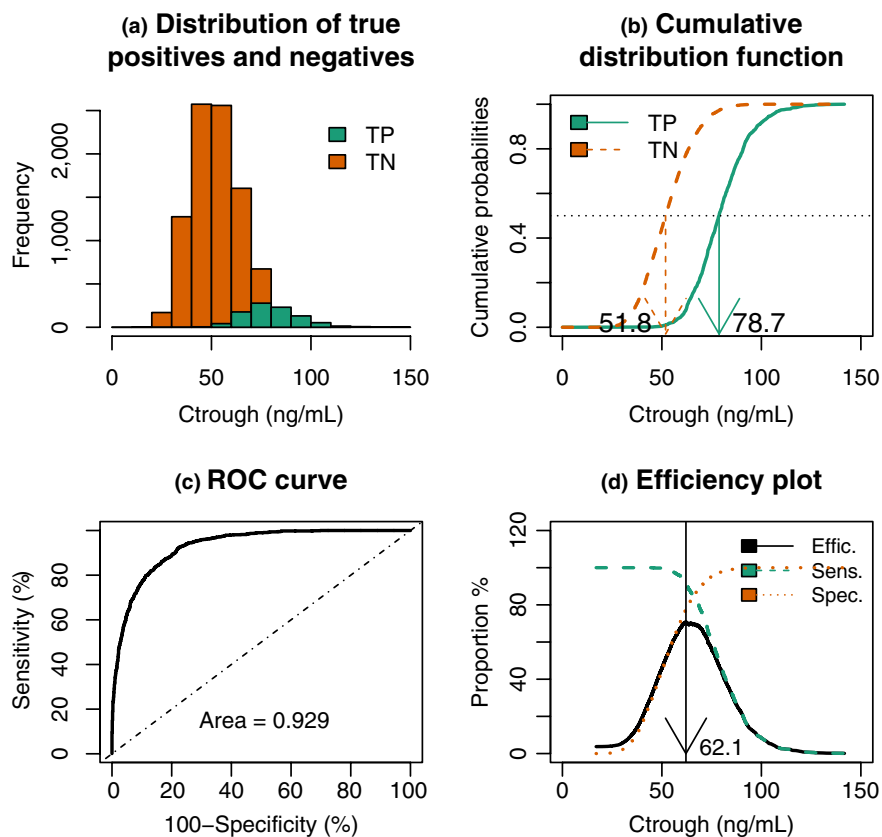
### (a) Distribution of true positives and negatives

### (b) Cumulative distribution function

### (c) ROC curve

### (d) Efficiency plot

**Figure 2** Receiver operating characteristic (ROC) and efficiency analysis for BaseScenario and a relative false-negative cost (FNC) of 10. (**a**) Probability histogram that shows the overlapping trough plasma concentration ($C_{trough}$) distributions of true-positive (TP) and true-negative subjects (TN) and their incidence. (**b**) Cumulative density function curves for TP and TN virtual subjects. (**c**) ROC curve for the ability to use 5 mg b.i.d. day 10 $C_{trough}$ as a prognostic indicator of whether a virtual subject will exceed the 250 ng/mL threshold if they were titrated up to a 10 mg b.i.d. dose. (**d**) An efficiency plot that incorporates the ratio of FNC to FPC costs for determining an optimal $C_{trough}$ cutoff.

compared to a fixed high-dose strategy. For BaseScenario, there was an approximate 92.2% reduction in the toxicity incidence, yet a 243% increase in the percentage of subjects achieving their efficacy target (**Figure 3**). HighIIV along with AEshallow followed a similar trend for more subjects achieving efficacy, which was in contrast to HighRUV and HighIOV, which did not perform as well (**Table 2**).

For HighRUVIIV, compared to HighIIV, PPV, NPV, specificity, AUCROC, and Youden's J statistic, and the percentage of subjects achieving efficacy and having toxicity were all worse for HighRUVIIV. When compared to HighRUV, HighRUVIIV performed better at PPV, sensitivity, AUCROC, and Youden's J statistic, however, it performed worse at NPV and the percentage of subjects achieving efficacy and having toxicity, with efficacy and toxicity being the relevant clinical context.

