Beyond exposure-response: A tutorial on statistical considerations in dose-ranging studies

Glen Laird | **Lei Xu** | **Meng Liu** | **Jin Liu**

Vertex Pharmaceuticals Incorporated, Boston, Massachusetts, USA

Correspondence
Glen Laird, Vertex Pharmaceuticals Incorporated, 50 Northern Avenue, Boston, MA 02210, USA.
Email: glen_laird@vrtx.com

Abstract

Dose-ranging studies are a crucial part of the phase II drug development process. They complement the understanding gained from exposure-response analyses. However, the statistical design issues related to dose-ranging studies are not always keenly understood and a poorly designed study can be costly for later development. In this tutorial, we review five key statistical principles in designing such a study. We also describe some popular statistical approaches, including pairwise comparison, modeling, and Multiple Comparison Procedure modeling in the context of principles.

It is a simple fact of drug development that phase III failure is common.[1] Although rates vary by therapeutic area, it is generally accepted that the overall failure rate is about 50%. This rate is seen despite phase III studies generally being carefully planned and powered studies. What are the sources of failure? A common reason cited is using the wrong dose.

A dose-ranging study complements at least two other key pieces of information regarding the choice of dose for phase III. One of these is the first-in-human study, which provides the sponsor an understanding of the safety of the drug over a range of doses and may define a maximum tolerated dose (MTD). The other is exposure-response modeling, which should give the sponsor some idea what exposure levels of the drug are needed to safely modulate the intended target of interest and, hopefully, have efficacy. It seems more plausible that an optimal concentration exists for a given patient than an optimal dose and the therapeutic window for a drug is inexorably linked to the concentration. Patient-to-patient variability in exposure may explain a considerable portion of the response variability at a given dose. We hope, by understanding the relationship among dose, concentration, and efficacy, to arrive at a reasonable dose for the population of interest. In practice, it is usual to include only one dose (or at most 2 doses) in a phase III study, so a rigorous bottom-line assessment of different

doses is likely needed before the major resources of a phase III study are committed. Of course, there is sometimes need for dose adjustments in specific subpopulations. Although the "single dose for all" paradigm is preferred, it does not always apply. Somewhat contrarian to the highest-dose question, defined in some cases by safety findings in the first-in-human study, the phase II dose-ranging study can help us understand the lower end of the dose spectrum with respect to target engagement, pharmacodynamic (PD) response, or efficacy. We highlight some principles and some considerations related to the statistical design of such a study.

PRINCIPLE 1: KNOW THE MAIN PURPOSE OF THE STUDY

A dose-ranging study could have a primary analysis based upon one of several different objectives. The statistical framework (e.g., power) may be different depending on the intended goal(s), even for the same doses and sample sizes, so it is important to understand which case(s) are relevant to a given study. An example given later will highlight different design choices and the associated changes in statistical power and inferential conclusions.

One goal in a dose-ranging study may be to establish proof of concept (POC) by simply showing that all doses (including placebo) are not the same. In this context, it may make sense to establish a plausible model that succinctly parameterizes a dose/effect curve. A statistically significant result on the correct parameter(s) would indicate that the dose-response curve is not a flat line and establish POC. If the modeling is done correctly, this may be possible to do with a relatively small sample size. A simple example of this would be the classic linear trend test in a situation in which the dose-response relationship actually is linear. Beyond statistical testing, one can hope that estimating model parameters well allows a sufficiently precise understanding of the dose-response relationship. A disadvantage of this approach is that it does not necessarily have statistical power to support a statement regarding the value of any particular dose. Another study may be necessary to have confidence to select particular doses. In addition, key modeling assumptions should not be strongly violated or else bias can be introduced. For these reasons, one could consider a larger study examining a series of comparisons between each dose and placebo. However, such a study may have issues with multiplicity. Another approach to consider if the correct form of the model is uncertain is multiple comparisons procedure and modeling (MCP-Mod). More details on MCP-Mod are provided under Principle 3.

Another goal in phase II may be to establish a minimum effective dose (MinED). This concept is often framed in terms of the smallest dose sufficiently superior to placebo or the lowest effective dose. Zhou et al.[2] offer a good overview of this approach with a comparison.

Alternatively, one could attempt to establish POC by showing that one particular dose (e.g., the highest dose) is superior to placebo. Other doses are included largely as a backup in case the top dose has safety problems or to facilitate dose-exposure and/or exposure-response modeling. This approach allows for an efficient pairwise comparison between the target dose and placebo. However, data from the other doses may be "wasted" in the sense of not contributing to the primary analysis. In addition, there could be type 1 error inflation if the top dose is not statistically significant from placebo or has safety problems precluding its selection, but another dose is selected based upon its unadjusted p value (see Principle 3). For these reasons, we generally would not recommend this approach unless one is confident that the particular dose chosen is both safe and among the most active.

