7
General Considerations in Dose–Response Study Designs

NAITEE TING

7.1 Issues Relating to Clinical Development Plan

As mentioned in Chapter 1, one important step in early clinical development of a new drug is to draft a clinical development plan (CDP). Various clinical studies are designed and carried out according to this plan, and the CDP is updated over time based on newly available information. Estimation of dose–response relationship should be one of the very important components in CDP.

Considerations and plans regarding dose finding should be in place starting from the nonclinical development stage. Across all phases of clinical development, information to help with dose selection is needed. The key stage for finding the appropriate range of doses should be around Phase II. But critical information to help design Phase II studies are obtained from nonclinical, and Phase I studies. In certain situations, the drug candidate belongs to a well-established drug class in which information from other drugs of the same class is available. Clinical scientists need to make best use of that available information to help design Phase II studies. Hence, one of the primary objectives in the earlier part of CDP should be to deliver useful data to help designing dose ranging and dose selection studies in Phase II. Based on information collected from Phase I clinical studies, a number of Phase II studies should be planned and carried out—proof of concept (POC), dose ranging, and dose-finding studies. Some of these studies are carried out to measure the clinical endpoints, while some others are implemented to characterize biomarkers. Choice of appropriate endpoints for each study should be considered in the CDP. Criteria to measure success should also be clarified in the CDP.

After the multiple dose pharmacokinetics (PK) is established for a drug candidate from Phase I studies, there is often an estimated Maximally Tolerated Dose (MTD). With the PK and MTD information available, a typical step to progress the drug development is to conduct a POC study. A commonly used POC study usually has two parallel treatment groups—a control (often placebo) group and a test treatment group using a high dose very close to MTD, or the MTD itself. In some situations, the test treatment group allows dose titration up to the MTD. The reason a very high dose (very close to MTD) is used for

POC is that the highest dose may provide the best hope to demonstrate drug efficacy.

Dose ranging studies usually include a placebo group, plus a few doses of the test drug—e.g., low dose, medium dose, and high dose. An ideal dose ranging study should cover a wide range of doses from low to high. Typically, these studies are parallel group with fixed doses. The main objective of a dose ranging study is to estimate the dose–response relationships for efficacy, and possibly for safety. Hence, in analyzing results from these studies, various dose–response models are often applied to help understand the underlying dose–response relationship.

On the other hand, dose selection studies or dose finding studies are mainly designed to confirm the efficacy of one or several doses. Although the design of a dose selection study is very similar to a dose ranging study (with placebo or active control, plus a few test doses), the data analysis tends to be hypothesis testing of each test dose against the control.

In the CDP, considerations should be made to determine whether studies could be conducted simultaneously or sequentially. In other words, in trials designed to study PK, this study can also provide safety information to help estimate MTD. Meanwhile, another study can be designed to learn the food effect. In these situations, we should try to maximize the amount of information that can be collected in each study, and minimize the time to achieve these objectives. On the other hand, a POC study or a dose ranging study cannot be designed without MTD information. Hence, these studies should be conducted after MTD information can be obtained from earlier studies. Therefore, the CDP needs to lay out the sequence of studies to be designed and executed over time. Estimation of the starting time of a new study should be based on critical information available to help design that study. In some cases POC and dose ranging studies are combined and in others dose ranging studies and dose selection studies are combined. All of these strategies need to be discussed while drafting the CDP.

Section 7.2 introduces some general considerations in designing clinical trials (not just for dose finding purposes). Section 7.3 discusses design considerations specifically for dose finding trials, and Section 7.4 provides concluding remarks.

7.2 General Considerations for Designing Clinical Trials

Figure 7.1 illustrates data collected from a typical dose–response study, often referred to as a "randomized, double-blind, placebo-controlled, fixed dose, parallel group, and dose–response design". In such a study, patients are randomized into predetermined dose groups (often including a placebo, a low dose, one or several medium doses and a high dose). Patients take the randomized dose for the planned study duration. The efficacy and safety data obtained are analyzed to evaluate the dose–response relationships. Note that "parallel group", "fixed dose", and "placebo-controlled" are some important features of this design. Each asterisk in Figure 7.1 represents the efficacy measurement from one subject. Suppose a higher

