

# Pharmacological modeling and biostatistical analysis of a new drug

Revathi Ananthkrishnan<sup>1</sup>  
Philimon Gona<sup>2</sup>

<sup>1</sup>Cambridge, MA, USA; <sup>2</sup>Boston University, Mathematics and Statistics Department, 111 Cummington St, Boston, MA-02215, USA

**Abstract:** Clinical research and clinical trials of experimental drugs to treat human diseases have gained greater importance in recent years. Phase I–IV clinical trials offer patients the opportunity to gain access to a new, more efficacious and safer medication to alleviate or cure their disease. There are potential side effects of every new drug; however, such trials and studies are crucial for drug development and testing in humans. The US Food and Drug Administration (FDA) regulated process of evaluating a new drug for treating a particular disease in humans is long, rigorous, and includes the stages starting from preclinical research through the entire human clinical trials process. This review synthesizes results from the above stages and describes the entire mechanism of the clinical study of a new drug for human disease. It emphasizes the associated mathematical modeling and statistical analyses, and bridges pharmacological modeling and biostatistics in clinical research and also provides a basic theoretical overview to biomedical experimentalists. The modern trend in clinical research involves a unified approach among several biomedical subspecialties and it is hoped that even more integrated studies of new drugs will continue to be carried out, leading to novel drugs that are highly effective in curing the associated condition.

**Keywords:** PK/PD pharmacological modeling, biostatistical analyses of clinical trials data, clinical trials, phases of clinical trials, types and designs of clinical trials

## Introduction

The clinical study of a new drug for a particular disease in humans is a scientifically rigorous process used to quantitatively evaluate the safety and efficacy of the drug. Firstly, several *in vitro* and animal tests are conducted with the drug to understand its clinical properties. These studies are known as preclinical studies. Preclinical studies are an essential starting point and provide preliminary evidence of how the drug may work in humans. However, these *in vitro* and animal studies alone are inadequate, and the drug needs to undergo further testing in humans. Only such comprehensive and extensive testing can clearly demonstrate the safety and efficacy of the drug. This combined clinical testing of the drug in humans starting from phase I through phase IV comprises a clinical trial. Considered the gold-standard for biomedical research, a randomized clinical trial (RCT) is an interventional and prospective study.<sup>1</sup> In its simplest form, the RCT consists of two randomly assigned treatment arms. One arm receives the active drug while the other receives a placebo. Both arms are monitored for their response to the drug and are statistically compared for both safety and efficacy.

This review integrates and presents analyses of preclinical research and clinical trials, focusing on the associated mathematical modeling and statistical analyses to

Correspondence: Revathi Ananthkrishnan  
3 Everett St, #3, Cambridge,  
MA 02138, USA  
Tel +1 908 2409647  
Email revathia@gmail.com

provide a relatively comprehensive understanding of the complete drug-testing process. It endeavors to bridge the gap between two fields by providing the biostatistician who is not familiar with preclinical pharmacological modeling and the preclinical modeler who is not familiar with biostatistical analysis of clinical trials data with an overview of the other area. Finally, it provides the biomedical experimentalist with a basic theoretical background of current modeling and analysis techniques used to interpret the experimental data from drug testing.

## Four phases of new drug development

The drug-testing process in humans usually takes place in three or four phases<sup>2</sup> (Table 1). A phase I study is performed following the completion of preclinical studies (described in detail below). The new drug is clinically tested on a small group of (usually) healthy humans to determine dosage ranges, tolerability and safety of the drug. A subsequent phase II study involves administering the drug to a larger number of subjects with the disease. The purpose of a phase II study is to determine the drug's safety and efficacy in

treating the disease. Following a phase II study is a phase III study, also known as a pivotal study, which involves a large number of subjects and is designed to thoroughly evaluate the drug's safety and efficacy. Experimental drugs that perform satisfactorily in phase III studies can be approved for use by regulatory agencies. Since phases I–III are conducted in controlled pharmacovigilant clinical settings using carefully selected subjects satisfying strict eligibility criteria, the long-term safety profile of the drug may not be known until a large number of patients from the general population has been treated with the new drug. Thus, long-term surveillance or phase IV postmarketing studies are sometimes conducted after the drug is put on the market.

A clinical study and associated clinical trials of a new drug for a human disease can be sponsored and conducted by either a federal entity such as the National Institutes of Health (NIH) in the United States, or by a biomedical research institute, a university, a hospital, or by a pharmaceutical company. The trial is conducted in a variety of settings such as university medical facilities, public or private hospitals or a combination of these.

**Table 1** The principal characteristics of the different phases of clinical trials

	Phase I	Phase II	Phase III	Phase IV
<b>Enrollment</b>	Small, >10.	Medium, >50.	Large, >1000.	Large, >1000.
<b>Goal</b>	Is the drug safe in humans?	Is the drug safe and efficacious in humans?	Is the drug more effective/safer than the standard treatment?	How does the drug work in the general population?
	Drug safety (toxicity, carcinogenicity) is investigated. PK/PD properties, MTD, dosing schedule are also investigated.	Trials can be exploratory to find the optimal dose and dosing schedule, using efficacy dose response curves and safety measures.	Trials are usually confirmatory to test or confirm efficacy/safety observed in phase II.	These consider long-term side effects, safety, efficacy, benefits of the drug outside the restrictive sample of the clinical trials.
<b>Participants</b>	Usually healthy subjects but in rare diseases or severe diseases, diseased subjects are studied.	Diseased subjects.	Diseased subjects.	Diseased subjects, and can include participant groups not studied when the drug was approved.
<b>Blinding and Randomization</b>	Not blinded, not randomized.	May or may not be randomized. May or may not be blinded.	Most are randomized and blinded.	Usually observational, surveillance studies but can be randomized studies.
<b>Sample size requirements</b>	Sample size may or may not be calculated. <sup>§</sup>	Sample size does not have to be based on rigorous statistical calculations.	Rigorous statistical calculations of sample size, based on responses or endpoints and efficacy measures, are required.	Usually not estimated for an observational study but may be calculated for a randomized study.
<b>Follow-up</b>	No follow-up; subjects enroll in subsequent phase II trial if subject is eligible and willing.	May or may not have long term follow up.	Usually have long term follow up.	Usually long term studies.

**Notes:** <sup>§</sup>Sample size is not formally calculated in trials using the 3 + 3 design of dose escalation where subjects are enrolled until the MTD is determined. However, some trialists perform simulations to estimate the number of evaluable subjects required to determine the MTD.<sup>34</sup>

**Abbreviations:** MTD, maximum tolerated dose; PD, pharmacodynamic; PK, pharmacokinetic.

## The clinical trial protocol

All clinical trials strictly follow a specific protocol or a prespecified and agreed upon set of clinical research rules. The protocol is a road-map prepared by the drug's sponsor to present the clinical trial objectives, scientific details of the new drug compound, its pharmacokinetic (PK) and pharmacodynamic (PD) properties and dosage, the eligibility criteria for subject participation. Calculations for the required sample size and statistical power, randomization procedures, the study design, the schedule of events, the specific outcomes to be measured and investigated, and some details regarding the statistical analyses of the data are all thoroughly described in the clinical trial protocol. Details of the clinical trial design include whether the study will be open-label, where everyone involved in the study including the subject is aware of the treatment assigned, or singly- or doubly-blinded, where either the subject or investigator or both do not know the treatment assigned to the subject. Blinding is a procedure used to limit observer bias<sup>3-5</sup> ie, theoretically, the observer will treat all study arms equally. Randomization, in theory, ensures the balance and comparability of subjects that are assigned to different treatment arms with respect to known and unknown confounding variables. The ultimate goal is to be able to attribute any observed efficacy or safety differences to the treatment rather than to some undefined external factors or lurking variables. The statistical team generates the randomization schedule, a list of preselected study identification numbers with each number randomly assigned to a treatment arm such as placebo or an active treatment group in the case of two treatment arms. The randomization schedule is then made available to participating sites by the sites calling into an automated interactive voice response system (IVRS) in order to assign a participant to a treatment arm.

