Research Article

Theme: Revisiting IVIVC (In Vitro-In Vivo Correlation)

Guest Editors: Amin Rostami Hodjegan and Marilyn N. Martinez

Evaluating *In Vivo-In Vitro* Correlation Using a Bayesian Approach

Junshan Qiu,^{1,3} Marilyn Martinez,² and Ram Tiwari¹

Received 21 November 2015; accepted 25 January 2016; published online 19 February 2016

Abstract. A Bayesian approach with frequentist validity has been developed to support inferences derived from a “Level A” *in vivo-in vitro* correlation (IVIVC). Irrespective of whether the *in vivo* data reflect *in vivo* dissolution or absorption, the IVIVC is typically assessed using a linear regression model. Confidence intervals are generally used to describe the uncertainty around the model. While the confidence intervals can describe population-level variability, it does not address the individual-level variability. Thus, there remains an inability to define a range of individual-level drug concentration-time profiles across a population based upon the “Level A” predictions. This individual-level prediction is distinct from what can be accomplished by a traditional linear regression approach where the focus of the statistical assessment is at a marginal rather than an individual level. The objective of this study is to develop a hierarchical Bayesian method for evaluation of IVIVC, incorporating both the individual- and population-level variability, and to use this method to derive Bayesian tolerance intervals with matching priors that have frequentist validity in evaluating an IVIVC. In so doing, we can now generate population profiles that incorporate not only variability in subject pharmacokinetics but also the variability in the *in vivo* product performance.

KEY WORDS: IVIVC; MCMC; probability matching prior; tolerance intervals; Weibull distribution.

INTRODUCTION

The initial determinant of the systemic (circulatory system) exposure resulting from the administration of any non-intravenous dosage form is its *in vivo* drug release characteristics. The second critical step involves the processes influencing the movement of the drug into the systemic circulation. Since it is not feasible to run *in vivo* studies on every possible formulation, *in vitro* drug release methods are developed as surrogates. Optimally, a set of *in vitro* dissolution test conditions is established such that it can be used to predict, at some level, the *in vivo* drug release that will be achieved for a particular formulation. This raises the question of how to assess the *in vivo* predictive capability of the *in vitro* method and the extent to which such data can be used to

predict the *in vivo* performance of a “new” formulation. To this end, much work has been published on methods by which an investigator can establish a correlation between *in vivo* drug release (or absorption) and *in vitro* dissolution.

An *in vivo/in vitro* correlation (IVIVC) is a mathematical description of the relationship between *in vitro* drug release and either *in vivo* drug release (dissolution) or absorption. The IVIVC can be defined in a variety of ways, each presenting with their own unique strengths and challenges.

1. One-stage approaches: For methods employing this approach, the *in vitro* dissolution and the estimation of the *in vivo* dissolution (or absorption) are linked within a single step. These methods reflect an attempt to address some of the statistical limitation and presumptive mathematical instabilities associated with deconvolution-based methods (1) and generally express the *in vitro* dissolution profiles and the *in vivo* plasma concentration vs time profiles in terms of nonlinear mixed-effect models. Examples include:

(a) Convolution approach: While this typically involves analysis of the data in two steps, it does not rely upon a separate deconvolution procedure (2, 3). Hence, it is considered a “one-stage” approach. In the first step, a model is fitted to the unit impulse response (UIR) data for each subject, and individual pharmacokinetic parameter estimates are obtained. The second stage involves

This article reflects the views of the author and should not be construed to represent FDA’s views or policies.

¹Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA.

²Office of New Animal Drug Evaluation, Center for Veterinary Medicine, Food and Drug Administration, Rockville, Maryland, USA.

³To whom correspondence should be addressed. (e-mail: junshan.-qiu@fda.hhs.gov;)

modeling the *in vivo* drug concentration-time profiles and the fraction dissolved *in vitro* for each formulation in a single step. This procedure allows for the incorporation of random effects into the IVIVC estimation.

(b) One-step approach: In this case, neither deconvolution nor convolution is incorporated into the IVIVC. Accordingly, this method addresses *in vivo* predictions from a very different perspective: using the IVIVC generated within a single step in the absence of a UIR to predict the *in vivo* profiles associated with the *in vitro* data generated with a new formulation (*i.e.*, the plasma concentration *vs* time profile is expressed in terms of the percent dissolved *in vitro* dissolution rather than as a function of time). Examples include the use of integral transformations (4) and Bayesian methods that allow for the incorporation of within- and between- subject errors and avoid the need for a normality assumption (5).

(c) Stochastic deconvolution: We include this primarily for informational purposes as it typically serves as a method for obtaining an initial deconvolution estimate. Typically, this would be most relevant when utilizing a one-stage approach, serving as a mechanism for providing insights into link functions (fraction dissolved *in vitro* *vs* fraction dissolved *in vivo*) that may be appropriate starting points when applying the one-stage approach. Although stochastic deconvolution is optimal when a UIR is available, this can be obviated by an identifiable pharmacokinetic model and a description of the elimination phase obtained from the dosage form in question. The *in vivo* event is treated as a random variable that can be described using a nonlinear mixed-effect model (6). A strength of this method is that it can be applied to drugs that exhibit Michaelis-Menton kinetics and biliary recycling (*i.e.*, in situations where an assumption of a time-invariant system may be violated). A weakness is that it typically necessitates a dense dataset and an *a priori* description of the drug's pharmacokinetics.

(d) Bayesian analysis: This method also addresses the *in vivo* events as stochastic processes that can be examined using mixed-effect models. Assuming that oral drug absorption is dissolution-rate limited, priors and observed data are combined to generate *in vivo* predictions of interest in a one-stage for a formulation series. Posterior parameter estimates are generated in the absence of a UIR (similar to that of the method by Kakhi and Chtendon, 2013). The link between observed *in vivo* blood level profiles and *in vitro* dissolution is obtained by substituting the apparent absorption rate constant with the *in vitro* dissolution rate constant. A time-scaling factor is applied to account for *in vivo/in vitro* differences. In so doing, the plasma profiles are predicted directly

on the basis of the *in vitro* dissolution data and the IVIVC model parameters (7).

II. Two-stage approaches: The *in vivo* dissolution or absorption is modeled first, followed by a second step whereby the resulting *in vivo* predictions are linked to the *in vitro* dissolution data generated for each of the formulations in question. A UIR provides the backbone upon which plasma concentration *vs* time profiles are used to determine the parameters of interest (*e.g.*, *in vivo* dissolution or *in vivo* absorption). These deconvolved values are subsequently linked to the *in vitro* dissolution data, generally *via* a linear or nonlinear regression. Several types of deconvolution approaches are available including:

1. Model-dependent: these methods rely upon the use of mass balance considerations across pharmacokinetic compartments. A one- (8) or two- (9) compartment pharmacokinetic model is used to deconvolve the absorption rate of a drug from a given dosage form over time.
2. Numerical deconvolution: a variety of mathematical numerical deconvolution algorithms are available, (*e.g.*, see reviews by 10, 11). First introduced in 1978 (12), linear systems theory is applied to obtain an input function based upon a minimization of the sums of squared residuals (estimated *vs* observed responses) to describe drug input rate. A strength of the numerical approach is that it can proceed with minimal mechanistic assumptions.
3. Mechanistic models: *In silico* models are used to describe the *in vivo* dissolution or absorption of a drug from a dosage form (13, 14). A UIR provides the information upon which subject-specific model physiological and pharmacokinetic attributes (system behavior) are defined. Using this information, the characteristics of the *in vivo* drug dissolution and/or absorption can be estimated. A range of *in silico* platforms exists, with the corresponding models varying in terms of system complexity, optimization algorithms, and the numerical methods used for defining the *in vivo* performance of a given formulation.

Depending upon the timeframe associated with the *in vitro* and *in vivo* data, time scaling may be necessary. This scaling provides a mechanism by which time-dependent functions are transformed such that they can be expressed on the same scale and back-transformation applied as appropriate (15). Time scaling can be applied, irrespective of method employed.