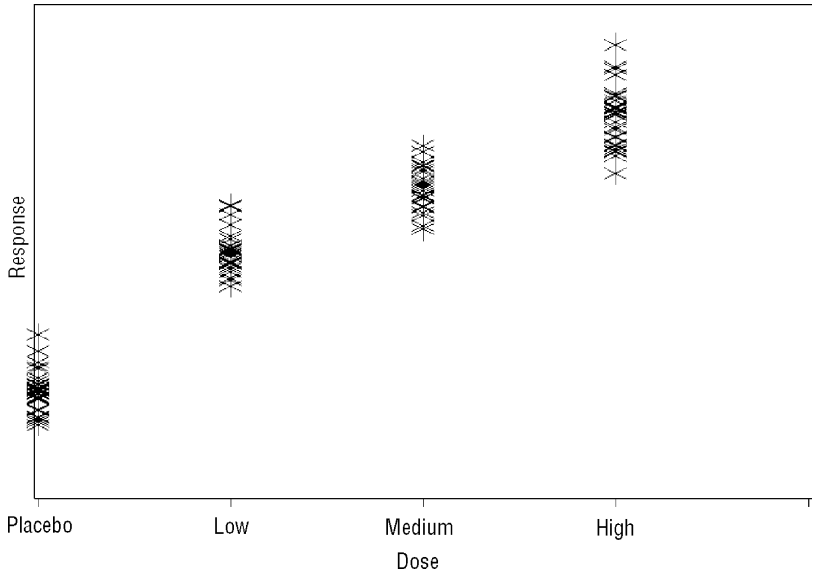

Figure 7.1. Observations from a simulated dose response study.

value indicates a better efficacy response, then Figure 7.1 indicates that as the dose of the test drug increases, the efficacy response improves.

Dose–response trials are typically conducted in Phase II, although occasionally they are done earlier or later. They are designed to explore a range of doses and to characterize the dose–response relationship. In this chapter, we will discuss some of the important clinical study design considerations for dose–response studies. Many of these points focus on Phase II (exploratory) study design, although some may be applicable to Phase III (confirmatory) studies, also.

7.2.1 Subject Population and Endpoints

Every clinical trial starts with a clinical question. Based on this question, clinical trial team members work together to draft a clinical trial protocol. This protocol serves as the design document for the trial. The results obtained from a clinical trial will help address the key clinical question. Hence, the most important study design consideration is to understand the objective(s) of the given study, and the trial is designed to collect the necessary clinical data to help answer these important clinical questions.

In designing a clinical trial, it is always important to collect and analyze data to address the primary objective. Typically, the primary objective can be studied by analyzing one or a few specific clinical variables (endpoints) from a well-defined study population. Hence, it is critical that in every study design, the subject population and the clinical endpoints be prespecified in the protocol. Primary and

secondary objectives should be aligned with the primary and secondary endpoints and populations.

As discussed in Chapter 1, clinical development is divided into four general phases (Phase I, II, III, and IV). In most of the Phase I studies, the purpose is to estimate pharmacokinetics (PK), pharmacodynamics (PD), and MTD. In Phase I, healthy and normal volunteers are recruited for trials to study drug candidates developed to treat non-life-threatening diseases. Endpoints used in Phase I include PK and PD parameters, as well as safety endpoints. Safety endpoints typically include adverse events, laboratory values and other measurements collected from examination equipment such as electrocardiogram (ECG). In all of the clinical studies, safety endpoints are collected, regardless at which phase the study is designed. This is because drug safety should be monitored closely in every stage of drug development.

In drugs developed for non-life-threatening diseases, a Phase II clinical trial is usually the first one to recruit patients with the disease under study. Patients for Phase II trials are recruited so that these patients may be most likely to benefit from the drug candidate and least likely to be exposed to potential toxicities. Endpoints used in Phase II studies include efficacy and safety endpoints. The efficacy endpoints may be clinical endpoints such as blood pressure, time to disease relapse, number of painful joints, visual acuity or surrogate markers such as white blood cell count, bone mineral density, among others.

Phase III studies are usually designed to recruit a wider patient population. This population could be very similar to the actual patients with the target disease. Clinical efficacy and safety endpoints are collected so that they are similar to the real world situation. Results obtained from Phase III studies are analyzed and reported to regulatory agencies for drug approval. In Phase III, we tend to have more relaxed inclusion/exclusion criteria with a hope to generalize well to clinical practice, but the heterogeneity of patient characteristics may reduce power of the trial.

The primary endpoint should be selected based on clinical relevance, directly related to study objectives. Other considerations may include the choice of scale (continuous, dichotomous, categorical), its potential impact on how analysis will be done, its impact on power, and its impact on interpretation.

In many situations, more than one efficacy endpoints are used to address the primary objective. When this is the case, it creates a multiple comparison issue in statistical analysis. Let the prespecified Type I error rate be α (usually a two-sided α is set at 0.05, or a one-sided α set at 0.025), then how should this α be spent for these multiple endpoints? What analysis should be performed so that the experiment-wise error rate (the Type I error is prespecified for the entire experiment) is controlled? All of these considerations will need to be addressed in the protocol and in the statistical analysis plan.

Often times, these multiple endpoints can be prioritized according to their importance in the clinical study. In this case, a stepwise test procedure can be applied to address the multiple endpoint issue by testing the most important endpoint first. If this null hypothesis is not rejected, stop. Otherwise, continue to test for the

second most important endpoint; then continue in this fashion until all prespecified endpoints are tested (please refer to Chapters 11 and 12 of this book). On the other hand, if two or three endpoints are equally important, then it is possible to combine these endpoints into a single score, and the primary analysis is performed on this composite score. It is also possible to apply multiple comparison adjustment to these equally important endpoints.

