Citation: CPT Pharmacometrics Syst. Pharmacol. (2016) 5, 427–436; doi:10.1002/psp4.12093

ORIGINAL ARTICLE

New Generalized Poisson Mixture Model for Bimodal Count Data With Drug Effect: An Application to Rodent Brief-Access Taste Aversion Experiments

Y Sheng[1,2*], J Soto[1], M Orlu Gul[1], M Cortina-Borja[3], C Tuleu[1] and JF Standing[1]

Pharmacodynamic (PD) count data can exhibit bimodality and nonequidispersion complicating the inclusion of drug effect. The purpose of this study was to explore four different mixture distribution models for bimodal count data by including both drug effect and distribution truncation. An example dataset, which exhibited bimodal pattern, was from rodent brief-access taste aversion (BATA) experiments to assess the bitterness of ascending concentrations of an aversive tasting drug. The two generalized Poisson mixture models performed the best and was flexible to explain both under and overdispersion. A sigmoid maximum effect (E_{max}) model with logistic transformation was introduced to link the drug effect to the data partition within each distribution. Predicted density-histogram plot is suggested as a model evaluation tool due to its capability to directly compare the model predicted density with the histogram from raw data. The modeling approach presented here could form a useful strategy for modeling similar count data types.
CPT Pharmacometrics Syst. Pharmacol. (2016) **5**, 427–436; doi:10.1002/psp4.12093; published online 29 July 2016.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Pharmacometric models for unimodal count data have been published. However, mathematical models for bimodal count data, including both drug effects and truncated features, are less well developed.

WHAT QUESTION DOES THIS STUDY ADDRESS?
☑ Are there any pharmacometric models that can include drug effects and nonequidispersion for bimodal count data?

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE
☑ This is the first time a pharmacometric model has been introduced for analyzing bimodal count data with drug effect. A new diagnostic plot, predicted density-histogram, was recommended for mixture distribution models.

HOW THIS MIGHT CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS
☑ This modeling approach could be used for analyzing bimodal data in other PD studies.

Count data can be encountered in both preclinical and clinical pharmacodynamic (PD) studies.[1,2] It consists of non-negative integer values that record the number of discrete occurrences often linked to explanatory variables. The Poisson model is widely used in analyzing count data but, when the count number is larger than 10, normal approximation with continuity correction is usually more convenient.[3] The restriction for a Poisson distribution that its mean and variance will coincide is not always met in real PD data. Thus, the negative binominal distribution, which can handle overdispersion, and the generalized Poisson model, which can treat both under and overdispersion, have been used.[4,5]

Count data collected in pharmacological studies are not always unimodal.[6] A wide range of shapes from underdispersion to overdispersion and from unimodal to bimodal can be observed in a real count dataset. Sometimes, nonequidispersion and bimodality occur simultaneously, thus increasing data analysis difficulties. Moreover, although there is no upper limit for count data, a maximum bound is often observed in biological studies for physiological or pathological reasons.[7] This limitation should also be carefully considered in the analysis.

In dose-response studies, drug effect is usually of the most interest to researchers. Because there is only one peak in unimodal count data, it would be straightforward to add drug effect to the centering parameter (e.g., see λ_1 in Eq. 2 below).[1] However, when the data exhibit bimodal peaks, incorporation of drug effect is not so straightforward because there may not exist a unique parameter characterizing the distribution. Furthermore, although some diagnostic tools have been developed for the count model,[2,8] most of them are built for unimodal data. The evaluation of the bimodal model is therefore a challenge.

The purpose of this article was to explore and implement different mixture distribution models by including both drug effects and truncated feature for bimodal count data. Underdispersion and overdispersion patterns were investigated and different evaluation methods were also tested and compared.

[1]Department of Pharmaceutics, UCL School of Pharmacy, London, United Kingdom; [2]Center for Drug Clinical Research, Shanghai University of Chinese Medicine, Shanghai, China; [3]Population, Policy, and Practice Programme, Institute of Child Health, UCL, London, United Kingdom. *Correspondence: Y Sheng (shengyc@gmail.com)
Received 23 February 2016; accepted 9 June 2016; published online on 29 July 2016. doi:10.1002/psp4.12093

Table 1 The summary of data from BATA experiments obtained from 10 rats

	Quinine concentration (m*M*)						
	0	**0.01**	**0.03**	**0.1**	**0.3**	**1**	**3**
No.[a]	1,080	718	720	722	720	720	720
Mean ± SD	43.8 ± 16.1	32 ± 22.3	31.8 ± 21.2	17.6 ± 17.9	8.0 ± 10.2	3.9 ± 4.7	3.0 ± 3.1
CV%	36.7	69.6	66.7	101.8	126.7	119.4	100.4
Median	49.5	44	41	9	4	2	2
Q_1/Q_3	46/53	4/51	6/50	2/33	1/10	1/5	1/4
Min/max	0/60	0/60	0/61	0/60	0/51	0/50	0/24

BATA, brief-access taste aversion; CV%, coefficient of variation.
[a]No. is the length of lick numbers.

METHODS

BATA experiment data

This investigation is motivated by an exploratory data analysis of the lick numbers obtained from the rodent brief-access taste aversion (BATA) experiments. The taste of an oral medication is a critical quality attribute for therapeutic adherence and successful treatment in patients, especially children. Palatability studies with human taste panels are carried out at the latter stages of formulation development if at all. There is a great need to develop a means to assess the taste of drugs and drug formulations at the early stages of drug development so that taste aversive compounds are highlighted early, formulations are optimized, and taste masking approaches are put in place. BATA model is an emerging *in vivo* screening tool with great promise in providing taste assessment of drugs at an early stage of oral drug product design. The details of this experiment have been described previously.[9] Briefly, 10 trained rats were tested in this study. After a 22-hour water-deprived period, each rat was placed in the lickometer and was presented randomly with different sipper tubes containing either deionized water or one of the six concentrations of a bitter drug, quinine hydrochloride dihydrate. The trial began when the rat took its first lick from the sipper tube, and ended 8 seconds later when the shutter closed. Each trial was intercepted by a water-rinse of 2 seconds to minimize carry over effects from the previous solution tested. Each quinine hydrochloride dihydrate concentration was presented four times and deionized water was presented six times per 40-minute session. The number of licks was electronically recorded by the lickometer. Experiments were repeated on eight different weeks intercepted by a 1-week washout period. A total of 5,400 records of lick numbers were taken from seven different quinine hydrochloride dihydrate concentration groups obtained from 10 rats. The lick number distribution followed neither a normal nor a unimodal pattern (**Table 1** and **Figure 1**).