The clinical trial protocol and any amendments need to be approved by regulatory agencies and Institutional Review Boards (IRBs) before any subject is enrolled. The role of the IRB is to protect human subjects from harm and the board is composed of individuals who do not stand to gain financially or otherwise from the new drug. The Food and Drug Administration (FDA) in the United States, or the appropriate agency elsewhere, decides whether to permit conducting the study based on a review of the sponsor's submitted Investigational New Drug Application (IND) that contains manufacturing details, preclinical results such as the pharmacology and toxicology data, and the protocol for human clinical trials. The FDA has 30 days to respond to the

IND and may request additional information or clarifications, or request the sponsor to address any concerns; the sponsor must address these issues adequately before the trial can start. If the sponsor does not receive a response from the FDA within 30 days of receipt of the IND, the sponsor can start the human clinical study (for differences in the drug approval process and drug dosing between FDA, European Medicines Agency [EMA; Europe] and Product Development and Management Association [PDMA; Japan] see<sup>6-8</sup>).

## The investigator brochure

The Investigator Brochure (IB) provides information about the experimental drug such as the drug's composition, chemical properties, manufacturing information, PK/PD and toxic properties observed in preclinical studies, safety and efficacy in humans, possible adverse reactions and also precautions to be taken with regard to its use. The IB is updated regularly as the study progresses and is used along with the protocol to provide additional information on the experimental drug.

## Statistical analysis plan

The relevant data from each subject are captured on paper or electronic case report forms (CRF or eCRF) that are specifically designed for the clinical trial. The clinical trial protocol, along with the CRFs or eCRFs, is the main source that statisticians use to design and write the statistical analysis plan (SAP). Regulatory agencies normally insist on an SAP that is written and approved prior to any data collection or analysis. Prespecified analyses are usually required to avoid the potential of data driven hypotheses testing should the investigating team see any data before the SAP is finalized. A phase II or III SAP typically includes a detailed explanation of the safety and efficacy evaluation to be performed, and shells of listings, tables and figures that are to be created to summarize trial results. The SAP may include shells of analysis tables with dummy place-holders 'xxx's used to denote the values of the statistics to be calculated using the clinical trial data. The SAP and the statistical analysis procedures used must follow the ICH harmonized tripartite guideline E9 on statistical principles for clinical trials.<sup>9</sup>

## Informed consent, confidentiality, and protection of human subjects

An important rule for every trial is that each subject entering the study must give written "informed consent" to participate in the study, being fully aware of its potential benefits and risks.<sup>10</sup> Subjects also enter the clinical trial with the full

understanding that they are expected to follow the clinical trial protocol and that they can choose to discontinue the trial with no consequences on the care they will receive. Subjects participate on their own will and are protected from any form of coercion regarding study participation or withdrawal of their consent. Detailed and specific conditions under which a subject may be discontinued from the study are provided. The personal details of every subject are kept strictly confidential by trial personnel. Each subject is identified by a unique study identification number given in the randomization schedule and is not identified by name. Every clinical trial in the United States is monitored by an independent committee of several experts (such as the IRB), which initially approves conducting the trial and later regularly oversees and reviews the research being carried out. This mandatory oversight helps ensure that Good Clinical Practice (GCP) guidelines are followed and helps avoid any unethical or unacceptable practices in a clinical trial. Another group that is set up in some trials to further protect subjects is the Data Monitoring Committee (DMC). The DMC reviews the trial data to examine the validity of the data, any safety issues or adverse events of the drug and whether the trial should be continued or stopped. In addition, the FDA can inspect any site of the clinical trial and can stop the site/investigator (or the entire trial itself) from participating if any problems are found.

## Stages of clinical studies of a drug

### Preclinical studies

Prior to conducting a human clinical trial of a new drug, pre-clinical studies are first carried out to learn about the drug's mechanism of action and the main regions of the body the drug acts on, its absorption and elimination properties, therapeutic and toxic dosage levels and the drug concentration versus drug effect properties. Preclinical studies usually involve *in vitro* and animal studies together with computational modeling. Animal models of disease refer to testing the experimental drug in animals with the disease under study in humans. For the animal testing to be relevant, the animal species where researchers believe the drug action—for the particular drug being tested and its administration route—to be most similar to that in humans is chosen.<sup>11</sup> Also, the cause and mechanism of the disease and the PK/PD and toxic properties of the drug must be similar between the animal species chosen and humans.<sup>12</sup> The experimental drug may also be tested in healthy animals (healthy animal models), with the adverse events, other safety measures and differences in PK<sup>13</sup> and PD parameter results being compared to findings from animals of the same species with the disease. Mice, rats,

nonhuman primates, cats, dogs, pigs or sheep are commonly used animals in experimental studies. Mice and rats comprise the vast majority of animals used in drug testing because they are abundant (due to rapid reproduction) and small in size, and have many genes in common with humans.<sup>14</sup> Nonhuman primates are close to humans in their genetic make-up, immune system,<sup>15</sup> brain complexity and tolerance for pain. Dogs are used because they have cardiovascular<sup>16</sup> and respiratory systems comparable to humans and cats are used because they have neurological diseases comparable to humans.<sup>17</sup> Pigs are similar to humans in their function and organ sizes, while sheep are docile in nature and are similar to humans in size and physiology.<sup>18</sup> The obtained experimental data are analyzed using PK/PD models, which are simulation studies of the action of the body on the drug and the action of the drug on the body, respectively.

### Pharmacokinetic studies

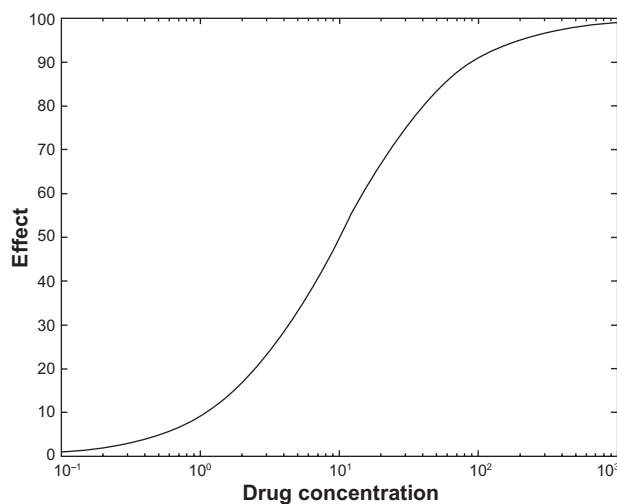
PK studies are used to determine the relationship between the drug dose administered to the subject and the drug concentration that is measurable in the plasma, urine or tissue. This relation is determined by the absorption, distribution, metabolism and excretion of the drug by the body (jointly referred to as drug transport). The primary PK parameters are the volume of distribution ( $V_d$ ) and clearance (CL).<sup>19</sup>  $V_d$  is a measure of how the drug is distributed in the body and CL is a measure of how quickly the body is able to eliminate the drug.  $V_d$  and CL are believed to follow allometric scaling laws that are based on the subject's body weight.<sup>20–22</sup>  $V_d$  and CL together determine the elimination rate constant of the drug, as described by the equation in one of the sections below.