**Effect of dose titration algorithm on the efficacy and utility assessments**
A summary of results for varying cutoff values (FNC:FPC ratio of 10:1) from ROC and utility index methodologies are provided in **Tables S1–S6**. Note that the highest utility index values were found in BaseScenario, HighIIV, and AEshallow (172, 175, and 157, respectively), whereas scenarios having high residual or IIVs had the lowest utility index values (142,

131, and 154 for HighRUV, HighIOV, and HighRUVIIV, respectively). The scenarios that had the biggest mismatches between the "optimal" $C_{trough}$ cutoff selected via ROC vs. utility index approaches were the scenarios with high residual or IOVs (**Table 2**). When the optimal $C_{trough}$ cutoff from the ROC approach was rounded to the nearest 10 ng/mL, the optimal $C_{trough}$ from the utility index approach, when compared to the ROC approach, was 0%, 13%, and 0% higher for BaseScenario, HighIIV, and AEshallow, respectively, and 31%, 36%, and 60% higher for HighRUV, HighIOV, and HighRUVIIV, respectively.

**Effect of varying the cost of FN and FP classifications**
Optimal cutoff and PPVs/NPVs, as a function of varying relative FNC to FPC, are depicted in **Figure 4**. Higher relative FNC costs resulted in higher NPVs, but lower optimal cutoff and PPVs. Scenarios with high RUV and IOV (HighRUV, HighIOV, and HighRUVIIV) experienced the largest drops in the optimal cutoff value for increasing FN classification costs.

**DISCUSSION**

Generally, therapeutic agents having narrow therapeutic windows make *a priori* selection of optimal dosing regimens
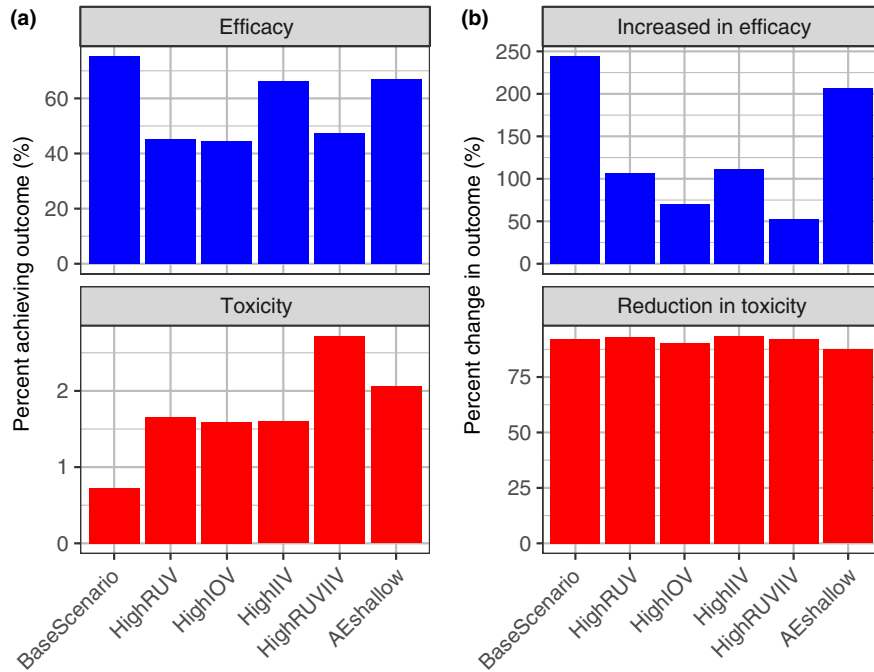
**Figure 3** The percentage of subjects achieving efficacy or having toxicity (**a**) when the titration algorithm is applied. Compared to BaseScenario, the other scenarios had lower efficacy and higher toxicity. Percent change in subjects achieving efficacy compared to a fixed low-dose strategy or percentage change in subjects having toxicity (**b**) when compared to a fixed high-dose strategy. In all scenarios, higher efficacy and lower toxicity was achieved when compared to the fixed-dosing regimens. AE, adverse event; IIV, interindividual variability; IOV, interoccasion variability; RUV, residual unexplained variability.
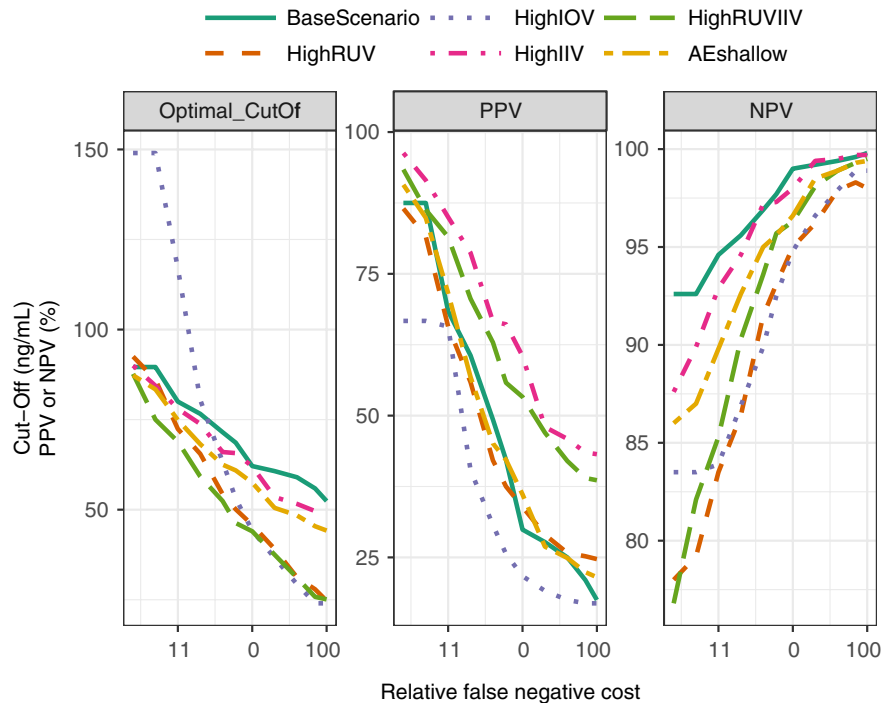