Because all the rats were only presented with the solutions for a very short period (8 seconds), physiologically, there would be a maximum limitation of lick numbers. In addition, the histograms of the lick numbers from the lowest four concentrations exhibited bimodal behavior and consisted of two distributions as observed. It seemed that the drug effect may influence the proportion of the different distributions but not affect each distribution itself.

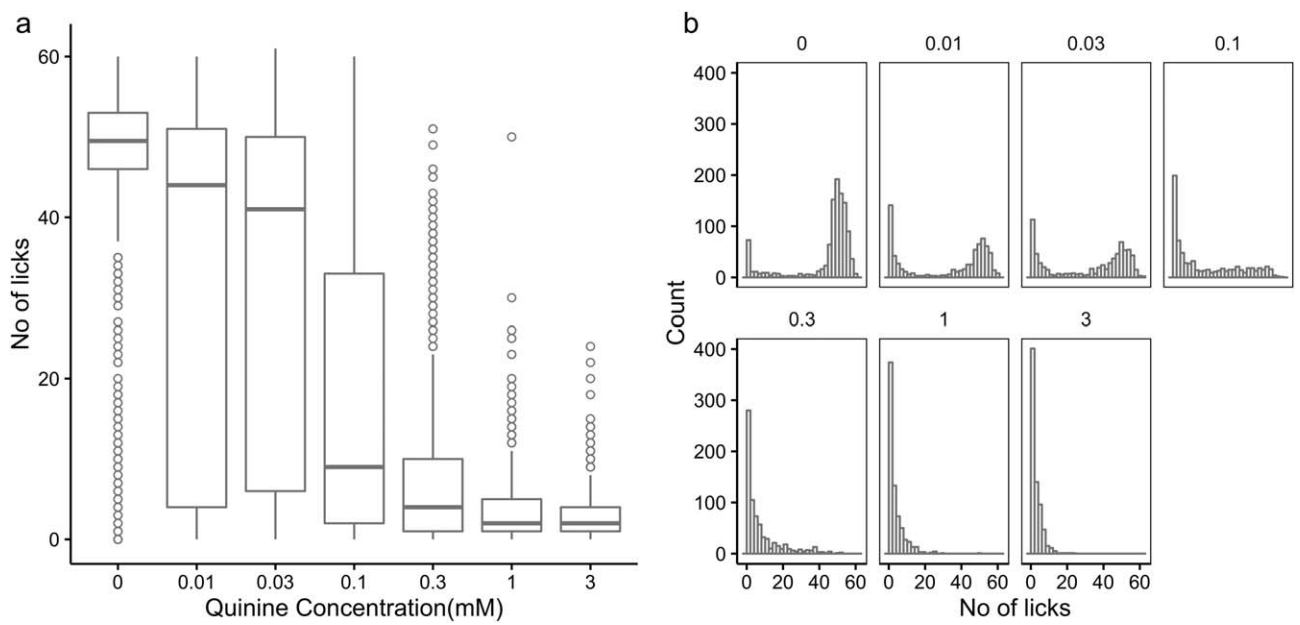

Figure 1 Box plot (**a**) and histogram (**b**) for the brief-access taste aversion (BATA) results from seven quinine concentrations.

Data analysis

Data modeling and simulation was performed with NONMEM version 7.3 (ICON, Ellicott City, MD) in conjunction with a gfortran (64-bit) compiler using Pirana version 2.9.0[10] as an interface. The first order conditional estimation with interaction method was the starting algorithm for analyses. Consequently, the Laplace integral approximation with stochastic approximation expectation-maximization and Monte Carlo importance sampling and the likelihood option was used throughout the model analyses. The R environment[11] for statistical computing version 3.2.2 was used for plots.

Models for the two distributions

Four different mixed distribution models were investigated: two Poisson mixture model (2PS), Poisson-normal mixture model (PSND), two negative binomial mixture model (2NB), and two generalized Poisson (2GP) mixture model. Except for the normal approximation in the PSND model, all the models consisted of the Poisson distribution or its extensions. All the mixture distribution models consisted of two distributions and a mixture probability (π) that indicated whether an observation was to belong to one of the two distributions:

$$P(Y_i=n)=\pi \cdot P_1+(1-\pi) \cdot P_2 \tag{1}$$

P_1 is the probability of the observation arising from the first distribution and P_2 is the belonging probability for the second distribution.

Models for the first distribution

Three models, Poisson, NB, and GP were tested for the first distribution in which most most of the lick numbers were less than 20 from **Figure 1b**: the probability of observation Y_i equal to n counts from Poisson model[5] is expressed as:

$$P_1(Y_i=n)=e^{-\lambda_1} \cdot \frac{\lambda_1^n}{n!} \tag{2}$$

where the mean and variance of Y_i are equal to λ_1.

The NB[5] has a further parameter (α_1) to handle overdispersion and the probability mass function is as follows:

$$P_1(Y_i=n)=\left[\frac{\Gamma(n+\frac{1}{\alpha_1})}{n! \cdot \Gamma(\frac{1}{\alpha_1})}\right] \cdot \left(\frac{1}{1+\alpha_1 \cdot \lambda_1}\right)^{\frac{1}{\alpha_1}} \cdot \left(\frac{\lambda_1}{\frac{1}{\alpha_1}+\lambda_1}\right)^n \tag{3}$$

where the mean of Y_i is λ_1, but the variance is $\lambda_1 \cdot (1+\alpha_1 \cdot \lambda_1)$.

The GP[12] also has a dispersion parameter (δ_1) to describe both underdispersion and overdispersion.

$$P_1(Y_i=n)=\frac{\lambda_1 \cdot (\lambda_1+n\delta_1)^{(n-1)} \cdot e^{-\lambda_1-n\delta_1}}{n!} \tag{4}$$

The mean of GP is $\lambda_1/(1-\delta_1)$ and the variance is $\lambda_1/(1-\delta_1)^3$, δ_1 can be negative or positive within the range of [max(-1,$-\lambda_1$/m),1], m≧4; the case $\delta_1=0$ corresponds to the Poisson model.

Truncated models for the second distribution

Although there is no upper limit on the Poisson distribution and its extensions, all the models for the second distribution, in which most of the lick numbers were within 40–60, were truncated to the observed maximum lick number of 61 in the results due to physiological reason. Truncated distributions ensure the sum of all the possible probabilities will be equal to 1.[7]

$$P_2(Y_i=n)=P_{truncated}=\begin{cases}\dfrac{P(Y_i=n)}{\sum_{m=0}^{max(Y_i)} P(Y_i=m)} & Y_i \in [0, max(Y_i)] \\ 0 & \text{otherwise}\end{cases} \tag{5}$$

The probabilities of second distribution $P_2(Y_i=n)$ were obtained from the truncated distribution. Besides the new set of parameters (λ_2, α_2 and δ_2), the formula for probabilities of the nontruncated second distribution $P_2(Y_i=n)$ from PS, NB, and GP models were the same as in the first one.