### Pharmacodynamic studies

PD studies are used to determine the association between drug concentration in the body and drug effect; they are commonly related by Equation 1 (the sigmoid  $E_{\max}$  model), which is based on the binding of the drug to a receptor.<sup>19,23</sup>

$$\text{Effect} = \frac{E_{\max} C^\gamma}{EC_{50}^\gamma + C^\gamma}, \quad (1)$$

where  $C$  is the drug concentration,  $E_{\max}$  is the maximum drug effect and  $EC_{50}$  is the concentration required to achieve half this maximum effect. The coefficient  $\gamma$  determines the slope or steepness of the curve in relation to that of the  $E_{\max}$  model. The  $E_{\max}$  model, a commonly used PD model, is a special case of the sigmoid  $E_{\max}$  model and is obtained by setting  $\gamma$  equal to 1.  $E_{\max}$  and  $EC_{50}$  are the primary PD parameters. Figure 1 illustrates  $E_{\max}$  in theophylline, a bronchodilator that is used



**Figure 1** A graph of drug effect versus drug concentration in the body based on the  $E_{\max}$  model (Equation 1 with  $\gamma = 1$ ). The x-axis is in the log scale. In this graph,  $E_{\max} = 100$  (say 100% increase – from baseline – in peak air flow to the lung) and  $EC_{50} = 10$  (mg/mL).

to treat asthma by increasing the peak flow of air from the lungs. As shown in the figure,  $E_{\max}$  is the maximum possible increase in the peak air flow due to the action of theophylline.<sup>23</sup> The concentration at which the drug effect is 80% of  $E_{\max}$  is called  $EC_{80}$ ;  $EC_{20}$  is defined similarly.

### Pharmacokinetic/pharmacodynamic models

PK/PD models are created to estimate PK parameters (absorption and elimination rates, volume of distribution and clearance) and PD parameters (maximum drug effect and the concentration at which the effect is half the maximum) using the obtained drug concentration versus time and concentration versus effect curves from an individual or a sample of individuals (population). The PK model for drug transport can be either a compartmental or a noncompartmental model (Figure 2). The compartmental model fitted to the plasma drug concentration data depends on the route of drug delivery, for example intravenous (IV), intramuscular, subcutaneous, oral, nasal, buccal, sublingual, vaginal, rectal or topical, and on how quickly the drug is distributed to various parts of the body. If the drug is rapidly distributed fairly uniformly throughout the body, the body can be approximated to behave as a single compartment. A two-compartment model is required if the data for the time course of plasma drug concentration (say from an IV bolus dose) show two phases: an initial fast phase of decay and a later slower phase of decay. This indicates that the drug is being distributed more quickly to some parts of the body such as the brain and heart than to other parts such as tissues and bone.<sup>24</sup> A two-compartment model includes a central

compartment to which the drug distributes rapidly and a peripheral compartment to which the drug distributes more slowly. Occasionally, a three-compartment model is used to analyze drug transport. A noncompartmental model makes no assumptions about the number of compartments and derives basic PK parameter estimates using two fundamental areas that are estimated from the PK data (see below). Compartmental models can be used for sparse sampled studies when population PK methods are applied and are not typically used for rich sampled studies, where noncompartmental models can be used. For a compartmental model, one can write an explicit equation for the time course of plasma drug concentration; however it can be a challenge to choose or create the right model (say one-versus two-compartments) that best fits all the data. In contrast, such a choice is not an issue with noncompartmental modeling, but the accuracy of the numerically obtained AUC, AUMC estimates depends on how frequently plasma samples are taken. Thus depending on the action of the body on the drug and the available data, appropriate PK models are fitted, and PK parameters are estimated using appropriate equations.

### One-compartment linear PK models

A drug administered by a single IV bolus dose enters the blood stream directly without any absorption involved and begins to be eliminated immediately in an exponential manner, as Equation 2 shows for the time course of plasma drug concentration for a one-compartment linear PK model.<sup>19</sup>

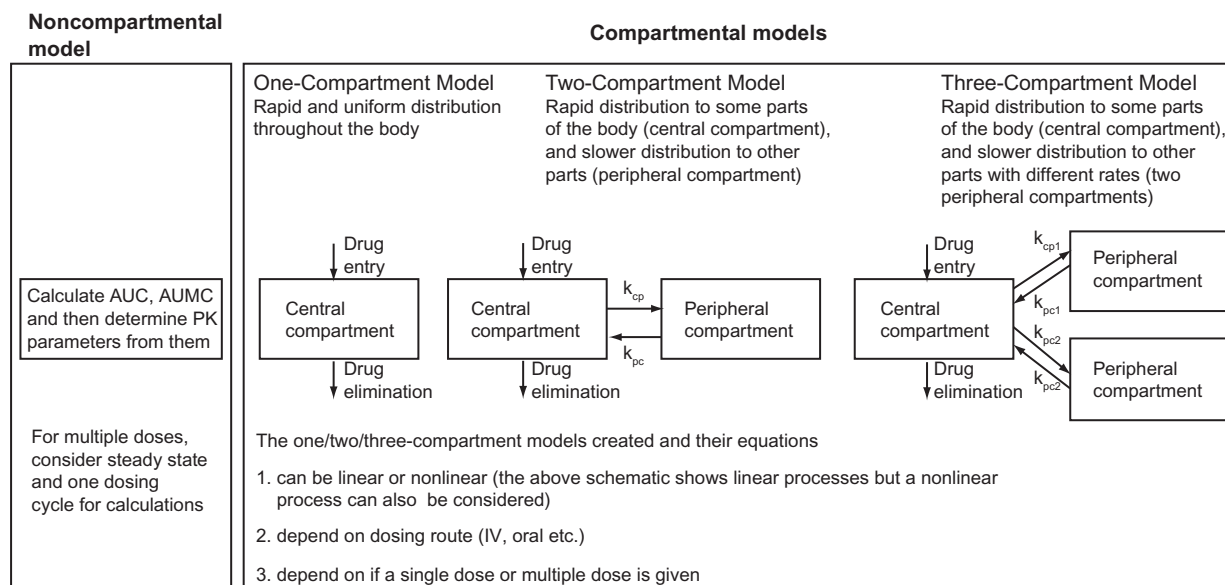
$$C(t) = C_0 \exp(-k_{el} \cdot t), \quad (2)$$

where  $C_0$ , the initial plasma drug concentration at time  $t = 0$ , is given by  $(\text{Dose}/V_d)$ , and  $k_{el}$  is the elimination rate constant. Equation 2 is the solution to the equation  $dC/dt = -k_{el} \cdot C$ , where drug elimination is assumed to follow first order kinetics, which is the assumption of a linear model. For a one-compartment linear model,  $k_{el} = CL/V_d$ .