Arguments both for and against each of these various approaches have been expressed, but such a debate is outside the objectives of the current manuscript. However, what is relevant to the current paper is that our proposed use of a Bayesian hierarchical model for establishing the IVIVC can be applied to any of the aforementioned approaches for generating an IVIVC. In particular, the focus of the Bayesian hierarchical approach is its application to the "Level A" correlation. Per the FDA Guidance for Extended Release Dosage Forms (16), the primary goal of a "Level A" IVIVC is to predict the entire *in vivo*

absorption or plasma drug concentration time course from the *in vitro* data resulting from the administration of drugs containing formulation modifications, given that the method for *in vitro* assessment of drug release remains appropriate. The prediction is based on the one-to-one “link” between the *in vivo* dissolution or absorption fraction, $A(t)$, and the *in vitro* dissolution fraction $D(t)$ for a formulation at each sampling time point, t . The “link” can be interpreted as a function, g , which relates $D(t)$ to $A(t)$, by $A(t) = g(D(t))$. To make a valid prediction of the *in vivo* dissolution or absorption fraction for a new formulation, $A^*(t)$, the relationship between the $A^*(t)$ and the *in vitro* dissolution fraction, $D^*(t)$, should be the same as the relationship between $A(t)$ and $D(t)$. In general, this is assumed to be true. Traditionally, mean *in vivo* dissolution or absorption fractions and mean *in vitro* dissolution fractions have been used to establish IVIVC via a simple linear regression. Separate tests on whether the slope is 1 and the intercept is 0 were performed. These tests are based on the assumption that *in vitro* dissolution mirrors *in vivo* dissolution (absorption) exactly. However, this assumption may not be valid for certain formulations. In addition, we should not ignore the fact that the fraction of the drug dissolved (absorbed) *in vivo* used in the modeling is not directly observable.

For the purpose of the current discussion, the IVIVC is considered from the perspective of a two-stage approach. In general, the development of an IVIVC involves complex deconvolution calculations for the *in vivo* data with introduction of additional variation and errors while the variation among repeated assessment of the *in vitro* dissolution data is relatively small. In this regard, we elected to ignore the variability among the *in vitro* repeated measurements. The reliability of the deconvolution is markedly influenced by the amount of *in vivo* data such as the number of subjects involved in the study, the number of formulations evaluated, and the blood sampling schedule (17), the model selection and fit, the magnitude of the within- and between-individual variability in *in vivo* product performance, and analytical errors. These measurement errors, along with sampling variability and biases introduced by model-based analyses affect the validity of the IVIVC. Incorporating the measurement errors, all sources of variability and correlations among the repeated measurements in establishing IVIVC (particularly at “Level A”) has been studied using the Hotelling’s T^2 test (18) and the mixed-effect analysis by Dune *et al.* (19). However, these two methods cannot uniformly control the type I error rate due to either deviation from assumptions or misspecification of covariance structures. O’Hara *et al.* (20) transformed both dissolution and absorption fractions, used a link function, and incorporated between-subject and between-formulation variability as random effects in a generalized linear model. The link functions used include the logit, the log-log, and the complementary log-log forms. Gould *et al.* (5) proposed a general framework for incorporating various kinds of errors that affect IVIVC relationships in a Bayesian paradigm featured by flexibility in the choice of models and underlying distributions, and the practical way of computation. Note that the convolution and deconvolution procedures were not discussed in this paper.

Since the *in vivo* fraction of the drug dissolved/absorbed is not observable directly and includes deconvolution-related variation, there is a need to report the estimated fraction of the drug dissolved (absorbed) *in vivo* with quantified uncertainty such as tolerance intervals. Specifically, use of a tolerance interval approach enables the investigator to make inferences on a specified proportion of the population with some level of confidence. Currently available two-stage approaches for correlating the *in vivo* and *in vitro* information are dependent on an assumption of linearity and time-invariance (*e.g.*, see discussion by 6). Therefore, there is a need to have a method that can accommodate violations in these assumptions without compromising the integrity of the IVIVC. Furthermore, such a description necessitates the flexibility to accommodate inequality in the distribution error across the range of *in vitro* dissolution values (a point discussed later in this manuscript). The proposed method provides one potential solution to this problem. Secondly, the current two-stage methods do not allow for the generation of tolerance intervals, thus the latter becomes necessary when the objective is to infer the distribution for a specific proportion of a population. The availability of tolerance limits about the IVIVC not only facilitates an appreciation of the challenges faced when developing *in vivo* release patterns but also is indispensable when converting *in vitro* dissolution data to the drug concentration vs time profiles across a patient population. In contrast, currently available approaches focus on the “average” relationship, as described by the traditional use of a fitted linear regression equation when generating a “Level A” IVIVC. Although typically, expressed concerns with “averages” have focused on the loss of information when fitting a simple linear regression equation (20), the use of linear regression to describe the IVIVC, in and of itself, is a form of averaging. As expressed by Kortejarvi *et al.*, (2006), in many cases, inter- and intra-subject variability of pharmacokinetics can exceed the variability between formulation, leading to IVIVC models that can be misleading when based upon averages. The use of nonlinear rather than linear regression models (*e.g.*, see 21) does not resolve this problem.

Both Bayesian and frequentist approaches envision the one-sided lower tolerance interval as a lower limit for a true $(1 - \beta)^{th}$ quantile with “confidence” γ . Note that the Bayesian tolerance interval is based on the posterior distribution of θ given X and any prior information while the frequentist counterpart is based on the data observed (X). In addition, Bayesian interprets “confidence” γ as subjective probability; frequentist interprets it in terms of long-run frequencies. Aitchison (22) defined a β -content tolerance interval at confidence, γ , which is analogous to the one defined *via* the frequentist approach, as follows:

$$\Pr_{X|\theta} [C_{X,\theta}(S(X)) \geq \beta] = \gamma,$$

where $C_{X,\theta}(S(X))$ denotes the content or the coverage of the random interval $S(X)$ with lower and upper tolerance limits $a(X)$ and $b(X)$, respectively. The frequentist counterpart can

answer the question: what is the interval (a, b) within which at least β proportion of the population falls into, with a given level of confidence γ ? Later, Aitchison (23) and Aitchison and Sculthorpe (24) further extended the β -content tolerance interval to a β -expectation tolerance interval, which satisfies

$$E_{X|\theta}[C_{X,\theta}(S(X))] = \beta.$$

Note that the β -expectation tolerance intervals focus on prediction of one or a few future observations and tend to be narrower than the corresponding β -content tolerance intervals (24). In addition, tolerance limits of a two-sided tolerance interval are not unique until the form of the tolerance limits is reasonably restricted.

Bayesian Tolerance Intervals

A one-sided Bayesian (β, γ) tolerance interval with the form $[a, +\infty)$ can be obtained by the γ -quantile of the posterior of the β -quantile of the population. That is,

$$a \leq q(1-\beta; \theta).$$

Conversely, for a two-sided Bayesian tolerance interval with the form $[a, b]$, no direct method is available. However, the two-sided tolerance interval can be arguably constructed from its one-sided counterpart. Young (25) observed that this approach is conservative and tends to make the interval unduly wide. For example, applying the Bonferroni approximation to control the central $100 \times \beta\%$ of the sample population while controlling both tails to achieve at least $100 \times (1-\alpha)\%$ confidence, $[100 \times (1-\alpha/2)\%]/[100 \times (\beta+1)/2\%]$ one-sided lower and upper tolerance limits will be calculated and used to approximate a $[100 \times (1-\alpha)\%]/[100 \times \beta\%]$ two-sided tolerance interval. This approach is only recommended when procedures for deriving a two-sided tolerance interval are unavailable in the literature due to its conservative characteristic.

Pathmanathan et al. (26) explored two-sided tolerance intervals in a fairly general framework of parametric models with the following form:

$$[d(\theta) - g^{(n)}, b(\theta) + g^{(n)}],$$

where θ is the maximum likelihood estimator of θ based on the available data X , $b(\theta) = q(1-\beta_1; \theta)$, $d(\theta) = q(\beta_2; \theta)$, and

$$g^{(n)} = n^{-1/2}g_1 + n^{-1}g_2 + O_p(n^{-3/2}).$$

Both g_1 and g_2 are $O_p(1)$ functions of the data, X , to be so determined that the interval has β -content with posterior credibility level $\gamma + O_p(n^{-1})$. That is, the following relationship holds,

$$\begin{aligned} P^\pi \{ F(b(\theta) + g^{(n)}; \theta) - F(d(\theta) - g^{(n)}; \theta) \geq \beta | X \} \\ = \gamma + O_p(n^{-1}), \end{aligned}$$

where $F(\cdot; \theta)$ is the cumulative distribution function (CDF), $P^\pi[\cdot | X]$ is the posterior probability measure under the probability matching prior $\pi(\theta)$, and $O_p(n^{-1})$ is the margin of error. In addition, to warrant the approximate frequentist validity of two-sided Bayesian tolerance intervals, the probability matching priors were characterized (See Theorem 2 in Pathmanathan et al. (26)). Note that g_2 involves the priors. The definition of g_2 is provided in the later section. The probability matching priors are appealing as non-subjective priors with an external validation, providing accurate frequentist intervals with a Bayesian interpretation. However, Pathmanathan et al. (26) also observed that probability matching priors may not be easy to obtain in some situations. As alternatives, priors that enjoy the matching property for the highest posterior density regions can be considered. For an inverse Gaussian model, the Bayesian tolerance interval based on priors matching the highest posterior density regions could be narrower than the frequentist tolerance interval for a given confidence level and a given β -content.