One of the more difficult challenges in phase II is choosing an optimal dose among several efficacious doses all of which are superior to placebo. One may be pushed into this context by ethical constraints that make lower doses problematic. The approach here may be the same as the placebo-controlled context, but the relevant effect sizes are likely to be considerably smaller compared with each other (among the active doses). Attempting such a study without proper statistical power can lead to faulty decisions in phase III. For example, in a phase II randomized

dose-ranging study of the JAK-2 selective inhibitor fedratinib (SAR302503),[3] 31 patients with intermediate-2 or high-risk myelofibrosis were randomized (1:1:1) to receive fedratinib 300, 400, or 500 mg once daily in consecutive 4-week cycles. Note the range of doses is quite narrow, perhaps related to ethical concerns in a serious disease, leading to considerable overlap in exposures. The study was undertaken in light of phase I results, "to further explore the clinical activity, safety, pharmacokinetics, and pharmacodynamics of fedratinib" at the selected doses. In this study, the primary efficacy end point was the percentage change in spleen volume evaluated by magnetic resonance image at the end of cycle 3 (week 12) from baseline. Although 31 patients may have been sufficient to detect a drug effect over placebo, it was not enough to fully power an efficacy comparison between these doses. The 500 mg dose showed somewhat better efficacy than 400 mg, but both 400 mg and 500 mg were taken forward to a phase III study. However, the 500 mg dose efficacy in phase III was similar to the 400 mg group and ultimately had safety problems, which paused development.[4]

PRINCIPLE 2: A WIDE RANGE OF DOSES IS OFTEN BENEFICIAL

To better characterize a dose-response relationship and identify the MinED that could help design the phase III study, it is recommended to explore a wide dose range of test drug. Assuming the maximum dose has been defined based on safety findings in phase I or preclinical safety margins, the question then becomes how low a dose level should be explored. Considering the wide range of possibilities the shape of a dose-response curve could take, adding a sufficiently low or subtherapeutic dose level is recommended to make the study robust enough to define the dose-response curve. To better demonstrate this, we assume there are two different types of monotonic dose-response curves that have the same defined maximally safe dose (Figure 1).

For a dose-response study with dose ranged from dose one to maximum dose, the wide range of doses is helpful to detect the true dose-response relationship, no matter what the shape of dose-response curve looks like (curve 1 or 2).

For a dose-response study with dose ranged from dose 2 to maximum dose, it may still be possible to detect the true dose-response relationship if curve 2 represents the true relationship. However, if curve 1 represents the true relationship, a flat dose response may be observed, and the establishment of dose-response relationship and identification of MinED may fail. Analogously, the estimation of key parameters (e.g., 90% effective dosage [ED_{90}]) are also affected by the choice of dose range. Part of the study design stage could include investigation (e.g., simulations) of this relationship.

Ting et al.[5] used studies for treatment of osteoarthritis as an example to show the importance of exploring a sufficiently low dose level [a]: a total of three dose-response

Dose-Response Curve
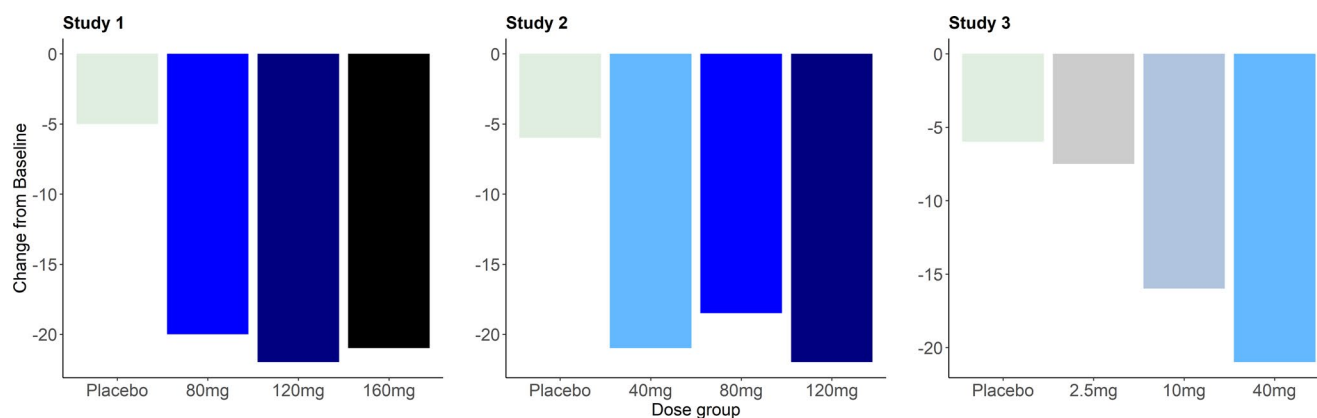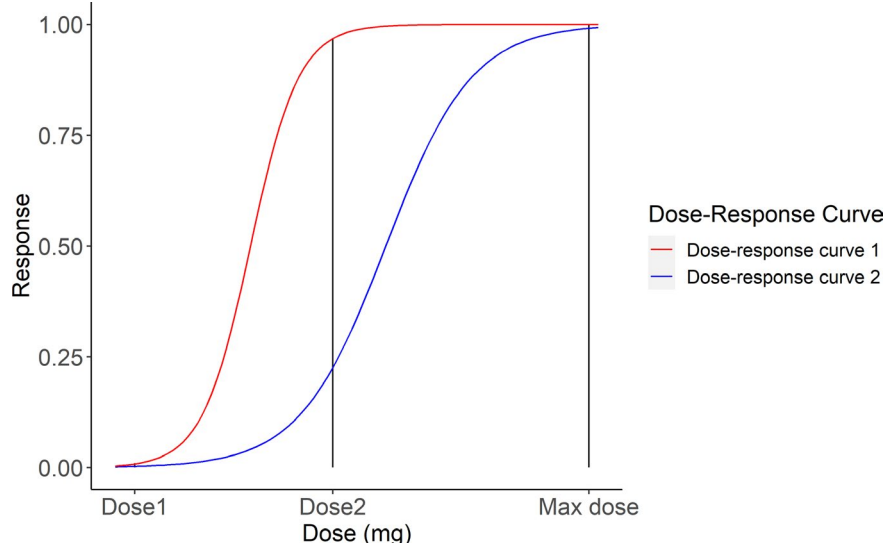— Dose-response curve 1
— Dose-response curve 2