For a continuous IV infusion, the plasma drug concentration over time is given by:

$$C(t) = \frac{k_0}{k_{el} \cdot V_d} [1 - \exp(-k_{el} \cdot t)], \quad (3)$$

where  $k_0$  is the infusion rate and the other parameters are as described above.<sup>24</sup> While drug elimination is assumed to follow first order kinetics in deriving Equation 3, the infusion process (rate constant  $k_0$ ) follows zero order kinetics. When the infusion is stopped, the plasma drug concentration decays exponentially (Figure 3). Depending on the drug and situation, a combination of an IV bolus



**Figure 2** Schematic of compartmental and noncompartmental models for plasma drug concentration.<sup>24</sup>  
**Abbreviations:** AUC, area under the concentration (versus time) curve; AUMC, area under the first moment curve.

dose and IV infusion can be employed for administration; for example, an IV bolus dose and a maintenance infusion, or a fast loading infusion and a slow maintenance infusion can be administered.

An orally administered drug is first absorbed before it is eliminated, and the time course of plasma drug concentration for a single oral dose is given by:

$$C(t) = \frac{F \cdot Dose \cdot k_a}{V_d(k_a - k_{el})} [\exp(-k_{el} \cdot t) - \exp(-k_a \cdot t)], \quad (4)$$

where  $k_a$  is the absorption constant,  $F$  is the bioavailability, the extent of drug available for use by the body, and the other parameters are as described earlier, and both the absorption and elimination process are assumed to follow first order kinetics.<sup>19</sup> The value of the peak of the drug concentration versus time curve (Figure 4) depends on  $F$ . For a slow release oral tablet, the time ' $t$ ' can be replaced by ' $t-t_{lag}$ ', where  $t_{lag}$  is the time between the oral administration and the onset of absorption of the drug.<sup>19</sup>

**Multiple dosing one-compartment linear PK models**

Normally, the subject is not given a single dose, but is given a drug dose at regular time intervals. For multiple doses, for example via IV administration or oral dosing, drug accumulation occurs in the body if the next drug dose is administered before the previous drug dose is eliminated. This accumulation leads to peaks and troughs in the plasma drug concentration at steady state (Figure 5), and the ratio

of the concentration maximum to minimum depends on the dosing interval and elimination rate constant. Shorter dosing intervals lead to smaller differences between peaks and troughs. The dosing schedule needs to be such that the maximum drug concentration is below the toxic level and the minimum drug concentration is above the therapeutic or minimum effective level.

The time course of plasma drug concentration for multiple IV dosing is,<sup>24,a</sup>

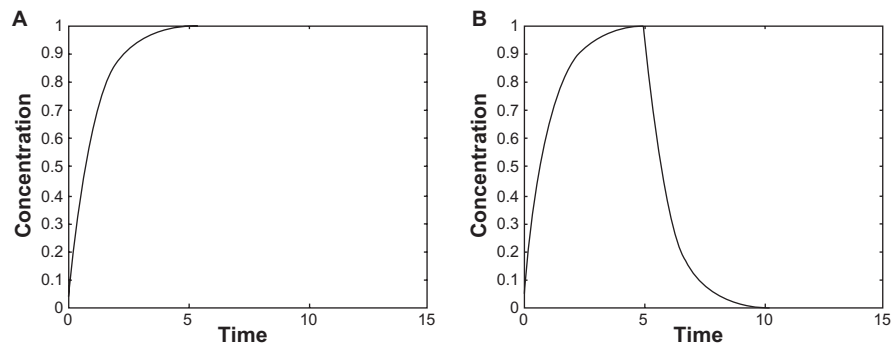
$$C_n(t) = \frac{Dose}{V_d} \left[ \frac{1 - \exp(-n \cdot k_{el} \cdot t_2)}{1 - \exp(-k_{el} \cdot t_2)} \right] \exp(-k_{el} \cdot t), \quad (5)$$

$$C_{max} = \frac{Dose}{V_d} \left[ \frac{1}{1 - \exp(-k_{el} \cdot t_2)} \right], \quad (6)$$

$$C_{min} = \frac{Dose}{V_d} \left[ \frac{\exp(-k_{el} \cdot t_2)}{1 - \exp(-k_{el} \cdot t_2)} \right], \quad (7)$$

where  $C_n(t)$  is the concentration at any time  $t$  after  $n$  doses,  $t_2$  is the dosing interval, and  $C_{max}$  and  $C_{min}$  are the steady state maximum and minimum concentrations (Figure 5).

<sup>a</sup>The reference for all the equations described in this section and those below until the end of the section on noncompartmental PK models is the introductory pharmacokinetics course website of Prof. David Bourne,<sup>24</sup> where there is a lucid explanation of these concepts.



**Figure 3** Time course of plasma drug concentration due to continuous IV infusion: **A**) at steady state (infusion never stopped); **B**) the infusion is stopped at a certain time, leading to an exponential decay in plasma drug concentration after that.

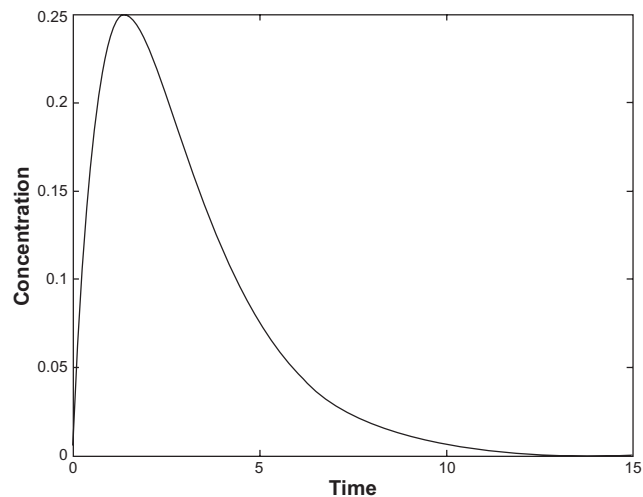
The time course of plasma drug concentration for multiple oral dosing is:

$$C_n(t) = \frac{F \cdot Dose \cdot k_a}{V_d(k_a - k_{el})} \left[ \left( \frac{1 - \exp(-n \cdot k_{el} \cdot t_2)}{1 - \exp(-k_{el} \cdot t_2)} \right) \exp(-k_{el} \cdot t) - \left( \frac{1 - \exp(-n \cdot k_a \cdot t_2)}{1 - \exp(-k_a \cdot t_2)} \right) \exp(-k_a \cdot t) \right], \quad (8)$$

$$C_{min} = \frac{F \cdot Dose \cdot k_a}{V_d(k_a - k_{el})} \left[ \left( \frac{1}{1 - \exp(-k_{el} \cdot t_2)} \right) - \left( \frac{1}{1 - \exp(-k_a \cdot t_2)} \right) \right], \quad (9)$$

$$\bar{C} = \frac{F \cdot Dose}{V_d \cdot k_{el} \cdot t_2}, \quad (10)$$

where  $C_n(t)$  and  $C_{min}$  are as defined above and  $\bar{C}$  is the average concentration in one dosing period during steady state



**Figure 4** Time course of plasma drug concentration for a single oral dose assuming a one-compartment linear PK model.

(Figure 5). It is not easy to calculate  $C_{max}$  due to the difficulty in calculating the time at which the concentration peak occurs.<sup>24</sup>

**Two-compartment linear PK models**

The equations for the time course of plasma drug concentration for a two-compartment linear PK model are derived using techniques similar to those used for a one-compartment linear PK model. For a single IV bolus dose, the time course of plasma drug concentration is as follows (Figure 6).