Implementation of Bayesian analyses has been hindered by the complexity of analytical work particularly when a closed form of posterior does not exist. However, with the revolution of computer technology, Wolfinger (27) proposed an approach for numerically obtaining two-sided Bayesian tolerance intervals based on Bayesian simulations. This approach avoided the analytical difficulties by using computer simulation to generate a Markov chain Monte Carlo (MCMC) sample from posterior distributions. The sample then can be used to construct an approximate tolerance interval of varying types. Although the sample is dependent upon the selected computer random number seed, the difference due to random seeds can be reduced by increasing sample size.

With the pros and cons of the methods developed previously, we propose to combine the approach for estimating two-sided Bayesian tolerance intervals by Pathmanathan et al. (26) with the one by Wolfinger (27). This article presents an approach featured by prediction of individual-level *in vivo* profiles with a “Level A” IVIVC established *via* incorporating various kinds of variation using a Bayesian hierarchical model. In the Methods section, we describe a Weibull hierarchical model for evaluating the “Level A” IVIVC in a Bayesian paradigm and how to construct a two-sided Bayesian tolerance interval with frequentist validity based upon random samples generated from the posterior distributions of the Weibull model parameters and the probability matching priors. In the Results section, we present a method for validating the Weibull hierarchical model, summarize the posteriors of the Weibull model parameters, show the two-sided Bayesian tolerance intervals at both the population and the individual levels, and compare these tolerance intervals with the corresponding Bayesian credible intervals. Confidence intervals differ from credibility intervals in that the credible interval describes bounds about a population parameter estimated as defined by Bayesian

posteriors while the confidence interval is an interval estimate of a population parameter based upon assumptions consistent with the Frequentist approach. As a final step, we generate *in vivo* profile predictions using “new” *in vitro* dissolution data.

Please note that within the remainder of this manuscript, discussions of the IVIVC from the perspective of *in vivo* dissolution are also intended to cover those instances where the IVIVC is defined in terms of *in vivo* absorption.

METHODS

Bayesian Hierarchical Model

Let $X[t, kj]$ represent the fraction of drug dissolved at time t from the k^{th} *in vitro* replicate in the j^{th} formulation (or dosage unit) and let $Y[t, ij]$ represent the fraction of drug dissolved/absorbed at time t from the i^{th} subject in the j^{th} formulation. An IVIVC model involves establishing the relationship between the $X[t, kj]$ and the $Y[t, ij]$ or between their transformed forms such as the log and the logit transformations. Corresponding to these transformations, proportional odds, hazard, and reverse hazard models were studied (19, 20). These models can be described using a generalized model as below,

$$L(Y[t, ij]) = h_1(\alpha) + Bh_2(X[t, kj]) + r[t, ij], 0 \leq t \leq \infty \quad (1)$$

where $L(\cdot)$ is the generic link function, h_1 and h_2 are the transformation functions, and $r[t, ij]$ is the residual error at time t for i^{th} subject and j^{th} formulation. Note that the *in vitro* dissolution fraction is assumed to be 0 at time 0. As such, there is no variation for the *in vitro* dissolution fraction at time 0. Thus, time 0 was not included in the analysis. Furthermore, this generalized model can be extended to include variation among formulations and/or replicates *in vitro*; variation among formulations, subjects, and combinations of formulations and subjects *in vivo*, $b_1[ij]$, and variation across sampling times, $b[t]$. Depending on the interests of the study, Eq. (1) can be extended as follows:

$$L(Y[t, ij]) = h_1(\alpha) + Bh_2(X[t, kj]) + b_1[ij] + r[t, ij], 0 \leq t \leq \infty \quad (2a)$$

$$L(Y[t, ij]) = h_1(\alpha) + Bh_2(X[t, kj]) + b[t] + r[t, ij], 0 \leq t \leq \infty \quad (2b)$$

$$L(Y[t, ij]) = h_1(\alpha) + Bh_2(X[t, kj]) + b_1[ij] + b[t] + r[t, ij], 0 \leq t \leq \infty \quad (2c)$$

Since, sometimes, the design of the *in vivo* study does not allow the separation of variations related to formulations and subjects, variation among combinations of formulations and subjects, $b_1[ij]$, should be used. In addition, the correlation between the repeated observations within the same subject and formulation *in vivo* and *in vitro* can be counted to some degree when modeling

both the random effects, $b_1[ij]$ and $b[t]$ in the same model. However, the correlation between these two random effects is usually not easy to specify, it can simply be assumed that the two random effects are independent. When generating a “Level A” IVIVC, we are dealing with establishing a correlation between observed (*in vitro*) vs deconvoluted (*in vivo*) dataset. Although the original scale of the *in vivo* data (blood levels) differs from that of the *in vitro* dataset, the ultimate correlation (% dissolved *in vitro* vs *in vivo* % dissolved or % absorbed) is generated on the basis of variables that are expressed on the same scale. It is from this perspective that if the within-replicate measurement error is small, it is considered ignorable relative to the between-subject, within-subject, and between-formulation variation. As such, the average of the fractions of drug dissolved at time t from the *in vitro* replicates for the j^{th} formulation, $X[t, j]$, was included in the analyses. This is consistent with the assumptions associated with the application of the F2 metric (28). We further extend the flexibility of the model in (Eq. 2) by modeling the distribution parameters of $Y[t, ij]$ and, the mean of $Y[t, ij]$:

$$Y[t, ij] \sim F(mu[t, ij], \theta \setminus mu[t]), 0 \leq t \leq \infty \quad (3)$$

$$L(mu[t, ij]) = h_1(\alpha) + Bh_2(X[t, kj]) + b[t], 0 \leq t \leq \infty \quad (4)$$

Here, F is the distribution function of Y with a parameter vector θ ; $mu[t, ij]$ is the model parameter which is linked to $X[t, kj]$ via the link function L and the model as in Eq. 4, and $\theta \setminus mu[t]$ denotes the parameter vector without mu at sampling time t . For the distribution of Y (i.e., F), a Weibull distribution is used as an example in this article. The link function L in log maps the domain of the scale parameter, $mu[t, ij]$, for the Weibull distribution to $[-\infty, +\infty]$. In addition, we assume that the distribution parameters vary across the sampling time points. The variation for the model of *in vitro* dissolution proportions at each sampling time point is $b[t]$ which is modeled as a Normal distribution in the example.

Weibull Hierarchical Model Structure and Priors

A Weibull hierarchical model was developed to assess the IVIVC conveyed by the data from Eddington *et al.* (29). We analyzed the data assuming a parametric Weibull distribution for the *in vivo* dissolution profile, $Y[t, ij]$. That is,

$$Y[t, ij] \mid \theta = (\gamma[t], mu[t, ij]) \sim \text{Weibull}(\gamma[t], mu[t, ij]), \text{ and}$$

$$\gamma[t] \sim \text{Uniform}(0.001, 20).$$

We started with a simple two-parameter Weibull model. If the model cannot explain the data, a more general Weibull model can be considered. The Weibull model parameters include the shape parameter at each sampling time point, $\gamma(t)$, and the scale parameter for each subject and formulation combination at each sampling time point, $mu[t, ij]$.