**Figure 4** Optimal day 10 trough plasma concentration ($C_{trough}$) cutoff concentrations, positive predictive value (PPV), and negative predictive value (NPV) as a function of varying costs of false-negative (FN) to false-positive (FP) classifications. The scenario with the least variability (BaseScenario) had the smallest reductions in $C_{trough}$ cutoff values as the cost of FN classifications increased. Whereas PPV is of utmost importance for low relative FN costs, NPV is influential for high FN costs. For a high FN cost it is critical that subjects at risk are not up-titrated, even if it means that subjects that might otherwise have been up-titrated are left at a lower dose. AE, adverse event; IIV, interindividual variability; IOV, interoccasion variability; RUV, residual unexplained variability.

difficult. It is complex because different degrees and/or types of variabilities may exist in PK-efficacy and PK-safety relationships. In this article, and using various scenarios, we demonstrated the utility of combined ROC and CTS-based analyses as a promising quantitative diagnostic approach for achieving the desired balance of safety and efficacy.

By classifying subjects as eligible for up-titration, individuals with the highest risk of exceeding the toxicity threshold (i.e., when high drug exposures are found at a low dose) are not up-titrated. Compared to the alternative of leaving everyone at a low dose level, the net effect of titration is higher drug exposure, PD response, efficacy, and percentage of subjects at the high dose level, all while reducing the incidence of toxicities. The current methods possess several moving parts providing flexibility in application. This flexibility is mainly driven by setting the cost of incorrectly predicting efficacy or safety, which can then be used to objectively and comprehensively derive an optimal titration rule.

In the current analysis of virtual compound A, the risks associated with a high dose were unacceptable, and yet a low dose was not anticipated to result in group-wise efficacy. For BaseScenario, HighIIV, and AEshallow, the diagnostics and simulations predicted an improved balance between safety and efficacy compared to a nontitrated dosing strategy. Generally, an AUCROC of greater > 80% indicates a good diagnostic for the PK or PD metric,[27] resulting in a high probability of correctly classifying subjects for a future intervention (i.e., up-titration). However, as illustrated with HighRUVIIV, AUCROC may be relatively high yet CTS may identify clinical relevance shortcomings in absolute efficacy and toxicity. Ultimately, proper clinical context is needed in the interpretation of AUCROC and Youden J statistics; the context being the actual percentage of subjects achieving efficacy or having toxicity.