In general, the Poisson distribution can be approximated by a normal distribution with parameters (μ,σ^2) when the count numbers <10. Therefore, the truncated normal model with the parameters μ and σ was also investigated for the second distribution. When using the normal approximation to the Poisson distribution, a continuity correction[13] should be introduced to the probability calculation.

$$P_P(Y_i=n) \approx P_N(Y_i \le n+0.5)-P_N(Y_i \le n-0.5) \tag{6}$$

where $P_P(Y_i=n)$ is the probability of $Y_i=n$ from Poisson distribution and $P_N(Y_i \le n+0.5)$ represents the probability of $Y_i \le n+0.5$ from a normal distribution. The probability of $Y_i=n$ from a Poisson distribution is approximately equal to the probability difference between $Y_i \le n+0.5$ and $Y_i \le n-0.5$ from a normal distribution. The PHI function of NONMEM can be used for calculating the normal probability from its cumulative distribution.

Similarly, right truncation of the normal distribution can be directly calculated by PHI function.

$$\begin{aligned}\sum_{m=0}^{max(Y_i)} P_P(Y_i=m) &\approx P_N(Y_i \le max(Y_i)+0.5) \\ &= PHI\left(\frac{max(Y_i)+0.5-\mu}{\sigma}\right)\end{aligned} \tag{7}$$

Model for the drug effect

From the histogram plot, there was the smallest proportion of the first distribution and a largest proportion of the second distribution in the water group. As the quinine concentration increased, the first proportion increased and the second proportion decreased. Meanwhile, the shape of the two distributions did not change appreciably. Therefore, the sigmoid E_{max} model with logistic transformation was examined for the effect on the mixture probability (π).

Table 2 Parameter estimates of the base 2PS, PSND, 2NB, and 2GP population models

Parameter	2PS	PSND	2NB	2GP	
OFV	39714	36892	34414	34123	Bootstrap[a]
ΔOFV	0	−2902	−5380	−5671	(2.5th, 97.5th)
E_{max}	29.3 (21.6)	7.58 (58.7)	14.5 (39.9)	21.7 (2)	(16.9, 35.8)
RIC_{50}	0.0798 (1.3)	0.0787 (2.1)	0.0497 (3.9)	0.0423 (0.3)	(0.0358, 0.0551)
E_0	−1.46 (12.9)	−2.57 (38.6)	−1.73 (5.2)	−1.57 (0.1)	(−1.60, −1.55)
λ_1	3.39 (16.7)	2.56 (7.4)	5.46 (8)	1.8 (2.1)	(1.4, 2.27)
λ_2	47 (7.3)		49.7 (2.1)	75 (0.9)	(65.6, 85.4)
μ		53.7 (10.9)			
σ		21.8 (19.8)			
α_1			1.08 (135.2)		
α_2			0.00214 (10.9)		
δ_1				0.693 (2.8)	(0.655, 0.742)
δ_2				−0.479 (3.9)	(−0.697, −0.303)
γ	0.543 (19.9)	0.544 (1.8)	0.711 (1.2)	0.701 (1.2)	(0.705, 0.715)
ω^2_{Emax}	75.2% (0.1)	23.8% (4.7)	20.8% (0)	14.8% (0)	(18%, 59%)
ω^2_{RIC50}	5% (28.2)	3% (7.3)	2.4% (25.2)	1% (35.6)	(0.3%, 2%)
ω^2_{E0}	68.3% (7.4)	56.3% (14.3)	5.8% (13.8)	2.2% (8.8)	(0.2%, 3%)
$\omega^2_{\lambda 1}$	48.6% (55.5)	52.6% (1)	39.7% (14.3)	42.9% (4.8)	(16%, 60%)
$\omega^2_{\lambda 2}$	8.9% (27.9)		8.2% (16.2)	18.4% (5.4)	(10%, 33%)
ω^2_{μ}		25.6% (9.2)			
ω^2_{σ}		20.7% (13.8)			
$\omega^2_{\alpha 1}$			56.4% (37.1)		
$\omega^2_{\alpha 2}$			46.7% (0.8)		
$\omega^2_{\delta 1}$				11.4% (26.3)	(2%, 14%)
$\omega^2_{\delta 2}$				56.3% (0.9)	(22%, 82%)
ω^2_{E}	40.9% (89.8)	58.7% (123.3)	25.4% (140.5)	27.2% (34.8)	(28%, 42%)
IC_{50}	11.57	0.23	0.64	1.18	

2GP, two generalized Poisson mixture model; PSND, Poisson-normal mixture model; 2NB, two negative binomial mixture model; 2PS, two Poisson mixture model; OFV, objective function value. All fixed effect parameters are represented with the relative standard error (%) in parentheses. All random effect parameters are represented as CV% (coefficient of variation) with the relative standard error (%) in parentheses. IC_{50} is derived from E_0, E_{max}, and RIC_{50}.
[a]Bootstrap confidence intervals were obtained from 1,000 simulated datasets.

$$E = E_0 + \frac{E_{max} \cdot C^{\gamma}}{IC_{50}^{\gamma} + C^{\gamma}} \tag{8}$$

$$\pi = \text{logistic}(E) = \frac{e^{E}}{1+e^{E}} \tag{9}$$

where $\text{logistic}(E_0)$ is the minimum proportion(π_0) of the first distribution and $\text{logistic}(E_0+E_{\max})$ is the maximum proportion (π_{max}) of the first distribution. γ is a slope coefficient. It should be noted that the IC_{50} here is not the true concentration producing half-maximal effect and is a hybrid parameter. Because the drug effect was assumed to affect only the mixture probability (π), the half-maximal effect (E_{50}) can be calculated as:

$$\pi_{50} = \frac{\pi_0 + \pi_{max}}{2} = \frac{\text{logistic}(E_0) + \text{logistic}(E_0 + E_{max})}{2} \tag{10}$$

$$E_{50} = \ln\left(\frac{\pi_{50}}{1-\pi_{50}}\right) \tag{11}$$

Then, the real IC_{50} (RIC_{50}), which producing half-maximal effect, can be expressed as the transformation of the E_{max} model.

$$RIC_{50} = e^{\frac{\ln\left(\frac{(E_{50}-E_0)\cdot IC_{50}^{\gamma}}{E_{max}+E_0-E_{50}}\right)}{\gamma}} \tag{12}$$

IC_{50} in the E_{max} model can also be transformed as:

$$IC_{50} = e^{\frac{\ln\left(\frac{(E_{max}+E_0-E_{50})\cdot RIC_{50}^{\gamma}}{E_{50}-E_0}\right)}{\gamma}} \tag{13}$$

RIC_{50} was used as an estimated parameter instead of IC_{50} in modeling, the new format of E_{max} model is as following:

$$E = E_0 + \frac{E_{max} \cdot C^{\gamma}}{\frac{E_{max}+E_0-E_{50}}{E_{50}-E_0} \cdot RIC_{50}^{\gamma} + C^{\gamma}} \tag{14}$$

Except for the slope coefficient γ, interindividual variability on all the model parameters was assumed to be log-normal distribution. In addition, the log-normally distributed interindividual variability was also added to the drug effect (E).