Correspondingly, the Weibull distribution has a density function in the following form:

$$f(x; \mu, r) = (r/\mu)(x/\mu)^{r-1} \exp\{-(x/\mu)^r\}.$$

Note that $\mu[t, ij]$ is further transformed to $Mut[t, ij]$ via the following formula:

$$Mut[t, ij] = \frac{1}{\mu[t, ij]^{\gamma[t]}}$$

to accommodate the difference of parameterization between OpenBUGS version 3.2.3 and Wolfram Mathematica version 9. The range of the uniform distribution for $\gamma[t]$ is specified to roughly match the range of the *in vivo* dissolution profile. Thus, the distribution of *in vivo* dissolution proportions can vary across the sampling time points. The log transformed scale parameter, $\log(\mu[t, ij])$, is linked to the average of the fractions of drug dissolved at time t , $X[t, j]$, via a random-effect sub-model as follows,

$$\log(\mu[t, ij]) = B \times (X[t, j] - 50)/50 + b[t], \text{ and}$$

$$b[t] \sim \text{Normal}(0, \tau).$$

$X[t, j]$ ranges from 0 to 100 and is centered at 50 and divided by 50 in the analysis. B is the regression coefficient for the transformed $X[t, j]$ in the random-effect sub-model, which includes an additive random effect $b[t]$ at each sampling time point. The random effect $b[t]$ accounts for the variation at each sampling time point of the observed values for the *in vitro* dissolution profile and follows a Normal distribution with a mean 0 and a precision parameter, τ . In the absence of direct knowledge on the variation in the time-specific random effect, we adopt a Gamma (0.001, 0.001) non-informative prior for the precision parameter. Both the regression coefficient, B ,

and the precision parameter, τ , are given independent “non-informative” priors, namely,

$$B \sim \text{Normal}(0, 0.0001), \text{ and}$$

$$\tau \sim \text{Gamma}(0.001, 0.001).$$

Note that a description of the variation across formulations and subjects is the primary objective for this effort. The variation across the replicates and the within-subject error are assumed ignorable relative to the formulation and subject-related variation. This Weibull hierarchical model is further summarized as in Fig. 1, where M is the number of sampling time points and N is the number of combinations of formulations and subjects.

The node “Ypred” is the posterior predictive distribution for the *in vivo* dissolution profile, which is used for checking model performance and making inference using only the new data for the *in vitro* dissolution. The node “Yc” is the empirical (sampling) distribution of samples from the Weibull distribution defined with the posteriors of the parameters “ r ” and “ μ ”. The 5 and 95% quantiles of Yc are the lower and upper limits of the 90% credible interval. Note that the credible interval could be at a population or an individual level. If samples are generated with population posteriors of $r[t]$ and $\mu[t]$, the corresponding credible interval is at a population level. If samples are generated with individual posteriors of $r[t]$ and $\mu[t, ij]$, the corresponding credible interval is at an individual level. A credible interval at an individual level will be wider than its counterpart at the population level. If no observations for certain t and/or ij are collected for Y , samples from the corresponding posteriors are used to infer the predictive distribution.

Prediction of *In Vivo* Dissolution Profile with *In Vitro* Dissolution Data for a New Formulation

One of the research interests is to use the established Bayesian hierarchical model to predict the *in vivo* dissolution or *in vivo* absorption profiles using *in vitro* dissolution data

Fig. 1. Weibull hierarchical model

generated for a new formulation. Whether a prediction refers to *in vivo* dissolution or absorption is determined by the design of the *in vivo* study and the deconvolution method employed. Either endpoint is equally applicable to the proposed tolerance interval approach. Since there is no *in vitro* dissolution data for a new formulation associated with our current dataset, we randomly selected one formulation and subject combination, formulation “Med” and Subject 1, and set the corresponding *in vivo* dissolution data as missing. With the Bayesian hierarchical model established based on the remaining data, the predictive distribution of the *in vivo* dissolution profile for Subject 1 dosed with formulation “Med” was created.

Bayesian Tolerance Intervals

Our approach for estimating the two-sided Bayesian tolerance intervals is inspired by Pathmanathan *et al.* (26) and Wolfinger (27). The steps are summarized as follows.

- For inference at the population level, the posterior mean of the model parameter, $\mu[t, ij]$, across the combinations of subjects and formulations, $\mu[t, .]$, and the posterior of $r[t]$ at each sampling time point were used to generate a random sample $Y^*[t]$ at size of 100, which follows a Weibull distribution with a scale parameter $\mu[t, .]$ and a shape parameter $r[t]$.
- For inference at the individual level, the posterior means of the model parameters, $\mu[t, ij]$ and $r[t]$, at each combination of subject, formulation and sampling time point were used to generate a random sample $Y^*[t, ij]$ at size of 100, which follows a Weibull distribution with a scale parameter $\mu[t, ij]$ and a shape parameter $r[t]$.
- Calculate the two-sided Bayesian tolerance interval *via* the approach by Pathmanathan *et al.* (26) at either the population or the individual level using the random sample $Y^*[t]$ or $Y^*[t, ij]$, correspondingly. Here, “individual” refers to the combination of subject and formulation. The two-sided Bayesian tolerance interval with β -content and γ confidence level, using the probability matching priors, was specified in the following form with equal tails

$$\left[q\left(\beta/2; \theta\right) - g^{(n)}, q\left(1-\beta/2; \theta\right) + g^{(n)} \right],$$

where the θ includes the maximum likelihood estimator of the scale parameter μ and the shape parameter r for the Weibull distribution with a density function

$$f(x; \mu, r) = (x/\mu)(x/\mu)^{r-1} \exp\{-(x/\mu)^r\}.$$

RESULTS

Weibull Hierarchical Model

Model Evaluation

Before making any inference based on the posterior distributions, convergence must be achieved for the MCMC simulation of each chain for each parameter. In addition, if

the MCMC simulation has an adaptive phase, any inference was made using values sampled after the end of the adaptive phase. The Gelman-Rubin statistic (R), as modified by Brooks and Gelman (30) was calculated to assess convergence by comparing within- and between-chain variability over the second half of each chain. This R statistic will be greater than 1 if the starting values are suitably over-dispersed; it will tend to one as convergence is approached. In general practice, if $R < 1.05$, we might assume convergence has been reached. The MCMC simulation for each model parameter was examined using the R statistic. The converged phase of the MCMC simulation for each model parameter of interest was identified for inferences.

Ideally, models should be checked by comparing predictions made by the model to actual new data. While data generated using new formulations were reported in the literature (31), these authors did not deconvolve that new dataset. Rather, they attempted to predict *in vivo* profiles for the new formulations based upon their *in vitro* dissolution profiles and the IVIVC generated with the same dataset used in this evaluation. Because we have reason to believe that unlike their original study, the underlying data reported by (31) included subjects that were poor metabolizers per our observation, we concluded it to be inappropriate to use the data from (31) for an external validation of our model. Accordingly, in the absence of data generated with a new formulation, the same data were used for model building and checking with special caution. Note because the predictions of Y , the *in vivo* dissolution profiles, were based on the observed *in vitro* data, deconvolved *in vivo* data, an assumed model, and upon posteriors that were based upon priors, this process involves checking the selected model and the reasonableness of the prior assumptions. If the assumptions were adequate, the predicted and the deconvoluted data should be similar. We compared the predicted and deconvoluted *in vivo* dissolution profiles to the corresponding observed *in vitro* dissolution data in Fig. 2.

Fig. 2. Estimated and deconvoluted *in vivo* vs *in vitro* dissolution profile

Red solid line denotes the estimated mean *in vivo* dissolution profile, blue solid lines denote the lower and upper bounds of the 95% credible intervals, and the black stars denote the deconvoluted *in vivo* dissolution profiles. Although there are some observations that fall below bounds as defined by the 95% credible interval, most of the observations are contained within those bounds.

To address the concern on using the same data for both model development and validation, a cross-validation approach was used to validate the established model. We randomly removed certain data points from the dataset and used the remaining data set for model development. Further, the removed data points were used to validate the model. For example, remove the data points for the combination of subject and formulation, *ij*, and calculate the residual vector, Residual [*ij*], of which each element is defined as

$$\text{Residual}[t, ij] < -Y_{\text{predi}}[t, ij] - Y_1[t, ij], \text{ for } t = 1 \text{ to } 9,$$

where Y_{predi} is the vector of predicted values at the individual level and Y_1 is the vector of removed data points for the combination of subject and formulation, *ij*. A boxplot of the residual vector by sampling time for Subject 1, with formulation “Med”, is used to show how close the predicted values from the established model are to the removed data points as in Fig. 3.

As shown in Fig. 3, residuals across the sampling time points do not significantly deviate from zero. Thus, it is concluded that the model established can predict the deconvoluted values with acceptable coverage and slightly inflated precision.

Summary of Posteriors

The Bayesian tolerance intervals were calculated based on the posteriors of the shape and scale parameters of the Weibull distribution at each sampling time and at each

Fig. 3. Boxplot of residuals

Fig. 4. Summary of distributions of posterior mean of scale parameter, $Mut[t, \cdot]$, which is derived via averaging over each subject and formulation at each time point

subject-formulation-sampling-time combination. The posteriors for the shape and scale parameters of the Weibull distribution were summarized *via* grouping by sampling time with respect to mean and 95% credible interval. The results are presented as in the forest plot (Fig. 4) for the scale parameters and as in the forest plot (Fig. 5) for the shape parameters. As shown in Figs. 4, 5, and 6, the distributions of the scale and shape parameters vary across the sampling time points. The distributions for both the

Fig. 5. Posterior distribution of scale parameter (Mut) for Subject 1 across the three formulations

Fig. 6. Summary of posterior distributions of shape parameter (r)

parameters at the first and the second time points are dramatically different from the ones for the rest of the sampling time points. In addition, the last 1000 MCMC simulation values of the model parameter of interest were saved for each parameter for establishing tolerance intervals later.