$$C(t) = A \exp(-a \cdot t) + B \exp(-b \cdot t). \quad \text{with } a > b \quad (11)$$

where  $A, B, a,$  and  $b$  are parameter estimates determined from experimental PK data using the method of residuals and these parameters can be used to estimate the rate constants  $k_{cp}$  and  $k_{pc}$  (which represent the drug distribution from the central to the peripheral compartment and vice-versa respectively) and the elimination rate constant  $k_{el}$  as follows.

$$k_{pc} = \frac{A \cdot b + B \cdot a}{A + B} \quad (12)$$

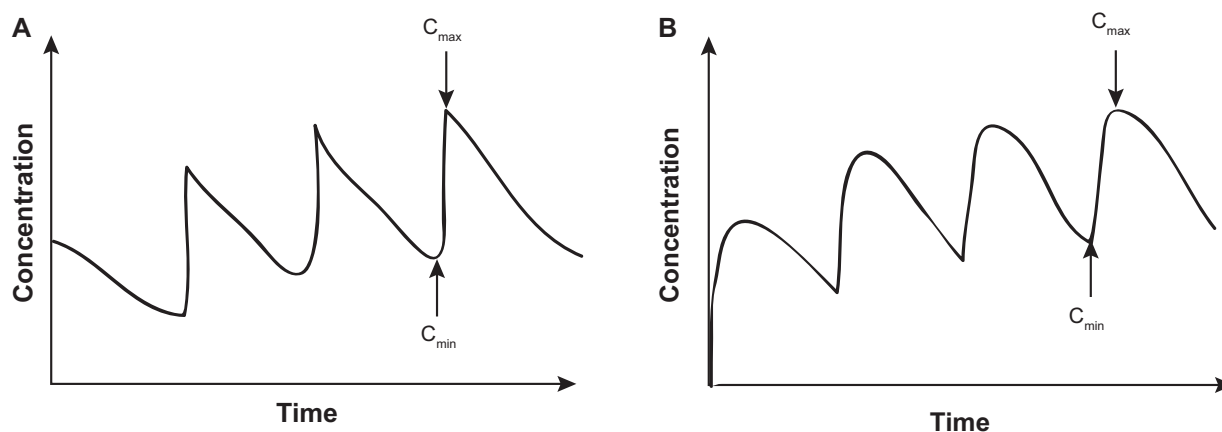
$$k_{el} = \frac{a \cdot b}{k_{pc}} \quad (13)$$

$$k_{cp} = a + b - k_{pc} - k_{el}. \quad (14)$$

The ratio  $k_{cp}/k_{pc}$  determines the nature of the plasma drug concentration versus time curve: for a high  $k_{cp}/k_{pc}$  ratio, the curve shows a clear distribution into the peripheral compartment. Also, for the same  $k_{cp}/k_{pc}$  ratio, the first decay phase is slower if the individual  $k_{cp}$  and  $k_{pc}$  magnitudes are smaller than when both are larger.<sup>24</sup>

For a single oral dose, the time course of plasma drug concentration is (Figure 6):

$$C(t) = A \exp(-a \cdot t) + B \exp(-b \cdot t) + C \exp(-k_a \cdot t), \quad (15)$$



**Figure 5** Schematic for the time course of plasma drug concentration for multiple dosing assuming a one-compartment linear PK model: **A)** IV administration, **B)** oral dosing.

where  $A + B + C = 0$ , and  $a$ ,  $b$  and  $k_a$  are parameter estimates determined using the method of residuals if they are sufficiently separated.

#### One-compartment nonlinear PK model

In the linear models discussed above, drug elimination is a first order kinetic process, per definition, but this may not always be the case. When the metabolism and elimination depend on the drug dose, in many cases the time course of plasma drug concentration for a one-compartment PK model (single IV bolus dose) is given by:

$$\frac{dC}{dt} = -\frac{V_m \cdot C}{K_m + C}, \quad (16)$$

where  $K_m$  is known as the Michaelis–Menten constant, and it is the concentration at which the metabolism rate is half the maximum rate,  $V_m$ . Equation 16 reduces to the first order linear equation  $dC/dt = -k_{el} \cdot C$  ( $k_{el} = V_m/K_m$ ) for low concentrations and to the zero order equation  $dC/dt = -V_m$  for high concentrations. The first order linear equation holds in most cases since  $K_m$  is generally much larger than  $C$ . However, for intermediate concentrations as may occur with multiple dosing, Equation 16, a nonlinear equation, needs to be solved.<sup>24</sup> A detailed analytic approach to solving Equation 16, starting with the solution to the equation  $dC/dt = -k_{el} \cdot C$ , is discussed by Beal.<sup>25</sup>

#### One-compartment linear PK model for drug concentration in urine, and drug concentration in tissue

In addition to the above analyses for plasma, it is useful to analyze the time course of drug concentration in urine, since PK parameters such as the elimination rate constant  $k_{el}$  can be estimated from the data. For a one-compartment linear PK model (single IV bolus dose) with two parallel pathways of elimination,

namely metabolism and excretion, the time rate of change of  $U$ , the amount of drug excreted in urine, is given by:

$$\frac{dU}{dt} = k_e \cdot Dose \cdot \exp(-k_{el} \cdot t), \quad (17)$$

where  $k_{el} = k_e + k_m$ ;  $k_e$  and  $k_m$  are first order rate constants for excretion and metabolism respectively. From Equation 17, the slope of the log of  $dU/dt$  yields  $k_{el}$ .

It is also important to monitor the time course of drug concentration in fat tissue since it is possible that the drug is absorbed into tissue and is not eliminated quickly from there, although it is being quickly eliminated from the blood.

#### Noncompartmental PK models

In noncompartmental PK models, the area under the curve (AUC) and AUMC (from the first moment curve) are first determined using the trapezoidal rule and are then used to estimate various PK parameters; for example, if both IV and oral data are available for the same drug, the following PK parameters can be calculated as<sup>24</sup>:

$$\begin{aligned} \text{Average elimination rate constant } k'_e & \\ &= \text{AUC/AUMC (from IV data)} \end{aligned} \quad (18)$$

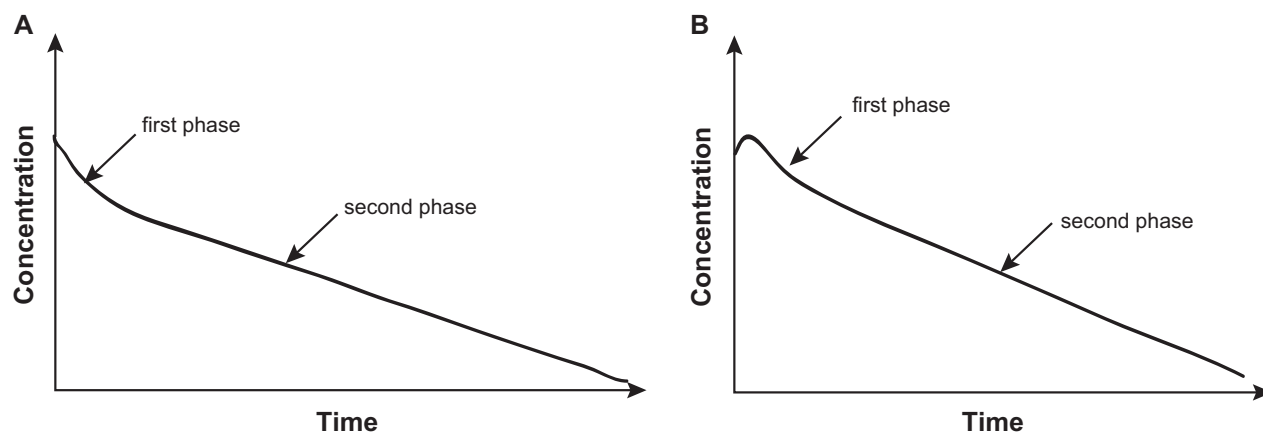
$$\text{Clearance CL} = \text{Dose/AUC (from IV data)} \quad (19)$$