7.2.2 Parallel Designs versus Crossover Designs

In a fixed-dose parallel group design, a patient receives the same treatment for the duration of the trial. In contrast, in a crossover design, each patient receives a sequence of treatments during two or more study phases. Multiple sequence groups are used. Each has a different order of treatments, to account for any trends (such as disease progression or seasonal variation). For instance, in a 2×2 crossover design, a subject is randomized into one of two sequences. For one sequence, the subject takes treatment A in the first study period and treatment B in the second period, usually after washout period between treatments. The treatment order is reversed for the other sequence. Sometimes more complicated crossover designs are utilized.

In many cases, drug efficacy takes some time to demonstrate. A trial designed to study efficacy may need each patient to go through several weeks to several months of double-blind treatment. With this length of treatment, it is often difficult to use crossover designs. Hence, a parallel study design is used in many of the Phase II/III clinical trials.

7.2.3 Selection of Control

Three types of treatment controls can be considered in clinical trial designs: (1) historical control, (2) placebo control and (3) active control. Historical controls are based on data from other studies or the published literature, and they are usually less credible than placebo or active controls. Hence, historical controls are rarely used in clinical trials for new drug development. An active control is a treatment that is already on the market. Usually this is the standard treatment available for the disease under study. Active control group may be more useful in later phase studies. The advantages and disadvantages of using an active control group depend on the disease under study, characteristics of the drug candidate being developed, and specific clinical inferences of interest (ICH E10, 2001). Studies with an active control, but without a placebo group, suffer from the additional burden of demonstrating that the treatment groups are effective (assay sensitivity), either through superiority to the active control or on the basis of some type of historical control information (Temple and Ellenberg, 2000; Ellenberg and Temple, 2000). An active control, however, may provide a reference from a treatment of 'known' effectiveness. In practice, the use or not use of an active control mainly depends on objectives of the study. However, in certain cases, it may also depend on clinical budget considerations.

In the early stage of clinical development of a new drug, it is a common practice (if deemed ethical) to compare the test drug with a placebo. This is important since detecting positive signals of effectiveness beyond that achieved with placebo is an important milestone for continuing development of this drug candidate. Accordingly, placebo plays an important role in a dose–response study—it represents a zero dose in the study. Patient response at zero dose is a basic standard for comparison with active doses. In typical dose–response studies, a few fixed doses (usually two, three or four) would be chosen. These doses plus placebo constitute the treatment groups for a randomized dose–response trial.

The basic principle is that the design needs to cover a range of doses, as wide as possible in most cases. Generally, the low end will be placebo (at dose 0), but sometimes the lowest dose may exceed zero (e.g., for ethical concerns). This raises at least two issues: (1) a narrower dose range reduces the power to detect a relationship, all other things being equal; (2) even if there is a significant dose–response slope in the right direction, we need to be able to argue that this slope reflects an improvement in all groups (rather than the case where a higher dose may be worse than placebo).

An active control group can be useful, for example, if the test drug did not show a difference from placebo, but the active control group demonstrates a superiority response compared with placebo. This provides evidence that the study drug did not work. However, if the active control does not show a difference from placebo, then one of two possibilities can be contemplated: either the placebo response is too high, or the conduct of the study was flawed so that nothing can be differentiated.

7.2.4 Multiple Comparisons

In typical dose–response studies, more than two treatment groups are included in a clinical trial. When this is the case, it is important to understand the questions related to the objectives of the study:

- To show a trend such that higher doses tend to have better responses? or
- To show a particular dose is better than placebo?

Depending on the objective of a study, appropriate multiple comparison adjustment need to be made so that the probability of making a Type I error can be controlled under α. In most Phase II studies, the objective is to estimate a trend of dose–response relationship. A modeling approach is commonly applicable for this purpose. Commonly used dose–response models include linear, quadratic, $E_{\max}$, logistic or others. Chapters 9 and 10 of this book discuss the modeling approach in analyzing dose–response data. In certain situations, a preplanned dose–response test with a positive slope can be considered as one of the pivotal proof of efficacy trials.

For Phase III, it is critical that during the study design stage, a multiple comparison procedure be prespecified. This is similar to the multiple endpoint issue: e.g., How is the Type I error α (or one-sided $\alpha/2$) controlled when more than one comparison is made? A number of multiple comparison procedures are available.

Commonly used procedures include Dunnett, Bonferroni, Hochberg, stepwise, and others (Hsu, 1996). Choice of procedures to be used for a particular study depends on the objective(s) of the study, the background disease for treatment, and how much prior knowledge is available at the time when the study is designed. Multiple comparison procedure is one of the most important statistical concerns in design and analysis of dose–response studies. This will be discussed in Chapters 10–13.