Prediction of *In Vivo* Dissolution Profile with *In Vitro* Dissolution Data

The predictive distribution of the *in vivo* dissolution profile was estimated with the established Bayesian hierarchical model. The Markov chain Monte Carlo

Fig. 7. Predicted and deconvoluted *in vivo* vs *in vitro* dissolution proportions

(MCMC) samples were generated from the posterior means of the model parameters with respect to each observed *in vitro* dissolution data point. The predictive distribution of the *in vivo* dissolution profile was characterized with respect to mean, and 95% predictive lower and upper limits at each sampling time point with the MCMC samples. As an example, the *in vivo* data for formulation “Med” and Subject 1 was assumed “unknown.” The predictive distribution of the *in vivo* profile for formulation “Med” and Subject 1 was summarized and shown in Fig. 7 with respect to mean (read line) and 95% lower and upper predictive limits (blue lines). In addition, the deconvoluted *in vivo* profile for formulation “Med” and Subject 1 (black stars) was also included to assess the predictive performance of the established Bayesian hierarchical model. As shown in Fig. 7, the deconvoluted *in vivo* dissolution proportions are close to the predicted means at each time point and fall into the 95% prediction interval. This symbolizes that the selected model can interpret the data sufficiently. Note that unlike a credible interval, which corresponds with the posterior distribution of a quantity of interest per the observed data and the prior information, the prediction interval corresponds with the predictive distribution of a “future” quantity based on the posteriors.

Bayesian Tolerance Intervals with Matching Priors

Random samples at size of 100 were generated from the Weibull distributions defined by the 1000 sampled posteriors of the shape and scale parameters at each sampling time and at each subject-formulation-sampling-time combination. Accordingly, two-sided Bayesian tolerance intervals with 90% content and 90% confidence for the *in vivo* dissolution profile were calculated using the approach by (26) at both the population and the individual levels. The results were plotted as in Figs. 8 (population level) and 9 (individual level). Note that the individual level inferences were based on the posteriors at the subject-by-formulation level, that is, using each set of $r[t]$ and $Mut[t, ij]$ to obtain Y_{pred} , as described in Fig. 1.

The comparison of these results underscores the importance of generating statistics at the individual rather than the population level when considering the IVIVC likely to occur in terms of the individual patient.

As shown in Fig. 8, the tolerance intervals generated at the population level cannot cover all the observations at each sampling time point. In seven out of nine time points, the 90% credible intervals at the population level are shorter than the corresponding Bayesian tolerance interval with 90% content and 90% confidence at the population level. The bounds of the credible intervals are directly related to the posterior distributions of the scale parameter (Mut) from Fig. 4 and the shape parameter (r) as shown in Fig. 6.

As shown in Fig. 9, the 90% individual tolerance interval succeeded in covering the observations from Subject 1 dosed with formulation “Fast”. Similarly, the 90% individual credible interval can cover the observations and is shorter than the corresponding population credible interval. As the variation decreases in the later sampling time points, the two-sided Bayesian tolerance intervals at either the population or the individual levels overlay

Fig. 8. Two-sided tolerance intervals (90% content and 90% confidence) for the *in vivo* dissolution profile in proportion (%) at the population level. Black open dots denote the deconvoluted *in vivo* dissolution profile in proportion; black bars denote the lower and the upper bounds of the two-sided Bayesian tolerance interval with 90% content and 90% confidence at the population level; red dotted bars denote the lower and upper limits of the 90% credible interval at the population level

with the credible intervals. However, the two-sided Bayesian tolerance intervals at the population level could be markedly narrower than the corresponding ones at the individual level at the earlier sampling time points due to the larger variation seen at the early time points. A similar trend is also shown in the credible intervals. In addition, the two-sided Bayesian tolerance intervals at the individual level are similar to the credible intervals. In general, the population credible intervals are shorter than the corresponding Bayesian tolerance intervals. The bounds of the credible intervals are directly related to the posterior distributions of the scale parameter (Mut) from Fig. 4. The same shape parameter (r) at each sampling time point as shown in Fig. 6 is shared when deriving the credible and tolerance intervals at the individual level.

DISCUSSION

Biological Interpretation of Analyses Results

The proposed method depends solely upon the observed *in vitro* dissolution and deconvoluted *in vivo*

dissolution profiles, avoiding direct interaction with the deconvolution/reconvolution process. Per the posterior distributions of the scale parameters for the Weibull model (Fig. 4), the variations of the parameters tend to decrease as the sampling time approaches maximum dissolution for any given formulation. It is greatest during periods of gastric emptying and early exposure to the intestinal environment. Similarly, given the relatively short timeframe within which these *in vivo* events occur, inherent individual physiological variability can lead to an increase in the variability associated with the deconvoluted estimates of *in vivo* dissolution. The noise is visualized in their posterior distributions and therefore there tends to be a wider credible interval associated with these early time points. Similar to the discussion associated with the scale parameters, the posterior distributions of the shape parameters (Fig. 6) reflect the inherent variability in the early physiological events that are critical to *in vivo* product performance.

As seen in Fig. 9, there may be situations where the upper bound of the tolerance limit will exceed 100%. This is

Fig. 9. Two-sided tolerance intervals (90% content and 90% confidence) for the *in vivo* dissolution profile in proportion (%) at an individual level (Subject 1, formulation “Fast”). Black open dots denote the deconvoluted *in vivo* dissolution profile in proportion; black bars denote the lower and the upper bounds of the two-sided Bayesian tolerance interval with 90% content and 90% confidence at the population level; black open triangles denote the lower and upper bounds of the two-sided Bayesian tolerance interval with 90% content and 90% confidence at the individual level; red dotted bars denote the lower and upper limits of 90% credible interval at the population level; red pluses denote the lower and upper limits of 90% credible interval at the individual level

inconsistent with a maximum deconvoluted *in vivo* value of equal to or less than 100%. As such, the scale parameter $Mu[t, ij]$ tends to be overestimated relative to the theoretical value, resulting in small odds for an upper limit greater than 100%. Nevertheless, since values of $Y_{pred} > 100$ lack biological relevance, it may be deemed appropriate to truncate the upper tolerance limit to 100% in these situations.

As seen in Figs. 2 and 8, there remain a few instances where observations fell outside the population tolerance and credible intervals. As values used as “observations” reflect the deconvolution estimates, this could either reflect the error in the estimated *in vivo* dissolution parameter (*i.e.*, resulting in deviations that might be considered experimental error), a bias in the *in vitro* dissolution dataset, failure to account for the need for time scaling, a need for modifications in our distribution assumptions, or a potential need to modify the current model.

Merits of the Approach

Bayesian hierarchical model is a powerful tool to incorporate multiple sources of variations into the analyses.

Particularly, with the Bayesian graphical modeling approach, it is straightforward to build a hierarchical model tailored to the need of particular study objectives, display the distribution over parameter space, and gain clear intuitions about how Bayesian inference works (especially for a complicated model). We implemented this approach in the evaluation of the “Level A” IVIVC data obtained from the work by Edington *et al.* (29) and demonstrated that inferences can be made at various levels of interest such as the population and the individual levels.

To overcome the controversy over using subjective priors in Bayesian analyses, we implemented probability matching priors in our analysis. The use of probability matching priors results in a Bayesian inference with frequentist validity. This feature connects the Bayesian and frequentist paradigms in a natural way, opens the dialog between these two fields, and addresses certain concerns on implementing the Bayesian approach to demonstrate treatment effects in drug approval. Further, covariates could be easily incorporated into the currently proposed model and make the model more suitable for certain scenario such as initial setting of IVIVC in Phase I, confirmation of IVIVC in Phase III, or IVIVC for a specific population. Moreover, the Bayesian IVIVC model can be

linked to not only tolerance limits but also to other inferences of interest.

Potential Applications and Considerations for its Novel Implementation to Overcome Hurdles in Data Analysis

• IVIVC

(a) *General Comments*: Formulation modification can occur throughout the lifetime of a drug product, ranging from changes instituted during early pre-approval stages to post-marketing changes. Typically, a determination of the *in vivo* impact of these modifications is addressed through a determination of blood level bioequivalence investigations. However, for human therapeutics, there are conditions under which *in vitro* dissolution data can be used to estimate product *in vivo* bioavailability characteristics, thereby supporting the approval of a new formulation (32). Furthermore, an IVIVC can support formulation development by predicting the targeted *in vitro* release characteristics necessary to achieve some targeted *in vivo* release profile.