FIGURE 2 Results of three dose-response studies

studies for treatment of osteoarthritis were performed. The results are summarized in Figure 2. For both study 1 (dose range: 80–160 mg) and study 2 (dose range: 40–120 mg), all doses explored were efficacious compared with placebo, but no significant difference in response were detected between the doses. Study 3 (dose range: 2.5–40 mg), with much lower doses explored, successfully demonstrated different responses between doses to help establish the dose-response relationship. This example points to the potential for additional cost and time wasted if sufficiently low doses are not included at the very beginning of phase II.

To demonstrate how lower doses improve the establishment of the dose-response curve in this example, we compared the dose-response curve (fitted by maximum effect [E_{max}] model) based on results from all three studies (dose range: 2.5–160 mg) versus results from study 1 and study 2 only (dose range: 40–160 mg).

When all three studies are included, as well as placebo, the fitted dose-response curve (curve 1 in Figure 3) increases sharply to the top.

With study 1 and 2 only (placebo and dose ranging from 40 to 160 mg), the fitted curve (curve 2 in Figure 3) tends to be smoother and reaches the maximum effect more slowly. It also gives a higher estimated E_{max}. Consequently, estimation based on curve 2 may be biased. For example, the dose that produces ED_{90} attributable to the drug was overestimated by ~ 3-fold based on curve 2 (i.e., ED_{90_2}) compared with curve 1 (i.e., ED_{90_1}).

Because of the importance of including a sufficiently low dose level in a dose-ranging study, multiple types of dose spacing design were proposed to determine the choices of low doses. Ting[6] has highlighted the advantages of using binary dosing spacing (BDS) design, which allocates more doses to the lower end of range and helpful in identifying MinED. For example, assume the MTD of 100 mg was already identified, and a dose-range study needs to include three test doses (low, medium, and high doses) between placebo (0 mg) and the MTD (100 mg). Following the rules of BDS, a midpoint will be picked between placebo and the maximum dose (i.e., 50 mg), and the high dose will be allocated between this midpoint and

FIGURE 3 Dose-response relationships including placebo data: ED_{90}_1 and ED_{90}_2 represents the estimated dose that produces the 90% of the maximum effect attributable to the drug based on curve 1 and curve 2, respectively
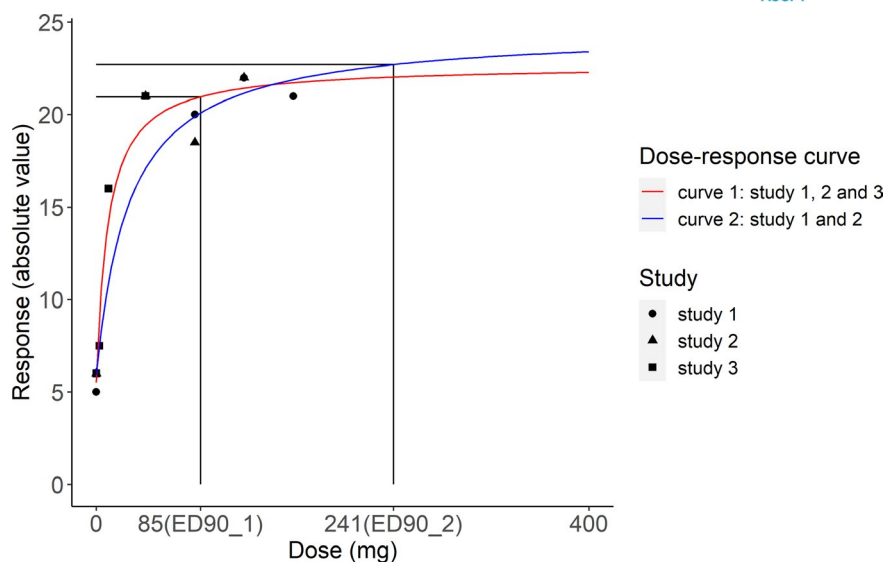
MTD (i.e., between 50 and 100 mg). A second midpoint will then be picked between placebo and the first midpoint, which will be at 25 mg. The low dose will then be allocated between placebo and the second midpoint (i.e., between 0 and 25 mg). The medium dose will be allocated between the first and second midpoint (i.e., between 25 and 50 mg).