$$\begin{aligned} \text{Steady State Volume } V_{ss} & \\ &= \text{CL} * \text{AUMC/AUC (from IV data)} \end{aligned} \quad (20)$$

$$\begin{aligned} \text{Average absorption rate constant } k'_a & \\ &= 1/(\text{AUMC/AUC (oral)} \\ &\quad - \text{AUMC/AUC (IV)}) \end{aligned} \quad (21)$$

$$\begin{aligned} \text{Bioavailability F} & \\ &= (\text{AUC (oral)} * \text{Dose (IV)}) / \\ &\quad (\text{AUC (IV)} * \text{Dose (oral)}). \end{aligned} \quad (22)$$





**Figure 6** Schematic for the time course of plasma drug concentration assuming a two-compartment linear PK model: **A)** single IV administration, **B)** single oral dose.

### PD models and immediate, delayed or cumulative drug effect

PD studies are also used to determine the time course of drug effect ie, if the effect is immediate, delayed or cumulative. For immediate drug effect, the drug effect versus time curve is classified into three types depending on the initial drug concentration (at time 0)  $C_0$ . If  $C_0$  is greater than  $EC_{80}$ , the curve shows little change even for large changes in concentration, while for  $C_0$  between  $EC_{20}$  and  $EC_{80}$ , the curve is close to a straight line, and finally for  $C_0$  less than  $EC_{20}$ , the curve is close to exponential.<sup>23</sup> A delayed drug effect is mainly due to the time it takes for a drug to travel to the site of action, and to act at the receptor site and produce an effect.<sup>23</sup> For a delayed drug response, a PD model – similar to a PK model that relates concentration and time – can be created to determine the drug effect versus time. The nature of such a curve depends on the equilibration half-life, whose main determinants are the volume of distribution and clearance. The drug response, for some drugs, is related to the drug's cumulative action, such as the AUC (single dose), average concentration at steady state (daily dose rate) or the area under the time versus effect curve (cumulative dose).<sup>23</sup>

In summary, understanding the PK/PD properties of the drug from preclinical studies is crucial in planning a human clinical trial. Any toxic and carcinogenic effects of the drug are investigated during the preclinical phase, and an initial safe dosage for humans and the optimal dosing frequency are generally estimated based on the obtained preclinical parameters by extrapolating them to humans using allometric scaling laws. It is also important in designing a human clinical trial to have an idea of the variability in the drug's PK/PD properties in different subjects due to

differences in their attributes, such as their age and their health condition. For example, the dosing regimen of a drug used in young children and the elderly needs to be altered from that of the rest of the population since their bodies transport the drug differently. Further, subjects with renal problems eliminate the drug more slowly than normal subjects do due to an altered clearance, and they need to be given a smaller drug dose, or the same dose with a longer dosing interval.

Standard computational tools used to estimate PK/PD parameters are NONMEM (NONMEM Manuals, especially Part 5;<sup>26</sup> see a sample program in the Appendix) and Win-NonLin. These tools help estimate PK/PD parameters from an individual or from a sample of individuals and the associated intra- and or intersubject variability. These PK models which consider both fixed effects (dose and time) and random effects (within-subject and between-subject variability) are called nonlinear mixed effects models.<sup>27</sup>

### Case studies from literature to determine the PK/PD properties of a drug

Below are two illustrations of PK/PD models in animals and humans.

- a) The first example concerns dasatinib, used to treat chronic myeloid leukemia (CML),<sup>28</sup> and shows what can be learnt about a drug's PK/PD parameters in humans by first analyzing its PK/PD properties in an animal model of disease.

#### Experimental data and modeling methods

The PK/PD properties of dasatinib were analyzed using K562 human CML xenografts grown in severe combined immunodeficient mice. The goal was to estimate the efficacious plasma dasatinib concentration in mice and then

extend the results of the animal model to humans to estimate the efficacious plasma dasatinib concentration in humans, as a means to facilitate the drug's clinical development. In order to do so, nonlinear mixed effects modeling was performed using the SAAM II software. The PK modeling used plasma drug concentration data from 5 mg/kg IV administration and 1.25, 2.5, and 5 mg/kg oral dosing of dasatinib.

## Results

A two-compartment linear PK model was used to fit the PK data from dasatinib IV administration and oral dosing. The following PK parameter estimates were obtained: the volume of distribution of the central compartment of 2 L/kg, the absorption and elimination coefficients of 0.3/h and 6.89/h respectively, and the inter-compartmental distribution constants of 4.4/h and 2.7/h. An indirect inhibitory  $E_{\max}$  PD model was fitted to the data obtained from 1.25 and 2.5 mg/kg dasatinib oral dosing to study the delay between the plasma drug concentration and the inhibition of phospho-BCR-ABL, an enzyme and a biomarker of CML. The parameter estimates obtained were the  $EC_{50}$  of 6.5 ng/mL, the rate constant of the de-phosphorylation of phospho-BCR-ABL of 2.52/s, the sigmoidicity factor of 4.15 and the  $EC_{90}$  (the plasma drug concentration needed to inhibit 90% of tumoral phospho-BCR-ABL) of 10.9 ng/mL. By accounting for the differences in plasma protein binding in mice and humans, the human  $EC_{90}$  was estimated to be 14.6 ng/mL. In summary, direct allometric scaling for dosage could not accurately predict the clinical efficacious regimen in this case, while a scaling for  $EC_{90}$  could. The total daily dose of dasatinib for humans recommended from this study is 140 mg, which can maintain the plasma drug concentration above the  $EC_{90}$  of 14.6 ng/mL for 8 to 10 hours (the preclinical model predicted a slightly smaller minimum time requirement of 6 to 8 hours).

b) The second illustration involves a PK study of trastuzumab (Herceptin<sup>®</sup>), a current standard therapy for human epidermal growth receptor-2 (HER2) positive breast cancer.<sup>29</sup>

## Experimental data and modeling methods

Four hundred seventy-six cancer patients in phase I, II, and III studies contributed to the PK data. The patients in the phase I study had advanced solid tumors, while the patients in the phase II and III trials had metastatic breast cancer. Patients in the phase I study received a single Herceptin<sup>®</sup> dose of 10–500 mg; patients in phase II obtained a 250 mg loading dose and a 100 mg weekly dose and those in the pivotal phase II and III studies received a loading dose of 4 mg/kg followed by 2 mg/kg weekly for up to 840 days. Herceptin<sup>®</sup>

was administered by IV infusion in all studies until disease progression. Three thousand eight hundred forty-nine unique and evaluable PK data points were obtained. Nonlinear mixed effects modeling conducted using NONMEM was used to analyze and interpret the population data.