The evaluation of an IVIVC has typically been based upon the use of a single dataset, oftentimes generated in a limited number of subjects who receive several formulations in a crossover study. The *in vitro* dissolution method reflects that which has the greatest *in vivo* prognostic capability based upon its relationship to the deconvoluted *in vivo* data. The IVIVC investigation usually consists of a relatively small number of subjects (*e.g.*, 10–24). As such, the power for detecting an IVIVC may be relatively low. Using multiple sets of IVIVC data can be a natural solution to this problem. Unlike the traditional linear regression approach for generating the IVIVC, the proposed approach can be easily modified to combine multiple sets of IVIVC data with relevant sources of variation incorporated into the analysis.

(b) *Benefits of Using This Bayesian Approach*: Typically, a “Level A” IVIVC involves the generation of a regression equation to describe the relationship between *in vivo* dissolution and *in vitro* dissolution. Inherent to this approach is an underlying assumption that variance is constant across all values of X (where in this case, X =the *in vitro* dissolution data). As we consider this assumption, it is important to keep in mind that for any given formulation, *in vitro* and *in vivo* dissolution are inextricably linked to time, thereby calling into question the validity of such an assumption as we consider the higher level of uncertainty often encountered during the early *vs* later time points. In other words, within any formulation, the impact of physiological variables (such as gastric pH, gastric emptying time, fluid volume within the gastrointestinal (GI) tract, *etc.*) often leads to a greater dispersion of the *in vivo* dissolution profile as compared to that occurring in the more distal portions of the GI tract (33, 34). This time-associated relationship in the magnitude of the variance can exist even in situations when the *in vivo* metric has been generated using

physiologically based pharmacokinetic (PBPK) models that have accommodated gastric emptying time into the *in vivo* predictions of product performance (35). Furthermore, in terms of the implementation of the F2 metric, it is recognized that there may be a greater magnitude of variability during the early *vs* later time points. The percent coefficient of variation can be as high as 20% at the earlier time points (*e.g.*, 15 min) but no greater than 10% at other time points, suggesting that there may also be time-associated differences in the variability from the formulation perspective (29).

The use of the Bayesian approach enables the description of subject-specific random effects and those covariates that can significantly affect the IVIVC. When utilizing a two-stage approach, those covariates would be incorporated when fitting the Weibull hierarchical model. From a slightly different perspective, when utilizing convolution-only-based techniques, the covariates can either be incorporated into the description of the subject specific random effects (*e.g.*, see Gaynor *et al.*, 2008) or included in the population description of the IVIVC as described in the two-stage approach.

While these differences have been ignored in the past (*i.e.*, when the IVIVC is defined by a regression equation), the novel approach described in this manuscript accommodates these potential fluctuations in the variability associated with the IVIVC relationship across time. This objective is accomplished by defining the relationship of percent dissolved *in vitro vs in vivo* dissolved (or fraction absorbed) as a series of relationships that is defined at each time point. In so doing, the deconvolution process converts *in vitro* dissolution data to a corresponding *in vivo* dissolution estimate, not by a singular regression equation but rather by the series of descriptors for the *in vivo/in vitro* relationship at each time point. Thus, the time-specific variability modeled by this method is more consistent with the site-specific variability known to exist across the numerous GI segments and the corresponding influence that these latent biological processes may have on *in vivo* product performance.

Use of the Bayesian approach also allows for the estimation of tolerance limits, providing predictions of the population distribution of *in vivo* product performance with some defined level of confidence. By inputting the tolerance limit estimates into mechanistic models, the resulting predicted range of *in vivo* dissolution profiles can be used to generate the population distribution of drug exposures likely to be achieved with a proposed formulation as described in Fig. 10. Thus, incorporation of the “Level A” IVIVC tolerance limits generated across a range of population proportions (*e.g.*, upper and lower 50, 60, 70 80, 90, 95, and 99% of the population, estimated with 90% confidence) into the PBPK model provides an opportunity to describe the distribution of exposures (or its corresponding pharmacodynamic (PD) consequences) likely to be achieved with a given formulation. This information can be invaluable for supporting drug

Fig. 10. Diagrammatic representation of the proposed approach for defining an IVIVC as a series of time-dependent relationships rather than as a single linear regression equation. The estimated IVIVC is then used for predicting the *in vivo* dissolution (or absorption) for a new formulation to infer how the formulation may perform in the targeted patient population

formulation assessment, be it from the perspective of drug development or drug regulation.

(c) *Applicability of Tolerance Limits:* Typically, the IVIVC is expressed in terms of means and confidence intervals. Since confidence intervals describe the uncertainty associated with estimates of a population mean, confidence intervals provide little information about the dispersion of *in vivo* dissolution across individual patients. If instead of confidence intervals, one tries to describe this dispersion through the use of estimated variation, there remains the problem of limitations imposed by the existing dataset, reliance on an assumption of a common and time-independent normal (or log-normal) distribution (despite the generation of these estimates on a relative small sample size), and the inability to ascribe some level of confidence to the targeted percentage of the population. This leads to the question of whether or not “average” is good enough and if not, what is the additional value provided by the development of statistical tolerance limits?

To answer this question, let us consider as an example, the increasing importance of carefully controlling systemic exposure when dealing with narrow therapeutic window drugs (36, 37). In these situations, the important question is not the “average” population exposure resulting from a new formulation but rather the proportion of individuals that may experience therapeutic failure or toxicity with a given formulation and dose. Recognizing the constraints associated with IVIVC predictions generated in normal healthy subjects, the use of tolerance intervals provides the opportunity to improve our predictions of population *in vivo* product performance. Incorporating other

sources of variability that can affect patient pharmacokinetics (e.g., polymorphisms or disease-associated variability in factors influencing clearance and distribution), these tolerance limits can be incorporated into PBPK models to describe the distribution of exposures likely to occur with a given formulation (based upon its *in vitro* dissolution data) across a patient population.

The following discussion shows how this population description can aid in formulation and dose selection in antimicrobial drug product development. For these anti-infective agents, dose is often linked to target attainment rate (38, 39). For example, the dose may be selected on the basis of bacterial susceptibility characteristics and the need to insure that 90% of the patient population will achieve the pharmacokinetic/pharmacodynamic (PK/PD) target. Akin to that described above for narrow therapeutic drugs, this concept can be used for formulation/dose assessment by integrating the *in vivo* dissolution into PBPK models to predict *in vivo* drug exposure across the population, determining if 90% of the population will likely achieve the targeted PK/PD metric. If confidence limits rather than tolerance limits are used in this assessment, a critical component of the prediction within a portion of population (e.g., ability to predict the distribution of *in vivo* dissolution characteristics within a portion of population) would be missing from this evaluation.

Thus, for some new formulations (or proposed set of *in vitro* dissolution characteristics), the development of tolerance limits can substantially improve our ability to predict the potential (portion of a population) for exposure-related likelihood of therapeutic success vs

failure and/or the exposure-associated safety concerns. Since the individual tolerance and credible intervals are subject-specific and unless we have information for a specific individual in mind (as per Fig. 9), the tolerance limits used for the reconvolution process would likely be based upon the population estimates at each sampling time of $r[t]$ and $Mut[t]$.

• Rare Disease Clinical Trial Evaluation

For clinical trials of rare diseases, such as amyotrophic lateral sclerosis (ALS) (40), it is challenging to conduct a large clinical trial and may be unethical to include a placebo group in such a trial. Borrowing information across the trials including sharing placebo arms becomes a viable solution to overcome the limitation of each individual trial. Databases for rare diseases have been developed to help scientists and clinicians to understand the rare disease, qualify biomarkers, validate clinical meaningful outcomes, and design better trials. The proposed approach can be extended for using the trials in the database collectively and for making decision based on the totality-of-evidence. Furthermore, to study certain types of disease mechanism such as neurodegeneration, borrowing information across diseases, and sharing certain disease commonalities may help in efforts to better understand the common disease modality. The current approach can be tailored to meet this goal as well.

With the passage of the Biologics Price Competition and Innovation (BPCI) Act in 2009, a new paradigm for approval of a biological product has been established to create an abbreviated licensure pathway for biological products. At the same time, this new paradigm triggers a need to integrate non-clinical or clinical data on the reference product from multiple historical studies into the evidence in a scientifically sound way, since in certain situations neither the randomized trials nor the single-arm trials alone will be enough to warrant reliable estimates of interest. One anticipated difficulty is to achieve an accurate estimate of a critical quality attribute (CQA) or a treatment effect given that a limited number of randomized trials is available relative to the number of single-arm studies. The proposed approach can be easily modified to integrate data from multiple studies given the assumptions on the extent of exchangeability of the studies. As such, this approach has the potential to be implemented in a biosimilar or interchangeable product application.