In addition, a rule of thumb that forces the dose range (i.e., the ratio of the highest tested dose over the lowest tested dose) to be at least 10-fold in the first dose-ranging study can help ensure a sufficiently low dose to be included in the study.[6]

PRINCIPLE 3: SELECT THE ANALYSIS METHOD WITH CAUTION

Evaluation of dose-response relationship of the investigational drug is often stated as the primary objective in the protocols of such dose-ranging studies and therefore the sponsor needs to prespecify a primary analysis on a selected primary end point, which is usually an efficacy end point or a biomarker end point but could also be a joint utility of both efficacy and toxicity (please see Principal 4).

Obviously, there are a large variety of statistical approaches available even for the same stated primary objective and the same primary end point, assuming the same true underlying dose-response relationship. They all serve the purpose of dose-response shape detection; however, caution should be urged in statistical analysis planning because how to analyze the study data could still have a significant impact on the study conclusion.

Preclinical and biological information as well as phase I results have always been fully leveraged to facilitate the design and analysis of a dose-ranging study. However, not every therapeutic area or every drug ends up with accurate predictions on the projected dose-response curve. For example, a linear trend test is a conventional statistical tool in a dose-ranging study but

whether it is powerful depends on the proposed dose range in which the plateau portion may or may not be successfully predicted and covered. Its power may also be lost if any unexpected curve, such as U shape or inverted U shape, is observed.[7] In some complicated disease areas or due to limited preclinical evidence, it is preferable to account for some level of uncertainty, in which case MCP-Mod becomes a popular analysis approach. Finally, multiplicity adjustment could still be an issue statistically. Although the control of familywise type I error in a dose-response analysis (as usual in phase II) may not be as critical as in a pivotal phase III study,[7] the multiplicity issue still deserves attention, because whether a multiplicity adjustment will be applied or how it is applied could lead to different study conclusions.

For illustration purposes, a hypothetical randomized double-blind placebo-controlled parallel dose-ranging study is used here, which mimics a scenario in which the dose-response curve is monotonic in a general sense only, but not strictly increasing along with the doses (Figure 4). A total of 120 subjects are equally allocated to placebo and 4 active doses (100, 300, 600, and 900 mg). The primary efficacy end point is a continuous response variable and a larger value indicates a better outcome. Several popular statistical methods could be prespecified to serve as dose-response analysis, which is the primary objective of the study, as follows:

1. Without any multiplicity adjustment, a simple naïve analysis of covariance (ANCOVA) analysis to provide mean response for each group as well as treatment difference between each active dose and placebo.
2. An ANCOVA analysis with a conservative Bonferroni procedure, which tests each dose versus placebo using split level of significance.
3. An ANCOVA analysis with Dunnett's test[8] as a multiple testing procedure, which accounts for comparison of each active dose against the same control group.
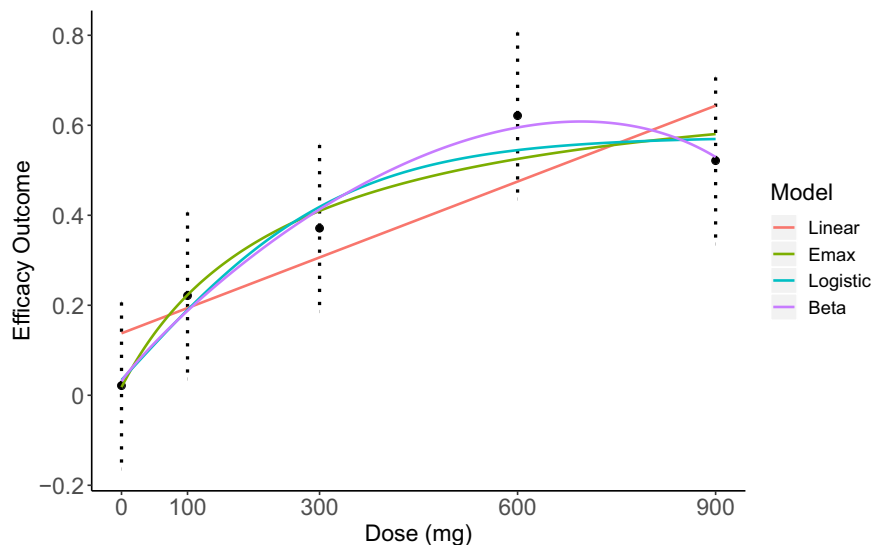
FIGURE 4 Efficacy outcomes of each treatment groups in the hypothetic study as well as four different dose-response curves fitted using multiple comparisons procedure and modeling method. Filled circles are mean responses for each group and dotted lines represent corresponding standard errors. E_{max}, maximum effect

TABLE 1 ANCOVA analysis results

Comparison	Estimate	SE	*t*-value	Nominal *p* value	Adjusted *p* value (Bonferroni)	Adjusted *p* value (Dunnett)
100 mg vs. placebo	0.20	0.262	0.764	0.4465	1.0000	0.8600
300 mg vs. placebo	0.35	0.262	1.337	0.1839	0.7356	0.4752
600 mg vs. placebo	0.60	0.262	2.292	0.0237	0.0948	0.0782
900 mg vs. placebo	0.50	0.262	1.910	0.0586	0.2344	0.1788

Abbreviation: ANCOVA, analysis of covariance.