7.2.5 Sample Size Considerations

During the design stage of any clinical trials, one important question is always How many subjects will be needed for this study? Generally, for a continuous variable, four important quantities are used to estimate the sample size: namely α, β, δ, and σ. Here α and β represents the probability of making a Type I error and a Type II error, respectively. These two quantities are prespecified probabilities to control for false positive and false negative rate, respectively. The quantity δ is the clinically important difference we want to detect from this study (often this is postulated as minimally clinically important difference), and σ is the common standard deviation for each treatment group (usually obtained from previous studies). Depending on the type of data to be used for analysis (continuous data, categorical data, time-to-event data), various formulas are available to calculate sample sizes using these four quantities.

When designing a dose–response study, the main concern in sample size calculation is how to handle multiple group comparisons. This issue is dictated by the objective of the study. A few examples of the objective of a dose–response study may be as follows:

- Testing to see if a specific dose of the study drug is different from placebo
- Finding the minimum effective dose
- Differentiating efficacy between active doses
- Checking to see if there is an increasing dose trend
- Demonstrating noninferiority between a particular dose and the active control

As a good clinical practice, it is important to keep a single and clear primary objective for a single study. Hence, the above examples can be considered mutually exclusive. The appropriate statistical method used to perform data analysis should be aligned with the primary objective. Sample size estimation, in turn, should be consistent with the data analysis method.

In the first example, if the trial were designed to differentiate a specific dose of test drug from placebo, then the sample sizing method would be similar to performing a two-sample t-test comparing the test drug against placebo. In most situations when analyzing dose–response studies, multiple comparison adjustments will be needed. Depending on the type of multiple comparison to be used for data analysis, sample sizes should be estimated based on the chosen method. For example, if a pre-determined stepwise multiple comparison adjustment will be used for analysis, then the two-sample t test at level α could be appropriate. On the other hand, if the Bonfferoni adjustment is proposed for data analysis, then the appropriate α adjustment will have to be made prior to

sample size estimation. In Phase II dose–response studies, the main purpose is to estimate a monotonic relationship and hence the sample size and power will help demonstrating a significant slope in a regression model.

There is another angle of sample size estimation: a study is powered to achieve a required amount of precision for an estimated quantity using a confidence interval approach (rather than testing Ho: effect = 0). The quantity could be an accepted range of responses at a given dose—or, more usefully, the dose to give a required range of response. This angle is not covered in this book.

A general discussion regarding sample size determination and power can be found in the (*Encyclopedia of Biopharmaceutical Statistics*, 2003). These considerations specifically for dose–response clinical trials will be covered in Chapter 14 here in this book.

7.2.6 Multiple Center Studies

Clinical trials are commonly conducted at a number of different investigator sites or centers. The main reason for this practice is to ensure timely enrollment of sufficient number of patients. Another benefit of multiple center studies is that results obtained from these studies can represent a wider variety of patient background. This means that a multiple center study including various type of centers are more desirable, and the conclusion is not heavily dependent on one single center. In other words, the conclusion of multiple center studies is more "generalizable", so that the interpretation of these results is more likely to be applied to a broader patient population.

Different centers may have different recruitment rates. As a result, this can cause an imbalance in the number of patients recruited from various centers. If this happens, some centers may fail to provide enough patients to be randomized to each treatment group, and the treatment-by-center interaction may become non-estimable. Therefore, we tend to limit the number of treatment groups in order to minimize the imbalance problem.

For example, in a dose–response study, there is often a need to include many doses in one study. As demonstrated in Figure 7.1, a typical dose–response study would include a placebo and three test doses (a total of four treatment groups). When this is the case, in a multiple center study, it is desirable to include at least four patients (one in each treatment group) from each center. However, in some cases, the center may fail to recruit up to four patients and this will cause imbalance in data analyses. There can also be situations where one particular dose is over (or under) represented in many centers. Then, when all centers are pooled together, the data causes another type of imbalance.

7.3 Design Considerations for Phase II Dose–Response Studies

Dose–response studies are usually carried out in Phase II. At this point, there is often a considerable amount of uncertainty regarding any hypothetical dose–

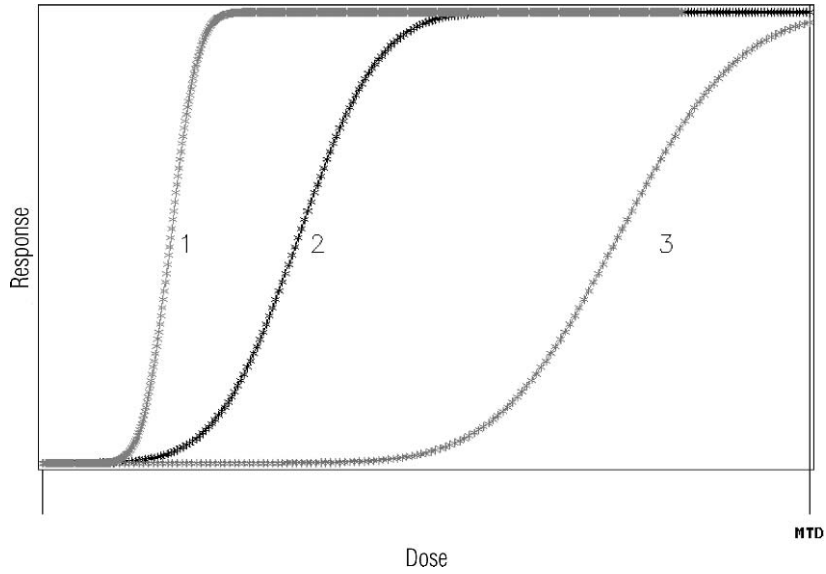

Figure 7.2. Several possible dose–response curves.