Questions for Consideration and Future Study

- (a) A first challenge is the need to explore some of the underlying reasons for some of the observations falling outside of the population tolerance and credible intervals. This work will necessitate further studies of both simulated and actual datasets. However, as shown in Figs. 2 and 8, this magnitude of this problem appears to be small and therefore should not negatively influence the applicability of this novel approach.

- (b) Type I error control has become one of the golden standards in validating a statistical approach for evaluating the treatment effects of a new drug. A Bayesian approach with probability matching priors can serve as an alternative way to control type I error. It has been well recognized that type I error control can be detrimental to the generation of conclusions based upon the use of Bayesian analyses because of the constraints it imposes on the use of Bayesian priors (41). Consequently, this has raised the question as to whether or not type I error control should be considered in Bayesian analyses. Furthermore, because the existence of probability matching priors is not guaranteed, Bayesian analyses face the hurdle of challenges in defining the priors to be used in the absence of probability matching priors. This is one of the benefits associated with the current proposed approach whereby a prior matching the highest posterior density region could be used as an alternative to the probability matching prior.

The performance of this new method will be tested by the users. Further feedbacks from the users will help improve the method.

CONCLUSIONS

A Weibull hierarchical model is used for evaluating the “Level A” IVIVC in a Bayesian paradigm and for the construction of a two-sided Bayesian tolerance interval with frequentist validity. A Bayesian hierarchical model is a powerful tool for incorporating multiple sources of variations and for accommodating potential fluctuations in the variability associated with the IVIVC relationship across time. This objective is accomplished by defining the relationship of percent dissolved *in vitro* vs *in vivo* dissolved (or fraction absorbed) as a series of relationships that is defined at each time point. In so doing, the reconvolution process converts *in vitro* dissolution data to a corresponding *in vivo* dissolution estimate, not by a singular regression equation but rather by the series of descriptors for the *in vivo/in vitro* relationship at each time point. Unlike the traditional linear regression approach for generating the IVIVC, the proposed approach can be easily modified to combine multiple sets of IVIVC data with relevant sources of variation incorporated into the analysis.

Corresponding tolerance limits generated with this method are based upon random samples generated from the (conditional) posterior distributions of the Weibull model parameters, followed by the use of the posterior means of these Weibull parameters and probability matching priors. The proposed method depends solely upon the observed *in vitro* dissolution and the deconvoluted *in vivo* dissolution profiles, avoiding direct interaction with the deconvolution/reconvolution process. Accordingly, it is equally applicable to one- and two-stage approaches for estimating the IVIVC.

Inter- and intra-subject variability of pharmacokinetics can exceed the variability between formulations, leading to IVIVC

models that can be misleading when based upon averages. The use of Bayesian approach enables the description of subject-specific random effects and those covariates that can significantly affect the IVIVC. The availability of tolerance limits about the IVIVC not only facilitates an appreciation of the challenges faced when developing *in vivo* release patterns but also is indispensable when converting *in vitro* dissolution data to the drug concentration vs time profiles across a patient population. By inputting the tolerance limit estimates into mechanistic models, the resulting predicted range of *in vivo* dissolution profiles can be used to generate the population distribution of drug exposures likely to be achieved with a proposed formulation. This information can be invaluable for supporting drug formulation assessment, be it from the perspective of drug development or drug regulation.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Meiyu Shen and Dr. Tristan Massie for their thoughtful comments and suggestions. We also would like to express our appreciation for the encouragement received from Drs. John Lawrence and Jim Hung.

APPENDIX

Modeling Software

OpenBUGS version 3.2.3 was used in hierarchical modeling with respect to model specification, diagnosis of model fit, summary of posteriors and generating posterior samples. Wolfram Mathematica version 9 was used to implement the method by Pathmanathan *et al.* (26) to establish the Bayesian tolerance intervals. With the program developed for this Bayesian approach, the implementation of this approach becomes much easier. Individuals wishing to obtain a copy of the codes used for generating the Bayesian two-sided tolerance limits in Mathematica, please contact Dr. Junshan Qiu at Junshan.Qiu@fda.hhs.gov.

Bayesian Tolerance Interval with Frequentist Validity

Both “*mu*” and “*r*” are in the positive domain. $g^{(n)}$ is calculated using the equation

$$g^{(n)} = n^{-1/2}g_1 + n^{-1}g_2 + O_p(n^{-3/2}),$$

here $g_1 = \frac{Mz_\gamma}{f^d + f^b}$ and $g_2 = \left\{ \frac{M}{f^d + f^b} \right\} \left\{ L_1(\pi) + L_2 + L_3(z_\gamma^2 - 1) \right\} + g_1^2 L_4$,

Where z_γ is the γ th quantile of the standard univariate normal distribution; $M = \left(c^{su} \hat{K}_s \hat{K}_u \right)^{1/2}$, where $s=1$ or 2 and $u=1$ or 2 , they denote the first and second element of the parameter vector for the Weibull distribution, $\theta = (\gamma, Mut)$; $c_{su} = -\left\{ \frac{\partial l}{\partial \theta_s} \frac{\partial l}{\partial \theta_u} \right\}_{\theta=\hat{\theta}}$, where $l(\theta)$ is the log likelihood; $\hat{K}_s = K_s(\hat{\theta}) = F_s(d(\hat{\theta}); \hat{\theta})^s - F_s(b(\hat{\theta}); \hat{\theta})$ and $\hat{K}_u = K_u(\hat{\theta}) = F_u(d(\hat{\theta}); \hat{\theta})^u - F_u(b(\hat{\theta}); \hat{\theta})$; $f^d = f(d(\hat{\theta}); \hat{\theta})$ and $f^b = f(b(\hat{\theta}); \hat{\theta})$; $L_1(\pi) = \frac{\hat{\pi}_s \lambda_s}{\hat{\pi}} \lambda_s$, $\hat{\pi} = \pi(\hat{\theta}) = \pi(\hat{\gamma}, \hat{Mut}) = \frac{1}{\hat{\gamma} \hat{Mut}}$, which is the prior

and has the probability matching probability for the Weibull distribution, $\hat{\pi}_s = \frac{\partial \pi(\hat{\gamma}, \hat{Mut})}{\partial \gamma} + \frac{\partial \pi(\hat{\gamma}, \hat{Mut})}{\partial Mut}$, and $\lambda_s = c^{su} \hat{K}_s$; $L_2 = \frac{1}{2}$

$$\left(\left\{ \frac{\partial}{\partial \theta_s} \frac{\partial}{\partial \theta_u} \frac{\partial}{\partial \theta_w} l(\theta) \right\}_{\theta=\hat{\theta}} c^{wu} \hat{K}_u c^{su} + c^{su} \frac{\partial}{\partial \theta_s} \frac{\partial}{\partial \theta_u} \left\{ \frac{F(d(\hat{\theta}); \hat{\theta}) - F(b(\hat{\theta}); \hat{\theta})}{M} \right\} \right)_{\theta=\hat{\theta}}; L_3 = \frac{1}{6} \left\{ \frac{\partial}{\partial \theta_s} \frac{\partial}{\partial \theta_u} \frac{\partial}{\partial \theta_w} l(\theta) \right\}_{\theta=\hat{\theta}} c^{su} \hat{K}_u c^{uu} \hat{K}_u c^{wu} \hat{K}_u + c^{su} \hat{K}_u c^{uu} \hat{K}_u \frac{\partial}{\partial \theta_s} \frac{\partial}{\partial \theta_u} \left\{ \frac{F(d(\hat{\theta}); \hat{\theta}) - F(b(\hat{\theta}); \hat{\theta})}{M} \right\}_{\theta=\hat{\theta}}; \text{ and } L_4 = \frac{1}{2} \left(\frac{\partial f(d(\hat{\theta}); \hat{\theta})}{\partial x} - \frac{\partial f(b(\hat{\theta}); \hat{\theta})}{\partial x} \right) / (f^d + f^b) - c^{su} \hat{K}_u \frac{\left(\frac{\partial f(d(\hat{\theta}); \hat{\theta})}{\partial x} - \frac{\partial f(b(\hat{\theta}); \hat{\theta})}{\partial x} \right)}{M}.$$