4. A linear trend test through a linear contrast of mean response with no multiplicity adjustment.
5. MCP-Mod[9] with a set of four dose-response candidate models (linear, E_{max}, logistic, and beta).

Methods 1, 2, and 3 conduct dose-response analysis through the conventional pairwise comparison, whereas methods 4 and 5 directly test certain dose response signals. Method 4 is a single test, which does not need any multiplicity adjustment, whereas methods 2, 3, and 5 all adopt some multiplicity adjustment with different focuses. Both methods 2 and 3 adjust for the multiple pairwise comparison of each active dose against the control, whereas the multiple comparison procedure component of MCP-Mod accounts for the multiple tests on a set of several different prespecified doses-response candidate models.

Among these statistical methods, the MCP-Mod approach is a special unified approach, which performs response signal testing with multiplicity adjustment as well as estimation of optimal dose through modeling techniques. A few initial guesses of model shapes are needed as dose-response candidates (4 are used in this example). The first step in MCP-Mod usually serves as the primary statistical inference with p values provided. A significant response signal is claimed if one or more prespecified models associate a significant adjusted p value. The second step could continue to leverage either

TABLE 2 MCP-Mod analysis results

Dose-response model	Test statistic	Adjusted *p* value (MCP-Mod)
Linear	2.248	0.0554
Emax	2.485	0.0312
Logistic	2.301	0.0485
Beta	2.342	0.0444

Abbreviation: MCP-Mod, multiple comparisons procedure and modeling.

model averaging or modeling selection criterion to estimate any optimal dose, such as the minimal effective dose.[2] It is also notable that the "dose" level in the MCP-Mod method could be quite flexible and it could be any univariate continuous variable that is able to differentiate dosing groups in an ascending order. Such flexibility would allow the dose response analysis utilizing all various dosing regimens, such as once daily or twice daily in the same study. The study design, including power calculation and real data analysis, can be conducted using the R package DoseFinding.[10]

ANCOVA analysis results as well as p values based on methods 1–3 are summarized in Table 1. Method 3 had a linear trend test t value 2.392 with p value 0.0184. The MCP-Mod analysis results are shown in Table 2. Note that the p value for the linear model under MCP-Mod is over 0.05 and larger than

that of the linear trend test under method 3. The different study conclusions based on methods 1–5 above are summarized as follows. Obviously, which statistical approach is used to analyze the data plays a key role in drawing these conclusions.

1. ANCOVA without any multiplicity adjustment: 600 mg is the only dose significantly better than placebo.
2. ANCOVA with Bonferroni adjustment: no active dose is significantly better than placebo.
3. ANCOVA with Dunnett's test: no active dose is significantly better than placebo.
4. Linear trend test: significant dose-response evidence identified.
5. MCP-Mod: significant dose-response evidence identified with three of four candidate models with significant p values; linear model is not significant.

PRINCIPLE 4: MULTIPLE FACTORS MAY PLAY INTO THE DESIGN OF A DOSE-RESPONSE STUDY

Dosing regimen

In designing dose-ranging studies, we need to know the frequency that a patient takes the test drug. More frequent dosing (say over twice per day) may reduce patient compliance. In contrast, less frequent dosing may result in diminished drug efficacy when approaching the end of the dosing interval.[11] Usually, the phase I pharmacokinetic (PK)/PD findings drive the design of the dosing interval. One important PK parameter is the half-life of a test drug, which estimates the time-taken to metabolize half of the test drug out of the body and then helps to examine how long the compound will stay in the human body. As shown in Figure 5, q.d. dosing is favored

for a drug with a relatively longer half-life; otherwise, b.i.d. dosing may be needed to keep the plasma concentration of a drug above the minimum effective level (1 unit in this case). With such information, we can propose a dosing frequency for dose-ranging study design. It is possible to study more than one frequency of dosing in a single study, which is usually realized by a factorial design (dose, frequency, and dose*frequency), as Fraser et al.[12] did in their study.

For this type of study design to investigate dose/dosing frequency, we can use the Jonckheere-Terpstra trend test if a priori knowledge of the order of treatment effects by expected exposures is available.[13] Alternatively, this type of study design can be assessed by pairwise comparison, where statistical procedures also need to be well-defined and documented to ensure that the type I error rate is not inflated. Suppose a simple case in a dose-response study with 3 treatment arms: 600 mg q.d., 300 mg b.i.d., and 300 mg q.d. compared against placebo. Other than the comparison of each dose with the placebo, it is also of interest in comparing 2 arms with 600 mg daily dose. Accordingly, we may want to conduct at least four pairwise comparisons:

1. 600 mg q.d. versus placebo;
2. 300 mg b.i.d. versus placebo;
3. 300 mg q.d. versus placebo;
4. 600 mg q.d. versus 300 mg b.i.d.