response curves. It is typical at this time that the MTD is known from Phase I studies, and it may also be assumed that efficacy is nondecreasing with increasing doses. Even so, the underlying dose–response curve can still take many possible shapes. Under each assumed curve, there are various strategies of allocating doses. For example, in Figure 7.2, the population dose–response curve can be assumed to take a variety of shapes. If we select doses to detect the ascending part of curve 3, then the planned doses should be on the higher range. On the other hand, if we need to select doses to detect the activities of curve 1, then the doses should be chosen on the lower end. Thus, the dose allocation strategy can be very different depending on the underlying assumed dose–response curves.

In general, when designing a Phase II dose–response clinical trial, we need to consider the following important points: dose frequency, dose range, number of doses, dose spacing, use of control (or lack of), sample size for each treatment, fixed dose or dose titration, and others. Some of these points are discussed in the subsections below.

7.3.1 Frequency of Dosing

In designing dose–response clinical studies, we need to know how often should a patient take the test drug (e.g., once a day, twice a day, or dose every 4 hours during the day). This is a question about dosing frequency, and it is usually guided by the Phase I PK–PD findings. One of the PK parameters is the half-life of a drug. The estimated half-life helps to estimate how long the drug will stay in human body. Using this information, we can propose a dosing frequency to be used for dose–response study design. In certain cases, we may study more than

one dosing frequency in a single study. When this is the case, a factorial design (dose, frequency, dose × frequency) can be considered.

However, in some drugs, the PD response may be different from PK. Recall that PD measures how the drug works in human body while PK measures how the body do to the drug. In this case, even from the PK half-life data, we think there is insufficient drug in the body after several hours of dosing, but there may still be enough drug in the tissue to help with PD responses. On the other hand, the PK may indicate that there are still plenty of drug in the body, but these drugs may not cause any effective PD responses. In some drugs, the concentrations for PD activities can be very different from that for PK activities. Hence, the dose frequency derived from PK may either overestimate or underestimate the concentration needed for PD response. In some cases, the best dose frequency may be derived in later phases of the drug development.

Another important guiding principle in selecting dosing frequency is based on the market assessment. For example, if the market requires a once daily dosing treatment, but the drug candidate under development has a twice-a-day PK profile, then some formulation change may be necessary. Figure 7.3 presents time–concentration curves of this situation. The horizontal line that is above the x-axis represents the efficacy concentration level (often based on PD information). Theoretically speaking, we need to keep the drug concentration staying above this line all the time for the drug to work. In order to achieve this concentration, two strategies are possible: we can either dose the subject twice-a-day (BID) with low dose (Figure 7.3), or once-a-day (QD) with the high dose (which is twice the

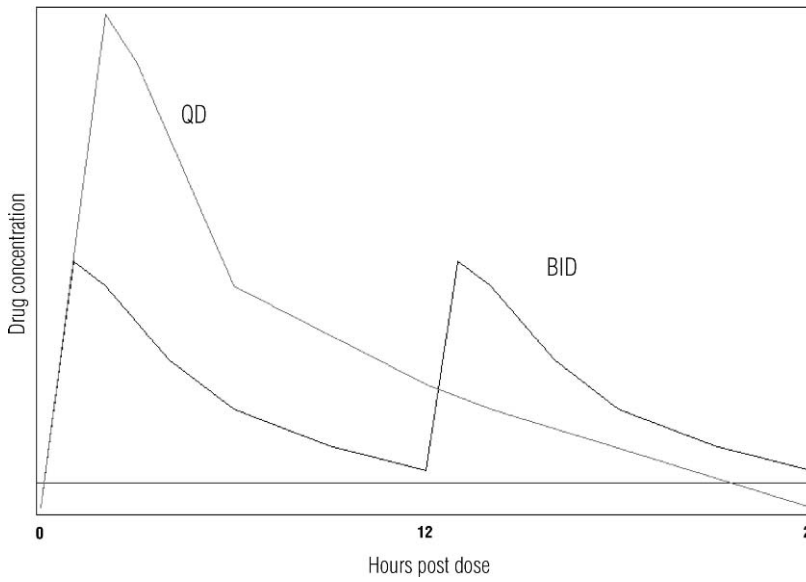

Figure 7.3. Once a day vs twice a day dosing.

dosage of the low dose, Figure 7.3). Note that in the first few hours post dosing, the high dose may result in a very high concentration, which could potentially cause severe adverse events. When this is the case, re-formulation of the drug may be needed so that when dosed as once a day, the C_{max} would not be too high, while the efficacy concentration can be maintained throughout a 24-hour period.