Model development and evaluation

Because the records were count numbers and presented a bimodal pattern, the two Poisson mixture model was

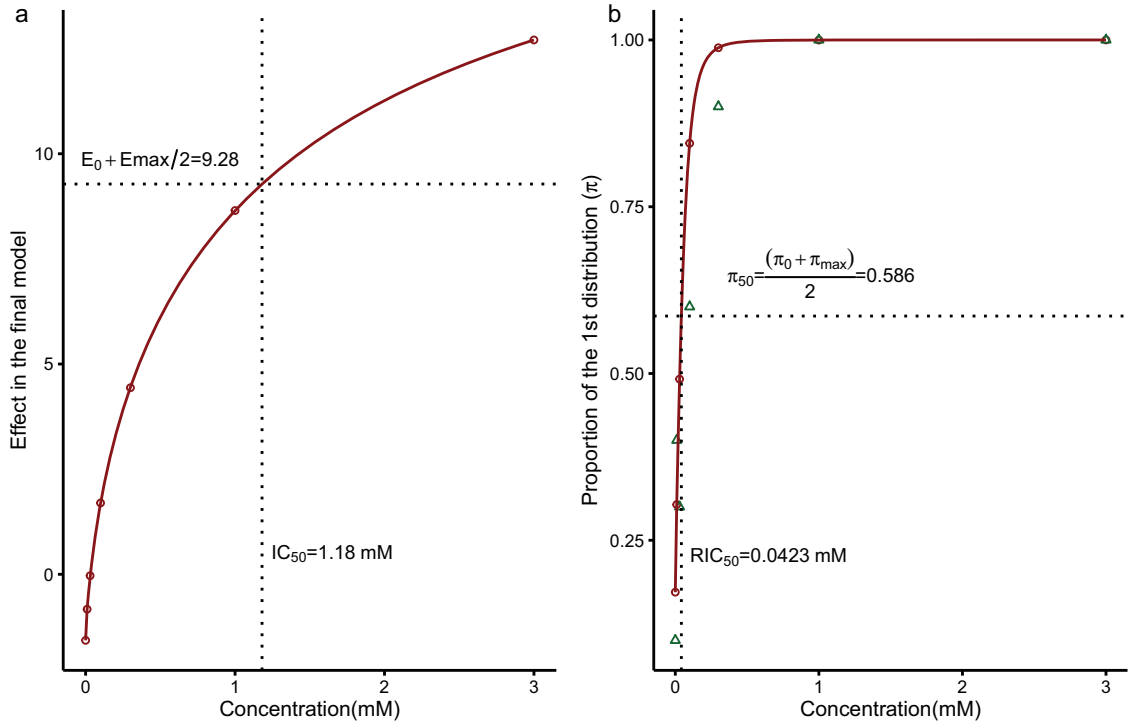

Figure 2 The effect on logistic vs. quinine concentration (**a**) and probability (π) of the first distribution vs. quinine concentration (**b**). Triangle is the proportion of less than 20 count from the original data.

regarded as a basic model. Model selection was based on the difference in the objective function value (ΔOFV) between all other models and the basic model. The accuracy of the final model parameters was evaluated by comparing the 90% confidence interval of the parameter estimates from the 1,000 bootstrap datasets with those obtained from the original dataset.[14]

Four kinds of visual assessment graphics were also explored during the modeling: (1) visual predictive check plot.[15] One thousand datasets were simulated and then 5,400,000 lick numbers with seven quinine hydrochloride dihydrate concentrations were generated for each mixture model. Results were divided into 14 equally spaced intervals based on the range of lick number. Because the count of numbers in each interval was considerably different, nine equal count intervals were also used for the assessment. Model evaluation then proceeded graphically through observing how the raw data are overlaid on the simulations. (2) Mirror histogram plot.[16] A randomly selected one of 1,000 simulated datasets from the visual predictive check (VPC) was used for producing the mirror plot. Histograms for each concentration were compared with them from the raw data. Agreement between the simulated and original histograms was assessed. (3) Suspended rootogram plot.[17] It is from the histogram but in square root scale. The bar of each bin represents the difference between the model predicted and the observed frequencies ($\sqrt{Predicted} - \sqrt{Observed}$). Then, the model performance can be assessed by comparing the deviations from x axis. (4) Predicted density-histogram plot. The model predicted probabilities of each count number in seven concentration groups were shown along the relative frequency histogram of the raw data. Relative frequency histogram instead of count can display the characteristics of raw data in the probability scale. The predictive probabilities were illustrated as a curve overlaid on the histogram. Therefore, this plot provides the direct comparison of the raw data and the model prediction.

Furthermore, a predictive check was also assessed for the final model. For each of 1,000 simulated datasets, the numbers of counts with 0, 1, ..., 5, and 50, 51, ..., 55, which were the maximum 12 numbers and represented about 60% of the raw data, were calculated for seven concentration groups. Then the means and the 5th and 95th percentiles were obtained from the simulated datasets and compared with the original dataset.