Similar to the scenario shown in principle 3, the type I error rate inflates unless we conduct a multiple comparison procedure (MCP). ANCOVA analysis with Bonferroni correction (method 2 in principle 3) can be used to compare, in a pairwise manner, the treatment arms, whereas ANCOVA analysis with Dunnett adjustment (method 3 introduced in principle 3) only works for the first three comparisons because it can only be used to compare each treatment arm to the placebo.
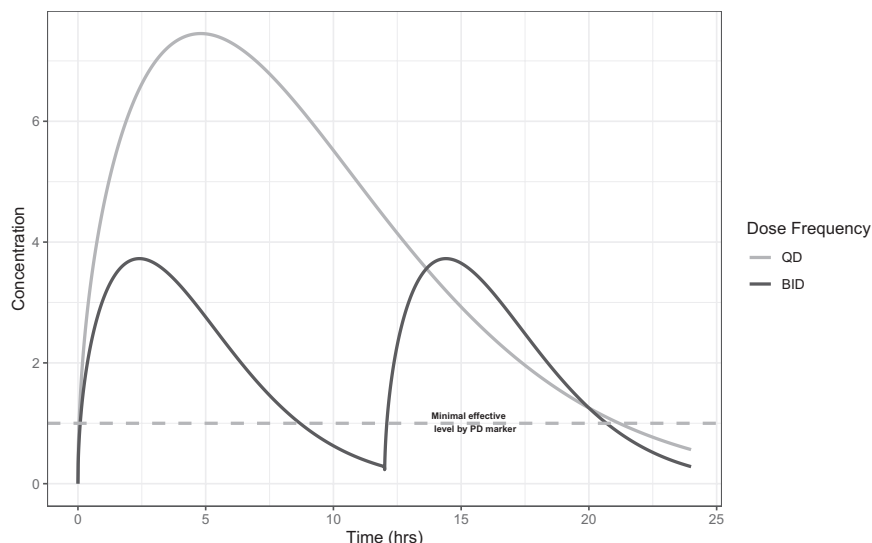

FIGURE 5 Plasma concentration of a drug with once a day and twice a day dosing

Additionally, the MCP-Mod approach presented in principle 3 is usually applied to determine the dose with a fixed frequency.[14] Suppose that all existing doses in the design of principle 3 are q.d., and we have 300 mg b.i.d. as an additional treatment arm. We cannot use the MCP-Mod to compare five treatment arms and the placebo directly. One possible solution is to combine dose and regimen into one model covariate when estimating the exposure-response relationship.[15] We may also either combine 600 mg q.d. and 300 mg b.i.d. (combining option) or exclude 300 mg b.i.d. (removing option) for the MCP-Mod analysis. The selection between "combining" and "removing" is based on the power calculation for each option with different assumptions for the relative treatment effect of 600 mg q.d. against 300 mg b.i.d. Specifically, for each option with each model and each prespecified relative treatment effect, we can calculate a power and then identify the minimum power for each option. If there is no significant difference between the minimum power of two options, the combining option is favored because the removing option excludes a treatment arm from the primary analysis. It is also recommended to consider the removing option for sensitivity analyses.

SAFETY

As introduced in the example detailed in principle 1, failure to identify unacceptable toxicity in phase II is one likely reason for an unsuccessful phase III study. There is a well-known asymmetry in all stages of drug development that favors more attention on the evaluation of efficacy rather than toxicity.[16] This asymmetry is reflected in the plans of the majority of phase II dose-finding studies: the prespecified dose-selection criteria is often based upon only the efficacy or a biomarker end point. Accordingly, it may be difficult to draw firm conclusions concerning drug toxicity due to the limited size of clinical data at the end of phase II.

Fortunately, several theoretical approaches based on the joint utility of efficacy and toxicity for dose-finding studies have been developed. For example, Yin et al.[17] proposed a Bayesian adaptive design to incorporate the bivariate outcomes, toxicity, and efficacy, of a new drug in early phase clinical trials. In this method, with the assumptions that both toxicity and efficacy are binary outcomes and that the toxicity probabilities are monotonic (constrained by a prior distribution), the relative degree toxicity against efficacy (i.e., the odds ratio) was used to quantify the association between them. In this adaptive design, the parametric functional form of the dose-response curve is not prespecified; instead, the dosage for the next cohort is determined by the observed data on the tried doses.

Similarly, Ivanova et al.[18] proposed to construct a utility function (also referred to as clinical utility index[19]) that incorporates the binary toxicity outcome (indicating the occurrence of an adverse event [AE]) with a continuous efficacy response. The utility function, which is determined by the prespecified weight of toxicity, could be various forms. They also developed two techniques to optimize the utility function. One is computing a maximum likelihood estimate of efficacy mean under the restrictions of a unimodal distribution of dose-efficacy function and a non-decreasing relationship of AE rate of interest with dose. The other method is for an adaptive design where a new individual is equally randomized to one of two designated adjacent doses. For each assignment, the difference in the utility function between the most recent designated pair of doses needs to be estimated, and the next designated pair is determined.