## Results

Model results showed that a two-compartment linear PK model with zero-order input fit the data the best. Population PK estimates obtained from the analyses were those for clearance, 0.225 L/day, and volume of distribution of 2.95 L of the central compartment of Herceptin<sup>®</sup>, and also its half life of 28.5 days. Interpatient variabilities for drug clearance (43%) and volume of distribution (29%) were found to be large, especially for drug clearance. Significant baseline factors for clearance and or volume of distribution were the number of metastatic sites, plasma level of extracellular domain of the HER2 receptor, and patient weight; however, compared to the large inter-patient clearance variability, these covariates were not clinically important. The estimate for clearance was unaffected by concomitant therapy (anthracycline plus cyclophosphamide, or paclitaxel).

## Phase I clinical trial

For most new drugs, the first human clinical study is a phase I trial.<sup>b</sup> A phase I trial is conducted after the drug's PK/PD properties are well understood from preclinical *in vitro* or animal studies. The experimental drug is tested on a small number (say one to two dozen) of (usually) healthy subjects. The trial is used to determine factors including the best route of drug administration, the minimum therapeutic dose and maximum tolerated dose (MTD) or safe drug dose in humans. Single dose studies are often performed to find the MTD. In this regard, many trials employ the classical 3 + 3 design of dose escalation, with the starting dose being a conservative estimate based on the preclinical results. However, presently some trials use Bayesian or other dose escalation methods to determine the MTD.<sup>31–34</sup> A phase I trial also evaluates the PK/PD properties of multiple doses of the drug, such as its absorption and elimination characteristics, its dosing schedule and its concentration versus effect properties. Thus,

<sup>b</sup>In rare cases, the first human trial is a phase 0 or a micro-dosing study to evaluate the drug's PK/PD properties, performed in special cases where an experimental drug shows a significant potential in the preclinical stage to warrant an acceleration of its development and a reduction in its costs.<sup>30</sup> If the results in humans are encouraging, further clinical trials and studies can proceed quickly.

a phase I study helps estimate the drug dosage and dosing schedule that could be used in the phase II, III trials.

## Preliminaries for phases II and III clinical trials

Several important concepts and terms related to phases II and III studies are explained before phases II and III trials are discussed.

### Types of phases II and III clinical trials

Different types of phases II and III clinical trials can be performed, as described below.

#### Superiority, noninferiority and equivalence trials

**Superiority trial:** The aim of a superiority trial, the most common type of phase II or phase III clinical trial performed, is to test whether an experimental drug is superior in efficacy to a placebo or to an existing standard treatment (control arm).<sup>35</sup> Assessment of superiority of the new drug is usually conducted by estimating the test statistic  $P$ -value, a quantitative statistical measure of the difference in the primary efficacy outcome between the experimental and control arms. The  $P$ -value quantifies the likelihood of the observed difference in effect between the treatment arms to be due to chance. The smaller the observed  $P$ -value, the more unlikely it is that this difference is due to chance, ie, a real difference exists. Typically, a  $P$ -value of  $\leq 0.05$  indicates a statistically significant difference (beyond chance) between the treatment arms. When the result of a superiority trial is statistically significant, the next step is to determine whether the treatment difference is also clinically meaningful, eg, a 1 unit (mm Hg) difference in systolic blood pressure may be statistically significant but is obviously not of any clinical value. Superiority trials with more than two treatment arms have also been conducted.<sup>36–38</sup>

**Noninferiority trial:** A noninferiority trial,<sup>35,39</sup> less common than a superiority trial, tests whether the efficacy of a new experimental drug is no worse than that of an existing standard treatment (control arm). In this trial design, efficacy is determined by comparing the difference in the proportion of events<sup>c</sup> (assume a negative outcome such as death) between the new drug arm ( $p_N$ ) and the control arm ( $p_C$ ) to a predetermined positive noninferiority margin  $\Delta$ , with the choice of  $\Delta$  being based on clinical relevance. Assessment of noninferiority is conducted using confidence intervals

of  $p_N - p_C$  rather than  $P$ -values.<sup>40</sup> Under the null hypothesis of a noninferiority trial, at 5% significance level, the lower bound of the 95% confidence interval of  $p_N - p_C$  should be  $\geq \Delta$ . The condition of noninferiority is met if the upper bound of the 95% confidence interval of  $p_N - p_C$  is  $< \Delta$  and implies that the efficacy of the new drug is no worse than that of the control. If the upper bound of the 95% confidence interval of  $p_N - p_C$  is also  $< 0$ , the condition of both a noninferiority and a superiority trial may be fulfilled. Even if the superiority condition is not met and only the noninferiority condition is met, it is possible for the new drug to exhibit better production costs and a superior safety profile than that of the standard treatment, and hence to be preferred to the standard treatment. Noninferiority studies can also be carried out for more than two treatment arms, and appropriate statistical testing procedures have been developed for some of these situations.<sup>41,42</sup>

**Equivalence trial:** In an equivalence trial, the aim is to confirm that there is no statistically significant difference in the effectiveness of the new and standard treatment/placebo (control arm).<sup>43</sup> Efficacy is determined by comparing the difference in the proportion of events between the new drug arm ( $p_N$ ) and the control arm ( $p_C$ ) to a predetermined interval, with the choice of the interval being based on clinical relevance. The two treatments are considered equivalent if the two-sided 95% confidence interval of  $p_N - p_C$  lies entirely within the interval (say  $(-\Delta, \Delta)$ ; note that in some cases the interval may be chosen to be asymmetric about zero). Equivalence trials can also be designed and analyzed for more than two treatments and some of the issues involved in proving equivalence in three rather than two treatment arms are discussed by Wiens and Iglewicz.<sup>44</sup>

#### Open label versus blinded trials

The term blinding in a clinical trial implies that some of the key persons involved in the trial, such as the subjects, physicians, and biostatisticians, are not made aware of which treatment the subject is undergoing. Blinding procedures are followed to eliminate selection bias and increase objectivity by avoiding the physician's knowledge of treatment allocation from influencing decisions related to the subject's treatment and care, and avoiding the subject's knowledge of treatment allocation from affecting his or her response to the treatment. There are different levels of blinding as described below.

**Open label trial:** In an open label trial, both the subject and the physician are aware of what drug or treatment the subject is assigned to receive. Such a trial may be randomized or not.

<sup>c</sup>The theory also works for a continuous outcome.

**Blinded trial:** The two main types of blinded clinical trials are:

**Single blinded trial:** In a single blinded trial, the subject is not aware of which treatment arm he/she is assigned to, but the physician is aware of which treatment the subject is to receive.

**Double blinded trial:** Typically, in a clinical trial, neither the subject nor the physician is aware of which treatment arm the subject is assigned to. The treatment code is held by an independent person, usually a statistician or someone not directly involved in the design, conduct, or analysis of the trial. The treatment code is only uncovered at study completion, after the data analyses are completed.

#### Nonrandomized versus randomized trials

**Nonrandomized trial:** In a nonrandomized trial, the treatment arm to which each eligible subject is assigned is not done in a random fashion but is chosen by the subject or investigator.

**Randomized trial:** Most phase II and phase III clinical trials are randomized, which means that eligible subjects are randomly assigned to one of the treatment arms. Using a computer program, the trial biostatistician generates a randomization schedule that assigns a treatment arm to each of the predetermined study numbers. If the trial design is balanced, implying that there is an equal number of subjects in each treatment arm, then each enrolled subject has, in principle, an equal a priori chance of being assigned to either treatment arm. This guarantees objectivity of a clinical trial since neither the participant nor the investigator knows a priori what treatment will be assigned to the participant.