RESULTS

Initially, the conventional first order conditional estimation with interaction method was tried, but the minimization problems occurred frequently. The stochastic approximation expectation-maximization and Monte Carlo importance method performed more robustly than first order conditional estimation with interaction. All four mixture models were successfully implemented by using stochastic approximation expectation-maximization plus Monte Carlo importance algorithm. The 2PS model gave a poor fit according to the model evaluation techniques described above. Although a lower objective function value was seen using the PSND

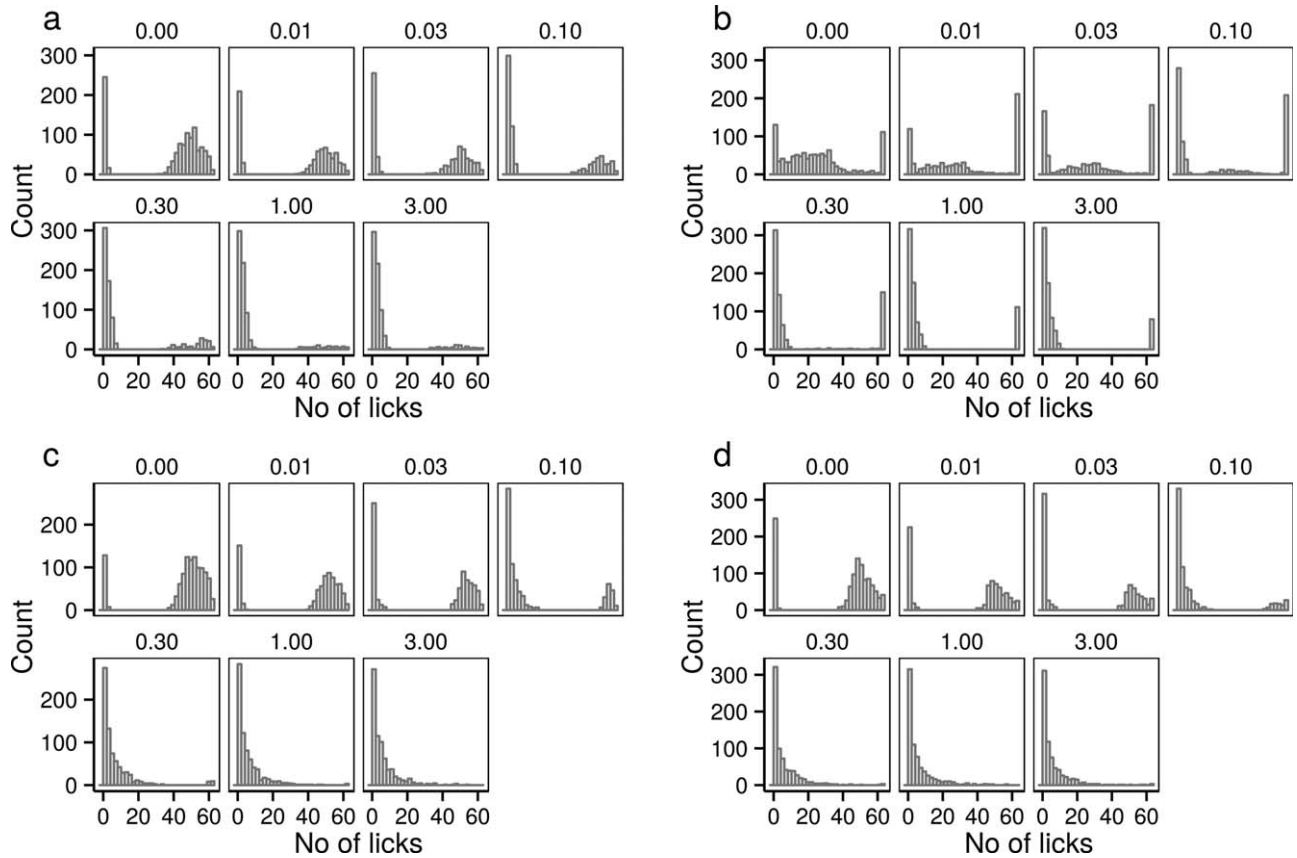

Figure 3 Mirror histogram plots of two Poisson mixture model (2PS) (**a**), Poisson-normal mixture model (PSND) (**b**), two negative binomial mixture model (2NB) (**c**), and two generalized Poisson (2GP) (**d**) for seven quinine concentrations.

model, the worse performance was observed in the graphical check. The 2NB model fitted the data better. However, the underdispersion, which likely occurred in the right part of the raw data distribution, was not well-described. Ultimately, the 2GP model, which had a minimum objective function value, was selected as the final model. Parameter estimates and objective function value (OFV) changes are given in **Table 2**. The raw data distributions of all the quinine concentration groups were best fitted by the 2GP model. The centers of left and right distributions were located at $1.8/(1\text{-}0.693) = 5.86$ and $75/(1 + 0.479) = 50.7$ with the variances of $1.8/(1\text{-}0.693)^3 = 62.2$, and $75/(1 + 0.479)^3 = 23.2$, respectively. Meanwhile, the drug effect that was modeled by sigmoid E_{max} logistic function, captured the two distribution's change from water to highest quinine concentration (**Figure 2b** and **Figure 5**). Except for the slope coefficient (γ) and intervariability of E_{max}, the bootstrapped confidence intervals were inclusive of all parameters and relatively narrow, which indicated the accuracy and precision of the parameters estimated from the final model. In addition, relative standard errors of the interindividual variability for E_{max} from the 2NB and 2GP models were estimated to be negligible.

From **Figure 1a**, the IC_{50} appeared to be in the range 0.03 to 0.1 m*M*, which corresponds with the real IC_{50} (RIC_{50}) values in all four models. However, the derived IC_{50} values were not in agreement with the original data plot and one of them was even larger than the maximum concentration tested. **Figure 2** illustrates the difference between IC_{50} and RIC_{50} from the final model.

The results of predictive checks for the final model are listed in **Supplementary Table S1**. The occurrences of no lick were overpredicted in low concentrations. For the low number of counts of 1 to 2, the observed data were underpredicted in the high concentration groups and overpredicted in the low concentration groups. For the count number in the range of 50–55, the agreements between model results and raw data were acceptable for all concentrations, although some overpredictions occurred in the higher concentrations. However, compared to the other three models, less disagreement was observed in predictive check (data not shown).