To include the remaining term $O_p(n^{-3/2})$ in the tolerance interval, the half of the interval is calculated (*HBTI*) as

$$HBTI = g_1 / (n^{0.5}) \times \exp((g_2 / g_1) / (n^{0.5})).$$

REFERENCES

1. Gaynor C, Dunne A, Davis J. A comparison of the prediction accuracy of two IVIVC modelling techniques. *J Pharm Sci.* 2008;97(8):3422–32.
2. Dunne A, Gaynor C, Davis J. Deconvolution based approach for level A in vivo-in vitro correlation modelling: statistical considerations. *Clin Res Regul Aff.* 2005;22(1):1–4.
3. Gillespie WR. Convolution-based approaches for in vivo-in vitro correlation modeling. *In Vivo-In Vitro Correlations* 1997 Jan 1 (pp. 53–65). Springer US.
4. Buchwald P. Direct, differential-equation-based in-vitro-in-vivo correlation (IVIVC) method. *J Pharm Pharmacol.* 2003;55(4):495–504.
5. Gould AL, Agrawal NG, Goel TV, Fitzpatrick S. A 1-step Bayesian predictive approach for evaluating in vitro in vivo correlation (IVIVC). *Biopharm Drug Dispos.* 2009;30(7):366–88.
6. Kakhi M, Chittenden J. Modeling of pharmacokinetic systems using stochastic deconvolution. *J Pharm Sci.* 2013;102(12):4433–43.
7. Kortjärvi H, Malkki J, Marvola M, Urtti A, Yliperttula M, Pajunen P. Level A in vitro-in vivo correlation (IVIVC) model with Bayesian approach to formulation series. *J Pharm Sci.* 2006;95(7):1595–605.
8. Wagner JG, Nelson E. Kinetic analysis of blood levels and urinary excretion in the absorptive phase after single doses of drug. *J Pharm Sci.* 1964;53(11):1392–403.
9. Loo JC, Riegelman S. New method for calculating the intrinsic absorption rate of drugs. *J Pharm Sci.* 1968;57(6):918–28.
10. Yu Z, Schwartz JB, Sugita ET, Foehl HC. Five modified numerical deconvolution methods for biopharmaceutics and pharmacokinetics studies. *Biopharm Drug Dispos.* 1996;17(6):521–40.
11. Süverkrüp R, Bonnacker I, Raubach HJ. Numerical stability of pharmacokinetic deconvolution algorithms. *J Pharm Sci.* 1989;78(11):948–54.
12. Cutler DJ. Numerical deconvolution by least squares: use of prescribed input functions. *J Pharmacokinet Biopharm.* 1978;6(3):227–41.
13. Lennernäs H, Aarons L, Augustijns P, Beato S, Bolger M, Box K, *et al.* Oral biopharmaceutics tools—time for a new initiative—an introduction to the IMI project OrBiTo. *Eur J Pharm Sci.* 2014;57:292–9.
14. Kesiosoglou F, Balakrishnan A, Manser K. Utility of PBPK absorption modeling to guide modified release formulation development of gaboxadol, a highly soluble compound with region-dependent absorption. *J Pharma Sci.* 2015.
15. Brockmeier D, Dengler HJ, Voegelé D. In vitro—in vivo correlation of dissolution, a time scaling problem? Transformation of in vitro results to the in vivo situation, using theophylline as a practical example. *Eur J Clin Pharmacol.* 1985;28(3):291–300.

16. Food, Administration D. Guidance for industry: extended release oral dosage forms: development, evaluation, and application of in vitro/in vivo correlations. Rockville: Center for Drug Evaluation and Research; 1997.
17. Morrison DF. Multivariate statistical methods. Singapore: McGraw-Hill; 1978. p. 128–34.
18. Dunne A, O'Hara T, Devane J. A new approach to modelling the relationship between in vitro and in vivo drug dissolution/absorption. *Stat Med*. 1999;18(14):1865–76.
19. O'Hara T, Hayes S, Davis J, Devane J, Smart T, Dunne A. In vivo–in vitro correlation (IVIVC) modeling incorporating a convolution step. *J Pharmacokinet Pharmacodyn*. 2001;28(3):277–98.
20. Cardot JM, Davit BM. In vitro–in vivo correlations: tricks and traps. *AAPS J*. 2012;14(3):491–9.
21. Mendell-Harary J, Dowell J, Bigora S, Piscitelli D, Butler J, Farrell C, Devane J, Young D. Nonlinear in vitro–in vivo correlations. In *In Vitro–in Vivo Correlations*. Springer US; 1997. pp. 199–206.
22. Aitchison J. Two papers on the comparison of Bayesian and frequentist approaches to statistical problems of prediction: Bayesian tolerance regions. *J R Stat Soc Ser B (Methodol)*. 1964;161–75.
23. Aitchison J. Expected-cover and linear-utility tolerance intervals. *J R Stat Soc Ser B (Methodol)*. 1966:57–62.
24. Aitchison J, Sculthorpe D. Some problems of statistical prediction. *Biometrika*. 1965;52(3–4):469–83.
25. Young DS. Tolerance: an R package for estimating tolerance intervals. *J Stat Softw*. 2010;36(5):1–39.
26. Pathmanathan D, Mukerjee R, Ong S. Two-sided Bayesian and frequentist tolerance intervals: general asymptotic results with applications. *Statistics*. 2014;48(3):524–38.
27. Wolfinger RD. Tolerance intervals for variance component models using Bayesian simulation. *J Qual Technol*. 1998;30(1):18.
28. Food, Administration D. Guidance for industry: Dissolution Testing of Immediate Release Solid Oral Dosage Forms. US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER). 1997. <http://www.fda.gov/downloads/drugs/guidance-compliance/regulatoryinformation/guidances/ucm070237.pdf>.
29. Eddington ND, Marroum P, Uppoor R, Hussain A, Augsburger L. Development and internal validation of an in vitro–in vivo correlation for a hydrophilic metoprolol tartrate extended release tablet formulation. *Pharm Res*. 1998;15(3):466–73.
30. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434–55.
31. Mahayni H, Rekhi GS, Uppoor RS, Marroum P, Hussain AS, Augsburger LL, *et al.* Evaluation of “external” predictability of an in vitro–in vivo correlation for an extended-release formulation containing metoprolol tartrate. *J Pharm Sci*. 2000;89(10):1354–61.
32. Food, Administration D. Guidance for industry: extended release oral dosage forms: development, evaluation, and application of in vitro/in vivo correlations. Center for Drug Evaluation and Research, Rockville. 1997. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070239.pdf>.
33. Chawla G, Bansal A. A means to address regional variability in intestinal drug absorption. *Pharm Technol*. 2003;27:50–68.
34. Kovačič NN, Pišlar M, Ilić I, Mrhar A, Bogataj M. Influence of the physiological variability of fasted gastric pH and tablet retention time on the variability of in vitro dissolution and simulated plasma profiles. *Int J Pharm*. 2014;473(1):552–9.
35. Kambayashi A, Blume H, Dressman J. Understanding the in vivo performance of enteric coated tablets using an in vitro–in silico–in vivo approach: case example diclofenac. *Eur J Pharm Biopharm*. 2013;85(3):1337–47.
36. Venitz J. Using exposure–response relationships to define therapeutic index: a proposed approach based on utility functions. Available from URL: www.fda.gov/ohrms/dockets/ac/02/slides/3898S1_06_Venitz.pdf.
37. Good DJ, Hartley R, Mathias N, Crison J, Tiruchera G, Timmins P, *et al.* Mitigation of adverse clinical events of a narrow target therapeutic index compound through modified release formulation design: an in vitro, in vivo, in silico, and clinical pharmacokinetic analysis. *Mol Pharm*. 2015.
38. Muller AE, Schmitt-Hoffmann AH, Punt N, Mouton JW. Monte Carlo simulations based on phase I studies predict target attainment of ceftibiprole in nosocomial pneumonia patients: a validation study. *Antimicrob Agents Chemother*. 2013;57(5):2047–53.
39. Dudley MN, Ambrose PG. Monte Carlo Simulation of new cefotaxime, ceftriaxone and cefepime susceptibility breakpoints for *S. pneumonia*, including strains with reduced susceptibility to penicillin. In Abstracts of the 42nd ICAAC, San Diego, CA. 2002:Abs-635.
40. Ringel S, Murphy J, Alderson M, Bryan W, England J, Miller R, *et al.* The natural history of amyotrophic lateral sclerosis. *Neurology*. 1993;43(7):1316–6.
41. Rosenblum M, Liu H, Yen E-H. Optimal tests of treatment effects for the overall population and two subpopulations in randomized trials, using sparse linear programming. *J Am Stat Assoc*. 2014;109(507):1216–28.