Four of the six scenarios were dedicated to the three commonly partitioned sources of variability in population PK/PD modeling: RUV, IOV, and IIV. RUV and IOV are of similar nature in that they are random either on a sampling level (RUV) or on a per-occasion level (IOV). IIV is different in that predictive value is retained within an individual subject between sample times and occasions, and, therefore, qualifying samples will have concordance with future samples. This enables more accurate subject-level classifications, depending on the exact scenario. In contrast, high RUV or IOV decreases the efficiency of optimal cutoff point selection from a single PK sample by lowering the predictive value for subsequent measurements or periods. Notably, the included high RUV or IOV scenarios demonstrated the relative lack of separation between TP and TN subjects (e.g., **Figure S3a**). For high IIV situations, and in contrast with high RUV or IOV scenarios, one can still conduct screening at a low dose and with relatively low risk allocate the subject to the proper dose level knowing that they will have similar profiles, adjusted for dose, in subsequent periods or assessments. Therefore, even in the presence of relatively high IIV, it is still possible to establish relatively high exposures within both groups and individuals. Note that the optimum cutoff is not necessarily where the specificity and sensitivity lines cross (**Figure S3b**) because of the dependence on the efficiency calculation where the relative cost of FN classifications is specified. Further, the methods can be generalized to different titration schema with different drivers (e.g., PK, efficacy, and safety). For example, if two separate toxicity thresholds for concentration existed, say for mild and severe toxicities, then it would be possible to place a low FNC on the mild toxicity and a high FNC on the severe toxicity.

When RUV and IOV variabilities are sufficiently low, or when the probability of transitioning from safety to toxicity is well-predicted, these titration-based methods may be useful when designing trials at multiple stages of drug development. Once the ROC analysis has optimized the cutoff value according to the FNC, CTS for the given cutoff will predict the scenario's concentration vs. time profiles (**Figure S7**), and derived from this the percentage of subjects having efficacy and toxicity (**Figure 3** and **Tables S1–S6**). If these cutoffs improve the benefit-risk balance, the analyses may provide justification for conducting further validating clinical studies. Therefore, it is the combination of ROC analysis and the simulation results that would provide the clinical relevance that demonstrates the practical application of the methods described. Although many scenarios have been explored, the generalizability of the method can be further tested on other envisioned scenarios. The six included in this report provide a representative sampling of commonly occurring model features encountered in drug development that undermine predictive value.

In conclusion, this report demonstrates that drugs having sufficiently low IOV or RUV may make good candidates for PK-based titration schemes based on only one $C_{trough}$ sample, even in the presence of high IIV. A drug development program with sufficient information will facilitate the construction of quantitative models that are conducive to ROC and simulation-based justifications for individualized dose titration schema.

**Supporting Information.** Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (www.psp-journal.com).

**Figure S6.** ROC and efficiency analysis for AEshallow and a relative false negative to false positive classification cost of 10.

**Figure S7.** Post-titration median plasma concentration vs. time profiles demonstrating that, compared to BaseScenario, the scenarios, including high residual and interoccasion variabilities, had the lowest cutoff values and exposures.