The PSND model can be simply recognized as inappropriate for the data from both mirror plot and predicted density-histogram plot. However, the mirror plots from the other three models were indistinguishable (**Figure 3**). Although some models seemed better than others from the VPCs based on the equally spaced interval, none performed well in the equally count interval categorical VPC (**Supplementary Figures S1 and S2**). Suspended

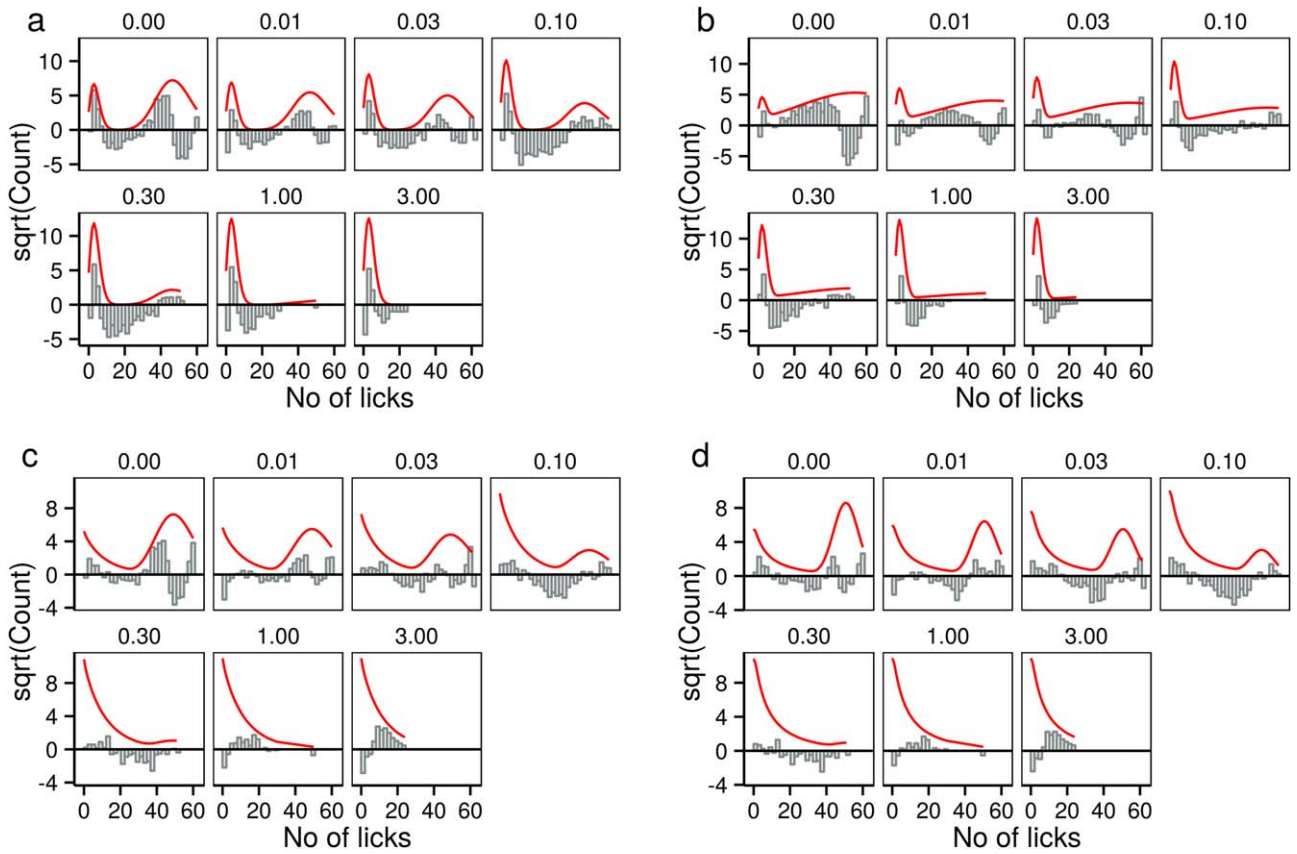

Figure 4 Suspended rootogram plots of two Poisson mixture model (2PS) (**a**), Poisson-normal mixture model (PSND) (**b**), two negative binomial mixture model (2NB) (**c**), and two generalized Poisson (2GP) (**d**) for all seven quinine concentrations. The solid red line is the model predicted frequencies and the gray area is the deviation between the predicted and the original frequencies, both are in square root scale.

rootogram plot seemed more informative than mirror plot and VPC. The deviations of 2PS and PSND models were large from the plot. Except in water group (gray bars in 2GP were closer to x axis than those in 2NB), suspended rootogram plots of all other concentrations were indistinguishable between 2NB and 2GP models (**Figure 4**). Nevertheless, the visual comparisons from predicted density-histogram plots were manifest and distinguishable (**Figure 5**). The 2GP model was the best one from the predicted density-histogram plots and it also agreed with the OFV comparisons.

DISCUSSION

Although discrete count data often appear in PD studies, bimodal count data is not so frequently encountered and appropriate analysis methods are less well developed for this case. Even if all the data are collected from one concentration, the dispersion and proportion of each distribution will cause difficulties in modeling and should be carefully considered. In particular, the drug effect and the maximum number of limitations should also be appropriately added in the analysis. This study compared four different bimodal models and illustrated how to model drug effect for bimodal count data from BATA experiments. All data from seven concentrations were well captured by the 2GP mixture model with logistic transformed E_{max} model for drug effect.

The two Poisson mixture model was a natural choice because bimodality had been exhibited in the histograms. Usually, the Poisson model is used for rare events.[18] If the count of the event during the certain time interval is a small number, a Poisson model is expected to perform well. In this study, data were centered in two ranges of 0–20 and 40–60. The two Poisson mixture model could not capture the high numbers of certain counts due to the equal mean and variance constraint. A normal distribution with continuity correction is usually considered a useful approximation for large count numbers. However, the PSND could not capture the large count distribution, which was possibly caused by interference from the few data in the middle. The predicted density-histogram plot from the 2PS model also suggested that the early peak was overdispersed and the later peak was underdispersed. As expected, the 2GP mixture model, which can treat both under and overdispersion, demonstrated superior flexibility and performed best among all models.

Drug effect was not obviously implied from the original data. After further checking, the boxplot and histogram of each concentration, the proportion of the first distribution where the number were assumed less than 20, were