Additionally, as reported by Ivanova et al.,[17] we can model efficacy and toxicity separately but consider them jointly through prespecified utility functions. The utility function, which is based on a plausible efficacy and toxicity region provided by study teams, defines the balance between efficacy and toxicity. Study teams first provide an acceptable balance of efficacy and toxicity as a "base case," say treatment with the efficacy of 20 and AE rate of 0.1 in Figure 5 (note here, the AE rate is only one possible toxicity signal. When building the utility function, one can use other toxicity signals as appropriate). They then decide dose with the efficacy of 21 (17) and AE rate of 0.2 (0) is as good as the base case so that the weight of toxicity can be defined as 10 because 1 unit increase (decrease) in efficacy is balanced by 0.1 increase (decrease) in toxicity in the AE rate. That is, the toxicity weight can be defined as the slope of the efficacy-toxicity line. It is noted that the weight is sensitive to the unit of efficacy.

With this definition, the dose-response relationship would change from dose-efficacy relationship to dose-joint utility of efficacy and toxicity when both efficacy and toxicity are considered in trial design. Let us rethink the fedratinib example in principle 1 and build a utility function hypothetically with the primary efficacy end point and a binary toxicity outcome that indicates the occurrence of Wernicke's encephalopathy (WE; the treatment emergent adverse event that caused its on-hold status in phase III). Figure 6 demonstrates dose-utility curves with 4 different levels of prespecified weights of toxicity (0, 1, 2, or 3). When the toxicity weight is 0, the dose-utility curve reduces to the dose-efficacy curve. Otherwise, based on the definition, the weight can be interpreted as one percent decrease in spleen volume (that suggests the treatment is efficacious) is equivalent to 0.1, 0.2, and 0.3 increase in WE rate for the toxicity levels 1, 2, and 3, respectively. From Figure 6, it is clear that the utility of dose 500 mg decreases with an increased toxicity weight, suggesting that, with a prespecified toxicity rate, the utility function can help choose an optimal dose balanced between efficacy and safety signals. Note that Ouellet[19] pointed

FIGURE 6 Dose-utility curve with considering different levels of toxicity. WE, Wernicke's encephalopathy
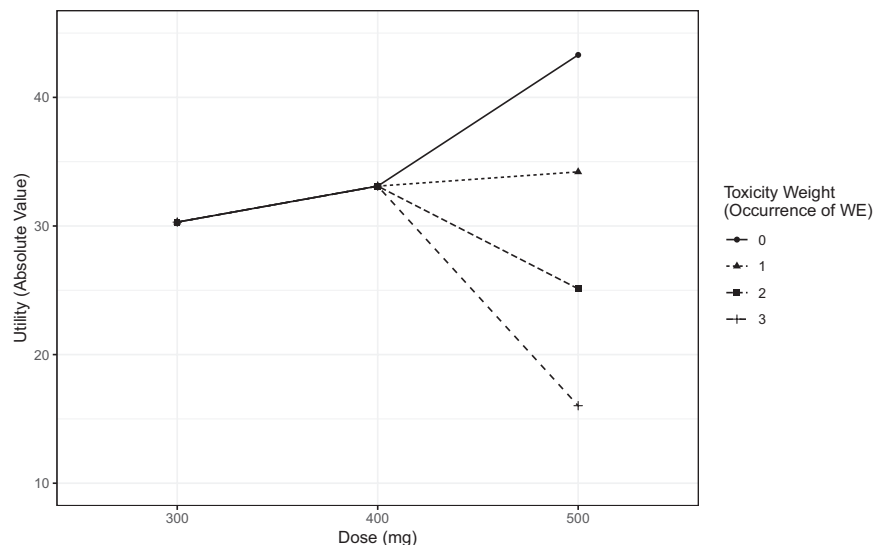
out that in addition to deciding an optimal dose that balances efficacy and safety, the utility function can be used with other considerations, such as pill burdens and cost of goods.

PRINCIPLE 5: DOSE-RESPONSE AND EXPOSURE-RESPONSE SHOULD BE VIEWED IN THE CONTEXT OF EACH OTHER

Besides the dose range, multiplicity, and other issues discussed in previous sections, one needs to take into consideration drug exposure. In situations where variability in PKs is limited, dose is generally a good predictor of drug exposure and the dose-response relationship aligns with the exposure-response relationship. In some situations, however, there can be extensive exposure variability in subjects administered the same dose. This can be due to a wide array of factors, such as differences in clearing organ (kidney or liver) function, body size, drug-drug interactions, or genomic differences in drug-metabolizing enzymes. For example, CYP3A is one of the most important drug-metabolizing enzymes in human beings. Studies have revealed that interindividual difference in expression and activity of CYP3A subfamily enzymes is large.[20] One can imagine that such difference may cause the dose-response relationship to look different than the exposure-response relationship; for example, two different doses may show similar responses if subjects taking the low dose happened to be "poor" drug metabolizers (i.e., low CYP3A activities due to the use of CYP3A inhibitors), thereby having similar drug exposure as those taking the high dose. A misleading conclusion could be made that the lower dose is as effective as the high dose. Ogasawara et al. demonstrated the importance of aligning dose-response and exposure-response results using the study of secukinumab in psoriasis as an example[21]; whereas a phase II dose-response