**Supplementary Materials S1.** NONMEM control file example for BaseScenario PK/PD.

1. Laer, S. *et al.* Development of a safe and effective pediatric dosing regimen for sotalol based on population pharmacokinetics and pharmacodynamics in children with supraventricular tachycardia. *J. Am. Coll. Cardiol.* **46**, 1322–1330 (2005).
2. Yim, D.S., Zhou, H., Buckwalter, M., Nestorov, I., Peck, C.C. & Lee, H. Population pharmacokinetic analysis and simulation of the time-concentration profile of etanercept in pediatric patients with juvenile rheumatoid arthritis. *J. Clin. Pharmacol.* **45**, 246–256 (2005).
3. Chan, P.L., Nutt, J.G. & Holford, N.H. Levodopa slows progression of Parkinson's disease: external validation by clinical trial simulation. *Pharm. Res.* **24**, 791–802 (2007).
4. Ozawa, K., Minami, H. & Sato, H. Clinical trial simulations for dosage optimization of docetaxel in patients with liver dysfunction, based on a log-binomial regression for febrile neutropenia. *Yakugaku Zasshi* **129**, 749–757 (2009).
5. Holford, N.H. Target concentration intervention: beyond y2k. *Br. J. Clin. Pharmacol.* **52**(suppl. 1), 55s–59s (2001).
6. de Jonge, M.E., Huitema, A.D., Schellens, J.H., Rodenhuis, S. & Beijnen, J.H. Individualised cancer chemotherapy: strategies and performance of prospective studies on therapeutic drug monitoring with dose adaptation: a review. *Clin. Pharmacokinet.* **44**, 147–173 (2005).
7. Zandvliet, A.S., Schellens, J.H., Beijnen, J.H. & Huitema, A.D. Population pharmacokinetics and pharmacodynamics for treatment optimization in clinical oncology. *Clin. Pharmacokinet.* **47**, 487–513 (2008).
8. Acuna, C., Morales, J., Castillo, C. & Torres, J.P. [Pharmacokinetics of vancomycin in children hospitalized in a critical care unit]. *Rev. Chilena Infectol.* **30**, 585–590 (2013).
9. De Gregori, S. *et al.* Clinical pharmacokinetics of morphine and its metabolites during morphine dose titration for chronic cancer pain. *Ther. Drug Monit.* **36**, 335–344 (2014).
10. Han, H. *et al.* Trough concentration over 12.1 mg/l is a major risk factor of vancomycin-related nephrotoxicity in patients with therapeutic drug monitoring. *Ther. Drug Monit.* **36**, 606–611 (2014).
11. Huang, L.S., Pheanpanitporn, Y., Yen, Y.K., Chang, K.F., Lin, L.Y. & Lai, D.M. Detection of the antiepileptic drug phenytoin using a single free-standing piezoresistive microcantilever for therapeutic drug monitoring. *Biosens. Bioelectron.* **59**, 233–238 (2014).
12. Hoffer, E., Akria, L., Tabak, A., Scherb, I., Rowe, J.M. & Krivoy, N. A simple approximation for busulfan dose adjustment in adult patients undergoing bone marrow transplantation. *Ther. Drug Monit.* **26**, 331–335 (2004).
13. Park, J.W. *et al.* Rationale for biomarkers and surrogate end points in mechanism-driven oncology drug development. *Clin. Cancer Res.* **10**, 3885–3896 (2004).
14. Caveney, E.J. & Cohen, O.J. Diabetes and biomarkers. *J. Diabetes Sci. Technol.* **5**, 192–197 (2011).
15. Jelic, I. *et al.* [Opatija study: observation of hemodialysis patients and titration of cera dose just switched from another erythropoiesis stimulating agent]. *Acta Med. Croatica* **66**, 157–164 (2012).
16. Padhi, D. & Harris, R. Clinical pharmacokinetic and pharmacodynamic profile of cinacalcet hydrochloride. *Clin. Pharmacokinet.* **48**, 301–311 (2009).
17. Chen, P. *et al.* Population pharmacokinetics and pharmacodynamics of the calcimimetic etelcalcetide in chronic kidney disease and secondary hyperparathyroidism receiving hemodialysis. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 484–494 (2016).
18. Paule, I., Tod, M., Henin, E., You, B., Freyer, G. & Girard, P. Dose adaptation of capecitabine based on individual prediction of limiting toxicity grade: evaluation by clinical trial simulation. *Cancer Chemother. Pharmacol.* **69**, 447–455 (2012).
19. Park, S.H., Goo, J.M. & Jo, C.H. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J. Radiol.* **5**, 11–18 (2004).
20. Florkowski, C.M. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* **29**(suppl. 1), S83–S87 (2008).
21. Perry, P.J., Miller, D.D., Arndt, S.V. & Cadoret, R.J. Clozapine and norclozapine plasma concentrations and clinical response of treatment-refractory schizophrenic patients. *Am. J. Psychiatry* **148**, 231–235 (1991).
22. Hosmer, D. & Lemeshow, S. *Applied Logistic Regression, 2nd edn* (John Wiley & Sons, Hoboken, NJ, 2000).
23. Chabaud, S., Girard, P., Nony, P. & Boissel, J.P. Clinical trial simulation using therapeutic effect modeling: application to ivabradine efficacy in patients with angina pectoris. *J. Pharmacokinet. Pharmacodyn.* **29**, 339–363 (2002).
24. Mouksassi, M.S., Marier, J.F., Cyran, J. & Vinks, A.A. Clinical trial simulations in pediatric patients using realistic covariates: application to teduglutide, a glucagon-like peptide-2 analog in neonates and infants with short-bowel syndrome. *Clin. Pharmacol. Ther.* **86**, 667–671 (2009).
25. Southworth, H. Simulating titration to goal clinical trials with statins. *J. Biopharm. Stat.* **14**, 451–467 (2004).
26. Doshi, S., Chow, A. & Perez Ruixo, J.J. Exposure-response modeling of darbepoetin alfa in anemic patients with chronic kidney disease not receiving dialysis. *J. Clin. Pharmacol.* **50**, 75s–90s (2010).
27. Hanley, J.A. & McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
28. Zweig, M.H. & Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.* **39**, 561–577 (1993).
29. Youden, W.J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
30. Beal, S., Sheiner, L., Boeckmann, A. & Bauer, R. *NONMEM User's Guides* (Icon Development Solutions, Ellicott City, MD, 1989–2009).
31. Bauer, R. *NONMEM Users Guide: Introduction to nonmem 7.2* (Icon Development Solutions, Ellicott City, MD, 2011).
32. R Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org> (2014).