In designing the first few Phase I trials to study PK, there is very limited information about how will the human body metabolize this drug candidate. At this stage, data observed from preclinical studies and animal experiments on this drug candidate are used to help guide designing these Phase I trials. In case the drug candidate belongs to a certain drug class where other drugs of the same class were already on the market, information obtained from these other drugs can be used to help guiding the study designs for this drug candidate.

During the development of a drug candidate, sometimes reformulation may be needed. There can be many different reasons why there is a need for drug reformulation, including to help absorption, and to change the half-life. It is critical to understand that after reformulation, the PK–PD properties of the drug candidate are different from what they were prior to reformulation. Hence, all of the dosing information and drug regimen obtained from studies before the reformulation will need to be changed and re-studied. This can potentially cause major re-work. Rework in drug research and development delays the development process and results in additional amount of investment.

In studying the PK–PD relationship, we should realize that the main point is whether C_{min}, C_{max}, or AUC drives the PD. This is a fertile area for collaboration between statisticians and pharmacokineticists. Models based on prior trial data (e.g., from preclinical data, clinical data of the drug candidate under study, other compound of the same class) can be developed to inform the decision.

7.3.2 Fixed-Dose versus Dose-Titration Designs

A fixed-dose design is in contrast to a dose-titration design. In a fixed-dose design, once a patient is randomized to a dose group, the patient would take the same dose of study drug throughout the entire dosing period. In a dose titration design, a patient is randomized to a dose regimen with a starting dose, then the dose for a patient can be changed over time. In a dose-titration study, subjects are randomized to start with a low dose, and depending on either patient's response to the drug, or a predetermined schedule, the dose is gradually increased until a suitable dose level is found. For example, in a "titration to response" design, each subject can receive more than one dose. A patient who responds to a low dose may stay on this low dose, and a patient who does not respond to a low dose after a prespecified treatment period may receive the next highest dose of the drug. This procedure is repeated until some designed criteria are satisfied. There are at least two ways of analyzing data obtained from this design:

1. If patients are titrated until a response occurs (e.g., sufficient efficacy or a tolerable level of adverse events), the response measure is the dose achieved

and the study generates an estimate of the distribution of doses required for a response. This can be useful—but it's not a dose–response relationship per se.
2. This same data (assuming multiple measurements of some parameter prior to response) can be analyzed with a mixed effects model in an attempt to tease out the dose–response relationship. This is based on the assumption that individuals vary in their dose–response parameters and we observe a censored set of data.

In this design, the time effect versus dose effects complicate matters.

In addition to the titration-to-response design, there are other types of titration designs. One example is a fixed titration design: doses are changed on a fixed schedule without regard to response. The time on dose is confounded with dose, so that either a model-based analysis is needed or preferably the dose groups are split at the titration times, keeping some patients on the same dose to allow estimation of the time effect. Another example is dose–response cross-over design—this can be considered as a type of titration design—with the sequence groups taking care of the time effect mentioned for the fixed titration design.

There are some advantages of a titration design. For example, a study with this design will allow a patient to be treated at the optimum dose for the patient; this dose allocation feature reflects the actual medical practice. However, the disadvantage of a titration design is the difficulty in data analysis. For example, if a patient responded to the test drug after doses are escalated, it is unclear whether the higher dose or the accumulation of the lower dose caused the response. In titration designs with multiple treatment groups, there may be overlapping doses—e.g., one treatment group is 10 mg escalating to 20 mg, while another group is 20 mg escalating to 40 mg. When this is the case, it is difficult to make inferences about the 20 mg dose group.

In some rare cases, instead of a dose–response study, a concentration response study is designed. A concentration response study assesses efficacy and safety measurements observed from subjects according to the plasma concentration of the study drug, but not the doses of the study drug. There are many practical limitations in using this type of designs, these include, among others, how to blind the patient and the physician, and how and when to measure the blood concentration.

Because of the issues with dose titration designs and concentration response designs, the parallel, fixed dose designs are, in general, the more commonly used designs for dose–response studies. Therefore, in many of the dose–response studies, patients are randomized to a few fixed dose groups and are compared with one or more control treatment groups.

7.3.3 Range of Doses to be Studied

As discussed earlier, drug efficacy can only be studied from patients with the target disease. Hence, at the beginning of Phase II there is no efficacy information on these patients to help define the dose range for study. It is desirable to obtain information that helps describing the efficacy and safety dose–response curves. Studies should

be designed to help estimate MaxED, MinED, and possibly to obtain additional information to support MTD. Although estimates regarding MTD should have been available prior to Phase II, more information will be helpful to re-confirm or to adjust MTD estimates obtained from previous trials. If the budget and timeline are permissible, the first dose ranging study should cover a wide dose range in a hope that this study will help identify the doses where most of the activities exist. The next study will then be designed to capture the dose–response relationship using information obtained from the first study.