study supported the selection of 150 and 300 mg doses of secukinumab for a phase III study, additional exposure-response analysis showed that the drug exposure depended on body weight: patients with a weight greater than or equal to 90 kg had lower secukinumab concentration (exposure) and thereby a lower clinical response rate compared with patients with a weight less than 90 kg.[21]

In fact, Ogasawara et al. have found that exposure-response analysis tends to be more frequently used to support the dose labeling.[21] This again suggests that the exposure-response relationship may provide additional rationales for dose selection.

SUMMARY

We have provided a summary of some key principles in the statistical design of dose-ranging studies, including comments on power. Factors contributing to a gain or loss of power (other than the omnipresent sample size and effect size) include the doses and models chosen as well as the multiplicity approach.

CONFLICT OF INTEREST
The authors declared no competing interests for this work.

REFERENCES
1. Grignolo A, Pretorius S. Phase III trial failures: costly, but preventable. *Appl Clin Trials*. 2016;25(8/9):36-42.

2. Zhou Y, Chen SU, Sullivan D, et al. Dose-ranging design and analysis methods to identify the minimum effective dose (MED). *Contemp Clin Trials*. 2017;63:59-66.

3. Pardanani A, Tefferi A, Jamieson C, et al. A phase 2 randomized dose-ranging study of the JAK2-selective inhibitor fedratinib (SAR302503) in patients with myelofibrosis. *Blood Cancer J*. 2015;5:e335.

4. Pardanani A, Harrison C, Cortes JE, et al. Safety and efficacy of fedratinib in patients with primary or secondary myelofibrosis: a randomized clinical trial. *JAMA Oncol*. 2015;1(5):643-651.

5. Ting N. Practical and statistical considerations in designing an early phase II osteoarthritis clinical trial: a case study. *Commun Stat Theory Meth*. 2009;38(18):3282-3296.

6. Ting N. Confirm and explore: a stepwise approach to clinical study designs. *Drug Informat J*. 2008;42:545-554.

7. Cadavid D, Mellion M, Hupperts R, et al. Safety and efficacy of opicinumab in patients with relapsing multiple sclerosis (SYNERGY): a randomised, placebo-controlled, phase 2 trial. *Lancet Neurology* 2019;18(9):845-856.

8. Dunnett CW. A multiple comparisons procedure for comparing several treatments with a control. *J Am Stat Assoc*. 1955;50:1096-1121.

9. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. 2005;61(3):738-748.

10. Bornkamp B, Pinheiro J, Bretz F. DoseFinding: Planning and Analyzing Dose Finding Experiments. R package version 0.9-16; 2018. https://CRAN.R-project.org/package=DoseFinding

11. Good Review Practice: Clinical Review of Investigational New Drug Applications. https://www.fda.gov/media/87621/download

12. Fraser G, Lederman S, Waldbaum A, Lee M, Skillern L, Ramael S. The Neurokinin 3 receptor antagonist, fezolinetant, is effective in treatment of menopausal vasomotor symptoms: a randomized, placebo-controlled, double-blind, dose-ranging study. *J Endocrine Soc* 2019;3(Suppl_1):OR33–OR36.

13. Jonckheere A. A distribution-free k-sample test against ordered alternatives. *Biometrika*. 1954;41(1/2):133-145.

14. Musuamba FT, Manolis E, Holford N, et al. Advanced methods for dose and regimen finding during drug development: summary of the EMA/EFPIA workshop on dose finding (London 4–5 December 2014). *CPT Pharmacometrics Syst. Pharmacol*. 2017;6:418-429.

15. Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. *Statist Med*. 2014;33:1646-1661.

16. O'Neill RT. A perspective on characterizing benefits and risks derived from clinical trials: can we do more? *Drug Inform J*. 2008;42:235-245.

17. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*. 2006;62:777-787.

18. Ivanova A, Liu K, Snyder E, Snavely D. An adaptive design for identifying the dose with the best efficacy/tolerability profile with application to a crossover dose-finding study. *Statist Med*. 2009;28:2941-2951.

19. Ouellet D. Benefit-risk assessment: the use of clinical utility index. *Expert Opin Drug Saf*. 2010;9(2):289-300.

20. Wilkinson GR. (2018), Cytochrome P4503A (CYP3A) metabolism: prediction of in vivo activity in humans. *J Pharmacokinet Biopharm*. 1996;24(5):475-490.

21. Ogasawara K, Breder CD, Lin DH, Alexander GC. Exposure- and dose-response analyses in dose selection and labeling of FDA-approved biologics. *Clin Ther*. 2018;40(1):95-102.e2.