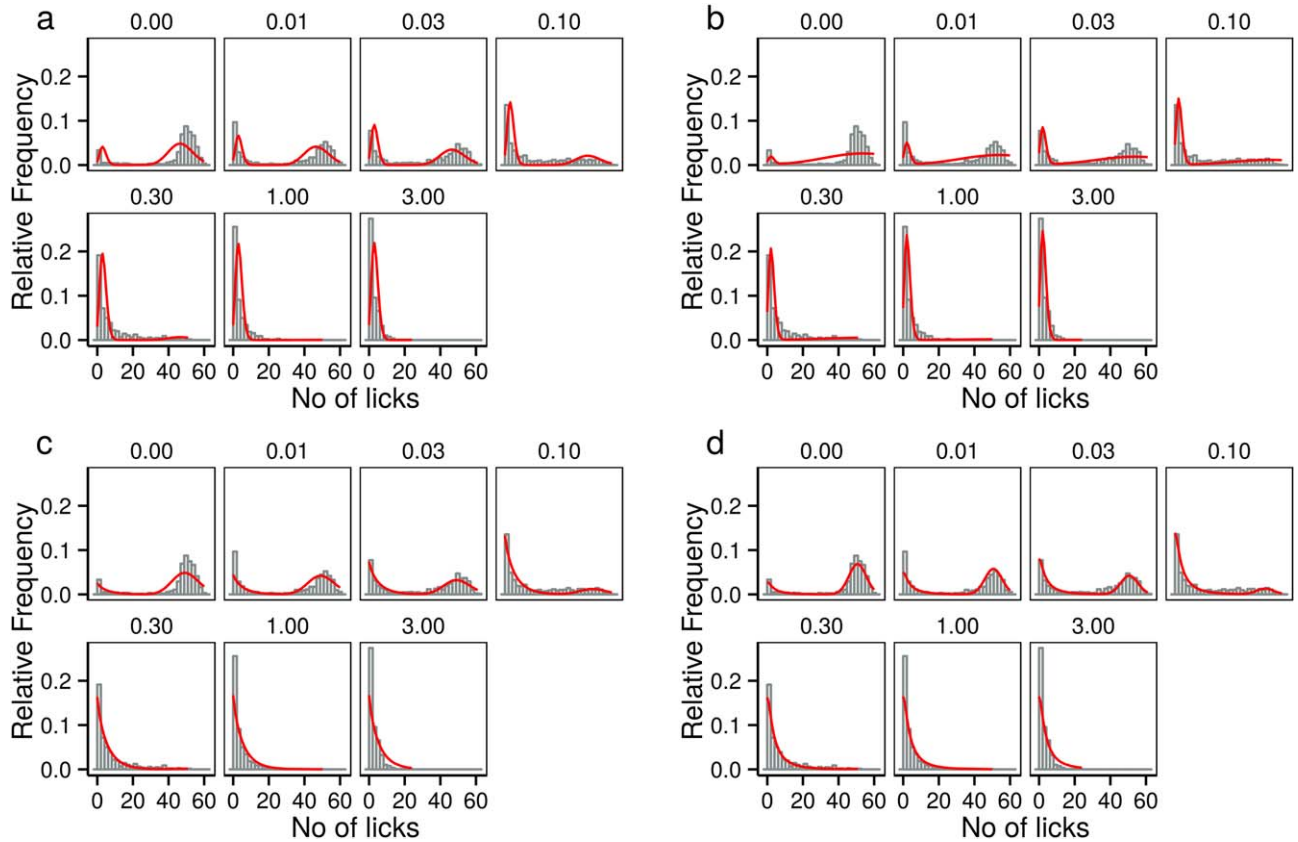

Figure 5 Predicted density-histogram plots of two Poisson mixture model (2PS) (**a**), Poisson-normal mixture model (PSND) (**b**), two negative binomial mixture model (2NB) (**c**), and two generalized Poisson (2GP) (**d**) for all seven quinine concentrations. The solid red line is the predicted probabilities and the gray area is the relative histogram of the original data.

calculated for each concentration. As shown in **Figure 2b**, the proportion vs. concentration profile strongly suggested the sigmoid E_{max} model was a promising candidate. A sigmoid E_{max} model can collapse into other drug effect models, but not vice versa. Therefore, no other drug effect model was tested. The final model also confirmed our assumption for the drug effect. Although there were some disagreements in **Figure 2b**, in consideration of the fact that the fixed cutoff value of 20 was roughly estimated before modeling and the corresponding proportions cannot be regarded as real proportions of the first distribution in observed data, the sigmoid E_{max} model with logistic transformation in the final model can adequately describe the drug effect. Another important notification from this analysis is that the IC_{50} in this kind of mixture model is not the real IC_{50} anymore and could be larger than the maximum concentration. The relationship between the IC_{50} in the model and the real IC_{50} had been presented. More attention should be paid when the maximum likelihood estimates are used. Consequently, the real IC_{50} can be used to assess the taste intensity of different drugs.

Some BATA studies treated the raw data with excluding 0 and 1.[19,20] In the water group, the lowest lick number could be rationally assumed as outliers because all the rats had been deprived water for 22 hours. Nevertheless, the proportion of lick number less than 2 was still above 5% in the water group and this would not be explained by the random chance. Although the experiment was randomized, some rats would not be thirsty at the end of a 40-minute session in which each concentration was presented four times. In addition, when the concentration of the bitter drug was increased, the low lick numbers should be the real bitter effect of the drug and could not be considered as the outliers any more. Thus, in the high concentration groups, it is impossible to differentiate the drug effect from the low lick number outliers. On the other hand, how to choose the cutoff value for the outliers is also a challenge and it would potentially affect the analysis. Consequently, keeping the whole data in the analysis is considered the best choice.

Visual check of the raw data is crucial before conducting analysis. The boxplot provided the whole concentration-response profile and the rough estimation for the real IC_{50}. It also suggested that there were substantially large variation and outliers in the raw data. To avoid the potential deceptive information from the boxplot, the histogram for each concentration was also plotted. From histograms, the data from most concentrations were distributed into two peaks rather than the one bulk in the boxplot. In addition, most of the outliers of the boxplot were not the real outliers anymore; indeed they carried the most important information of the dataset, which suggested the bimodal distribution. Thus, these outliers should not be automatically removed from the dataset and deserve special consideration.

Similarly, graphic assessment is also the key to model evaluation. In this study, four kinds of plots were investigated. The VPC is a common evaluation approach and had been successfully introduced for assessing the count model and categorical model.[15] In this study, because there were 62 different count numbers from the raw data and some of them were seldom observed in all concentrations, VPC plots for each count number was not reasonable. Consequently, two different VPCs with equal interval and equal count boundaries were explored for all models. Due to most of the data being within 0–20 and 40–60 but relatively few in between, neither VPC captured the raw data satisfactorily. Besides, the performance of VPC also was dramatically influenced by the selection of interval. Although the 2PS model had the largest OFV, it looked the best among all four models from equally spaced interval VPC. Very low proportions in the middle range of lick numbers across all concentrations occupied near the half of all panels in the equally spaced interval VPC and then concealed the unsatisfied model performance. However, when the intervals were modified as equal count, large discrepancies between observations and simulations appeared. Similarly, the suspended rootogram plot did not clearly discriminate in evaluating different models. It exaggerates low counts to make them more visible whereas peaks are less emphasized than in a histogram.[17] The mirror plot can deliver a general message of model prediction but has two main constraints. First, only one simulated dataset is used in it. This data only represents uncertainty for the interindividuals. Therefore, even if the model is an excellent fit, this simulated data may be far from the original data due to the simulation sampling. Second, it is unlikely to distinguish the better one from two models even if their OFVs are significantly different.