Note that nonclinical information on the candidate and perhaps both clinical and nonclinical data for related compounds often provide a minimum drug concentration profile that is expected to be required for efficacy and safety. Together with the PK profile, this provides a target dose range, which we would want to explore, and possibly, a minimum dose expected to have little or no efficacy that we might want to include.

In a dose–response design with placebo, low dose, high dose and several doses in between, the dose range is defined as the range between the lowest and the highest dose. Dose range can be expressed as the ratio of highest dose over lowest dose—as a rule of thumb, in the first dose ranging study, the range should be at least 10-fold. In many cases, when the dose range is too narrow, the dose–response study failed to deliver the necessary information for efficacy or safety, and re-work will be needed after these studies. Costs of re-work can be tremendous at times. These costs may include costs of additional studies and costs of delaying the drug get to the market, in addition to all the resources foregone in conducting the current studies.

7.3.4 Number of Doses to be Tested

In order to cover a wide range of doses, it is desirable to study as many doses as possible. However, the number of doses that can be tested in a given study is limited. There are practical constraints in determining the number of treatment groups. Most trials with sufficient number of patients are multicenter trials. As mentioned earlier, different centers may have different recruitment rates, and imbalance in number of patients between treatment groups may exist. If this happens, the treatment-by-center interaction may become nonestimable. By increasing the number of treatment groups, the risk of imbalanceness increases. In order to minimize this risk, we tend to limit the number of treatment groups in each study. If we need to have more dose groups in a Phase II setting, we can prespecify that the primary model for data analysis is a main effect model, and that the treatment-by-center interaction is not to be tested or estimated.

Another practical issue is dosage form. Sometimes there are only limited dosage forms available in the early stage of clinical development. When this is the case, the number of doses to be used in a study may also be restricted. For example, if the tablet strengths are 10, 20, and 50 mg, respectively, then it is very difficult to study doses of 1, 3, or 25 mg, respectively. For some studies, the technique

to achieve blinding is to produce matching placebos for each treatment group. The more doses that are included in a study, the more matching placebos may be needed. For example, if 1 mg tablets, 5 mg tablets and 10 mg tablets are used in the same study and these three types of tablets look different, then a placebo for each dose will be needed; i.e., three types of matching placebo to be used in this study. The number of dose groups may also be limited by practical considerations of how many pills we might reasonably expect a subject to take for any given dose. For these reasons and others, clinical studies designed with more than six or seven treatment groups are rare.

7.3.5 Dose Allocation, Dose Spacing

As depicted in Figure 7.2, with very limited information about the drug candidate, after allocating a placebo control, a high dose that is close to MTD, it will be very difficult to select the medium or low doses. The challenge is that at an early stage, there is no information as to what the underlying dose–response curve should be. Is it curve 1, curve 2, curve 3 or some other form? When there is very limited data to help allocating doses, we may consider the potential use of other information such as preclinical and related compounds. This is much more than an issue for statisticians, we should preferably work with the pharmacokineticists, clinicians, and pharmacologists. Dose allocation also depends upon the primary question: detecting an effect, estimating the slope near the MTD, finding the lowest dose with effect of at least some minimally clinically important difference, fitting a specific type of model, and so forth.

After the number of dose groups is chosen, it is still a challenge to determine the high dose, low dose, and spacing between test doses. Typically, the high dose is a dose selected around or below the MTD, but choices of lower doses are often challenging. Wong and Lachenbruch (1996) introduce cases using equal dose spacing from low to high doses; that is to divide the distance from placebo to highest dose by the number of active doses, then use that divided distance as the space between two consecutive doses (e.g., 20, 40, 60 mg, respectively). Others may consider some type of log dose spacing; e.g., 1, 3, 10, and 30 mg, respectively, for the design.

Hamlett et al. (2002) proposed to use binary dose spacing (BDS) design for dose allocation. If the study includes two test doses and placebo, BDS suggests to pick a mid-point between placebo and MTD, then allocate a dose above the midpoint and another dose below. If the study uses three test doses and placebo, BDS suggests to keep the high dose as the one selected in the two-dose case. Then pick a second midpoint between placebo and the first mid-point, allocate the low dose below the second midpoint, and the medium dose between the two mid-points. When more doses are used, BDS picks more mid-points to the lower end and allocates doses accordingly. BDS provides a wide dose range, helps identify MinED, avoids allocating doses too close to the MTD, allows a log-like dose spacing, it is flexible, and easy to implement.

7.3.6 Optimal Designs

The dose levels and the number of subjects at each level can be chosen mathematically (using mathematical theory, simulation, or some other tools) in order to optimize a statistical criterion such as small errors of estimation. This set of dose levels and number of subjects at each level is called the statistically optimal experimental design, and the design depends on the chosen criterion and the underlying model for the dose–response curve.