Although the classical goodness-of-fit, which includes observations vs. predictions, is not suited for the likelihood results from NONMEM, the model predicted likelihood is comparable with the relative frequency histogram because both are within the same range of 0–1. Predicted density-histogram plots can directly compare the model predicted density with the histogram from the raw data. Not only can this illustrate the model's shape but also it can conveniently distinguish the different models performances. Hence, the predicted density-histogram plot displayed its capacity in appraising such likelihood estimations.

Bimodality is also observed in clinical data. Some performance scales such as the walking index for spinal cord injury II,[21] reduction in the depression rating scale after repetitive transcranial magnetic stimulation treatment[22] and pain tolerance times[23] can also illustrate bimodality. Our proposed link model for drug effect, which introduced effect on the proportion of each bimodal part, is not limited in count data. It can be used for all kinds of bimodal data if necessary.

CONCLUSION

In this study, the 2GP mixture model was introduced for modeling bimodal count data. Compared to the other three mixture models, 2GP models demonstrated flexible characteristic to treat both under and overdispersion. Drug effect was also successfully described by sigmoid E_{max} model with logistic link through all seven concentration groups in BATA experiments. In addition, the relationship between the traditional IC_{50} and the real IC_{50} had been explored. After comparing with other graphics, the predicted density-histogram plot can be recommended as an appropriate tool to evaluate model fitted using likelihood estimation methods. The modeling strategy presented here could be used for similar bimodal data analyses in the future.

Acknowledgments. Y.S. received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement 602453. J.F.S. received funding from a United Kingdom Medical Research Council Fellowship (grant G1002305).

Author Contributions. Y.S. and J.F.S wrote the manuscript. J.S., M.O.G., and C.T. designed the experiment. J.S. performed the experiments. Y.S., M.C.B., and J.F.S analyzed the data.

Conflict of Interest. The authors declared no conflict of interest.

1. Ahn, J.E., Plan, E.L., Karlsson, M.O. & Miller, R. Modeling longitudinal daily seizure frequency data from pregabalin add-on treatment. *J. Clin. Pharmacol.* **52**, 880–892 (2012).
2. Plan, E.L. Modeling and simulation of count data. *CPT Pharmacometrics Syst. Pharmacol.* **3**, e129 (2014).
3. Coxe, S., G. West, S. & S. Aiken, L. Generalized Linear Models. In: Little, T.D. (ed.) *The Oxford Handbook of Quantitative Methods, Vol. 2: Statistical Analysis*, Chap. 3, pp 26–51 (Oxford University Press, New York, NY, 2013).
4. Plan, E.L., Maloney, A., Trocóniz, I.F. & Karlsson, M.O. Performance in population models for count data, part I: maximum likelihood approximations. *J. Pharmacokinet. Pharmacodyn.* **36**, 353–366 (2009).
5. Trocóniz, I.F., Plan, E.L., Miller, R. & Karlsson, M.O. Modelling overdispersion and Markovian features in count data. *J. Pharmacokinet. Pharmacodyn.* **36**, 461–477 (2009).
6. Kowalski, K.G., McFadyen, L., Hutmacher, M.M., Frame, B. & Miller, R. A two-part mixture model for longitudinal adverse event severity data. *J. Pharmacokinet. Pharmacodyn.* **30**, 315–336 (2003).
7. Plan, E.L., Elshoff, J.P., Stockis, A., Sargentini-Maier, M.L. & Karlsson, M.O. Likert pain score modeling: a Markov integer model and an autoregressive continuous model. *Clin. Pharmacol. Ther.* **91**, 820–828 (2012).
8. Plan, E.L., Karlsson, K.E. & Karlsson, M.O. Approaches to simultaneous analysis of frequency and severity of symptoms. *Clin. Pharmacol. Ther.* **88**, 255–259 (2010).
9. Soto, J., Sheng, Y., Standing, J.F., Orlu Gul, M. & Tuleu, C. Development of a model for robust and exploratory analysis of the rodent brief-access taste aversion data. *Eur. J. Pharm. Biopharm.* **91**, 47–51 (2015).
10. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e50 (2013).
11. R.D.C. Team. R: a language and environment for statistical computing. Vol. 3. <http://www.r-project.org> (Vienna, 2015).
12. Yang, Z., Hardin, J.W., Addy, C.L. & Vuong, Q.H. Testing approaches for overdispersion in Poisson regression versus the generalized Poisson model. *Biom. J.* **49**, 565–584 (2007).
13. Florescu, I. & Tudor, C. *Handbook of Probability* (John Wiley & Sons, Inc., Hoboken, NJ, 2013).
14. Ette, E.I. Stability and performance of a population pharmacokinetic model. *J. Clin. Pharmacol.* **37**, 486–495 (1997).
15. Bergstrand, M., Hooker, A.C. & Karlsson, M.O. Visual predictive checks for censored and categorical data. Vol. 18, pp 4–5. <http://www.page-meeting.org/?abstract=1604> (2009).
16. Harling, K., Hooker, A.C., Ueckert, S., Jonsson, E.N. & Karlsson, M.O. Perl speaks NONMEM (PsN) and Xpose. Vol. 19, 2010. <http://www.page-meeting.org/default.asp?abstract=2193> (2011).
17. Wainer, H. The suspended rootogram and other visual displays: an empirical validation. *Am. Stat.* **28**, 143–145 (1974).

18. Jobson, J.D. *Applied Multivariate Data Analysis. Springer Texts in Statistics* (Springer New York, New York, NY, 1991).
19. Brasser, S.M., Mozhui, K. & Smith, D.V. Differential covariation in taste responsiveness to bitter stimuli in rats. *Chem. Senses* **30**, 793–799 (2005).
20. Guenthner, C.J., McCaughey, S.A., Tordoff, M.G. & Baird, J.P. Licking for taste solutions by potassium-deprived rats: specificity and mechanisms. *Physiol. Behav.* **93**, 937–946 (2008).
21. Patrick, M. *et al*. Consumer preference in ranking walking function utilizing the walking index for spinal cord injury II. *Spinal Cord* **49**, 1164–1172 (2011).
22. Fitzgerald, P.B., Hoy, K.E., Anderson, R.J. & Daskalakis, Z.J. A study of the pattern of response to rTMS treatment in depression. *Depress. Anxiety*; e-pub ahead of print (2016).
23. Thompson, T., Keogh, E., French, C.C. & Davis, R. Anxiety sensitivity and pain: generalisability across noxious stimuli. *Pain* **134**, 187–196 (2008).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://www.wileyonlinelibrary.com/psp4)