The statistical principles of optimal experimental design can be applied to many studies. Optimal design techniques can be used with various statistical models. For a given study objective and a reasonable model, optimization techniques allow one to determine the statistically best set of doses and number of subjects to be used at each dose. These designs help to estimate the parameters of the model; for example, slope and ED50 (the dose which achieves 50% of efficacy response) in logistic regression models, and intercept and slope in linear regression models. The doses used might not necessarily be those intended for use in the label, but they provide a basis for estimation of the dose–response curve so that the response at any dose can be predicted with validity and precision. Depending on the optimality criteria chosen, the doses studied may not necessarily be equally spaced or have equal numbers of subjects at each dose. Information on the shape of the dose–response curve should be attained where possible from PK–PD studies and early Phase II studies, which can help the design of the late Phase II studies.

Pukelsheim (1993) proposes a comprehensive set of statistical approaches to optimal experimental design. Wong and Lachenbruch (1996) review dose–response designs and use optimal design criteria for linear and quadratic regression. They also use simulation to illustrate the effect of optimal design criteria on spacing of doses and the numbers of subjects at each dose.

The key to optimizing a design is availability and use of prior information—based on candidate information and related compounds. A second key is to take into account the uncertainty in an a priori model. A goal might be to obtain a design that will work adequately no matter where the dose–response curve sits on the dose scale (over the range deemed most likely). On the other hand, the focus may be on average success—weighted average success, integrated over the prior distribution of the dose location uncertainty. Clinical simulation can be a useful tool here.

7.4 Concluding Remarks

Dose finding or dose selection happens mostly during Phase II or Phase III clinical development. The primary challenge for designing a Phase II dose–response clinical study is the lack of knowledge about how the drug works because this is the first time the test drug is studied in patients with the target disease. Again, Phase II studies are designed primarily for exploratory purposes and hence the main statistical method is estimation, and scientists tend to use model approaches in analyzing data collected from these studies. The main challenge in Phase III is to guess the correct

dose or range of doses and be able to demonstrate it. Phase III studies are for confirmatory purposes. Multiple comparison procedures are commonly used to test statistical hypotheses for each dose comparing with placebo. This chapter introduced some considerations and difficulties in designing dose–response clinical trials.

A changing environment is pushing scientists, especially statisticians, to be more creative in designing dose-finding studies. Recently, FDA discusses the Critical Path initiative, which pressure the sponsors to reduce the development cost and speed up the time line. Within many pharmaceutical companies, there is a strong push to do more creative Phase II programs, aimed at assessing the dose response for safety and efficacy so that the drug candidate enter Phase III with the right dose range (or to stop developing compounds in Phase II if they're not likely to measure up). The potentially more creative designs and analyses serve a role in the regulatory review—for justifying the dose selection and possibly even for Phase III pivotal trials. Hence we hope to encourage readers to think about some of the newer strategies based on considering a wider range of potential designs, using prior information to inform the design, and basing at least some of the interpretation on model-based analyses that can take advantage of prior information and pharmacologically-reasonable assumptions about the underlying dose–response relationship.

At the end of Phase III, in the preparation of an NDA, the sponsor drafts summary of clinical efficacy (SCE) and summary of clinical safety (SCS). Traditionally, this often includes simple pooling of similar studies and side-by-side presentation of results. There is much more that can be done. One objective is to perform dose–response oriented meta-analyses of individual patient data, to combine all the relevant information about dose response and in particular, how it depends upon the indication, concomitant disease conditions, patient demographics, as well as time factors to accommodate different trial lengths. These were useful in the FDA discussions. This sort of meta-analysis can be built prospectively into the clinical development plan. In this regard, we hope to promote the collaboration of statisticians with PK–PD scientists, as natural 'partners in quantification'. This means broadening the perspective and understanding the difference between 'learning' and 'confirming' objectives for design and analysis of trials (and the entire programs).

Clinical trial simulation is a very useful tool to help with dose–response study designs. Simulation can be used to examine the impact of dose spacing, number of groups, and method for data analysis. There's not one right answer about number of groups as it depends upon the specific trial objectives, the data characteristics, and the dose–response relationship itself. An example of a clinical trial simulation is provided in Chapter 8.

References

Ellenberg, S.S., and Temple, R. 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments (Part 2). *Annals of Internal Medicine* 133(6):464–470.

Chow, S. 2003. *Encyclopedia of Biopharmaceutical Statistics*. 2nd ed. New York: Marcel Dekker.

Hamlett, A., Ting, N., Hanumara, C., and Finman, J.S. 2002. Dose spacing in early dose response clinical trial designs. *Drug Information Journal* 36(4):855–864.

Hsu, J.C. 1996. *Multiple comparisons—Theory and methods*. New York: Chapman-Hall.

ICH E10. 2001. Choice of control group and related issues in clinical trials. *Harmonized Tripartite Guideline*. http://www.fda.gov/cder/guidance/index.htm

Pukelsheim, F. 1993. *Optimal Design of Experiments*. New York: Wiley.

Temple, R., and Ellenberg, S.S. 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments (Part 1). *Annals of Internal Medicine* 133(6):455–463.

Wong, W.K., and Lachenbruch, P.A. 1996. Tutorial in biostatistics: Designing studies for dose response. *Statistics in Medicine* 15:343–359.