In some situations, a balanced design may not be practical or desired. In such situations, the randomization schedule can be adapted to unbalanced designs or more complex or iterative designs, as described below.

One alternate method to perform randomization is block randomization.<sup>40</sup> As an example, a block randomization with a block size of four and with two subjects each in treatments A and B, can be performed in six ways (AABB, BBAA, ABBA, ABAB, BABA, BAAB). In small sample size trials, block randomization leads to a greater chance of balance among treatment arms with respect to the number of subjects assigned to each arm.<sup>45</sup>

Sometimes, an unequal randomization between two treatment arms, such as a 1:k randomization where  $k > 1$ , is carried out to either reduce trial costs<sup>46</sup> or simply to learn more about the experimental drug being administered<sup>47</sup> by assigning more subjects to the experimental drug arm than to the control arm. Unequal randomization can be implemented

using a block randomization scheme, and the block size is always a multiple of the randomization ratio addition; for example, for a 1:k randomization, a minimum block size of  $1 + k$  is required.

Subjects can also be randomized based on gender, age, and other characteristics, using a procedure known as stratified randomization.<sup>48</sup> Stratified randomization is often implemented in a two step process that first involves a block randomization scheme and then involves the creation of a randomization list per stratum. This process creates a balance of the stratification attributes among the treatment arms. Statistical analyses of such a clinical trial should account for this randomization design by including in the model the stratification factor(s) used in the randomization.

A more recently developed, complex randomization scheme is called dynamic randomization.<sup>48-50</sup> In this method, the probability of a subject being randomized to a certain treatment arm (say A or B) depends on the number of subjects that are already in each arm and their attributes being matched, and the attributes of the subject being randomized. For example, if there are more subjects in arm A than in arm B when a particular subject is being randomized, then the subject may be randomized to arm B with a predetermined probability. However, if the randomization is being matched by gender and if the subject is female and there are less females in arm A than in arm B, then the subject may be assigned to arm A with a certain probability. Thus, the decision as to which arm to assign the next subject is iteratively re-evaluated each time a new subject is randomized. Dynamic randomization can be particularly efficient when the pool of eligible subjects is not very large, or when the number of stratification factors is large.

In summary, a RCT is considered the gold-standard. Randomization not only guarantees that the treatment arms are homogenous but also that known and unknown confounding factors are evenly distributed among the treatment arms. In addition to the statistical efficiency brought upon by randomization, the homogeneity among treatment arms provides confidence that should one treatment arm be found to perform differently than the others, the difference in outcome will be due to the difference in treatments, and not due to any inherent differences in the treatment arms themselves.<sup>45,51</sup>

Note that blinding, randomization, consecutive recruitment<sup>52</sup> and independent assessment of patient response (example tumor response) are some of the main methods to reduce bias in a clinical trial but may not be able to completely eliminate it.<sup>1</sup> The factors that cause bias in clinical

trials can be broadly classified as bias due to selection, observation and statistical procedures.<sup>53</sup> A further detailed categorization of the different types of bias is discussed by Chow and colleagues (see Tables 2.3.1. and 2.3.2 in Chow and colleagues).<sup>53</sup> In addition, publication bias, which is the tendency of researchers, pharmaceutical companies, and editors to handle the reporting of experimental results that are positive differently from results that are negative, also needs to be considered.<sup>54</sup>

### Clinical trial design

Some of the designs used in phase II and phase III clinical trials are described below.<sup>55</sup>

#### Parallel arm design

The parallel arm design is the most common design for a phase II or III trial.<sup>56</sup> Typically, in this design type, each subject is randomized at the beginning of the study to either the experimental drug or the standard treatment/placebo, and continues to receive the same drug throughout the study. Additional treatment arms, which serve to test different dosages of the experimental drug, can be included. The advantage of this design is that it has a low bias due to the lack of any sequence effects while the disadvantage is that the variability between subjects can influence the outcome.<sup>57</sup> However, this disadvantage can be minimized by performing a careful sample size calculation with an accurate estimate of the standard deviation that considers the between subject variability.<sup>57</sup>

#### Factorial design

The factorial design is a hybrid of the parallel arm design, but in this design more than one experimental drug is tested simultaneously.<sup>56,58</sup> In a simple form of this design, some subjects receive treatment A; others receive treatment B; yet others receive a combination of treatments A and B, and the remaining subjects serve as control and receive neither treatment A nor B. This design helps investigate the efficacy of a combination of pairs of treatments. The design's major disadvantage is the implicit assumption that the treatments do not interact but act independently in the body.

#### Cross-over design

The cross-over design is mostly used in chronic conditions, and is a much more uncommon design in phase II or III trials.<sup>2,59</sup> In a cross-over trial of two treatments, a subject who is randomized to receive treatment A is first administered treatment A for a predetermined length of time and is then switched to receive treatment B for the same duration, with a

washout period in between the two treatments. The opposite sequence is used for a subject who is first randomized to treatment B. The advantage of this design is that a smaller number of subjects is enrolled than for a corresponding parallel arm design study. The statistical power and efficiency of this design are achieved through each subject serving as his or her own control, resulting in smaller standard errors of the measures of efficacy. The main disadvantage of this design is that, for valid results, the effects of treatment A and treatment B should not interact with each other or carry over. Otherwise there will be a mixture of effects, which will obscure any efficacy differences that may exist. Also if period effects exist, ie, if there is a change in the response to treatment in the two periods, the conclusions drawn about the treatment effect may not be valid.

### End-points of interest in clinical trials

Generally, two kinds of endpoints are prespecified to assess the data and compare treatment arms: safety endpoint and efficacy endpoint.<sup>40</sup>

A safety endpoint is a clinically relevant measure or an outcome of the study that can be used to assess the safety of the experimental drug. Safety endpoints are defined based on assessments such as adverse event observations, serious adverse events (monitored by pharmacovigilance teams), or based on the change in value on treatment of a particular lab or physical exam parameter. In addition to the primary safety endpoint(s), secondary and tertiary endpoints can be defined.

An efficacy endpoint is a clinically relevant measure or an outcome of the study that can be used to assess the effectiveness of the experimental drug. Disease-free survival, time to tumor progression, reduction of a specific symptom, and changes in biomarker levels are examples of primary efficacy endpoints. Primary, secondary, and tertiary efficacy endpoints are often prespecified because analysis of outcomes that are not prespecified can be viewed with suspicion. Note that as the number of statistical tests increases in an analysis, so does the probability of finding a significant result purely by chance, ie, when no true effect exists. This type of statistical error is called a type I error.<sup>d</sup> To reduce the chances of a type I error, statistical tests should be limited and be confined to *a priori*-defined hypotheses. If multiple testing is conducted, the nominal significance level  $\alpha$  should be adjusted such that

<sup>d</sup>While a type I error relates to the probability of detecting a difference in the outcome variable between the treatment arms, when there is no difference between them, a type II error relates to the probability of failing to observe a difference in the effect or outcome variable between the treatment arms, given that the effect exists.

the overall type I error is maintained at the chosen significance level. If secondary efficacy endpoints are defined for purely exploratory reasons, then no adjustment for multiple testing is required. If multiplicity adjustment for primary or secondary endpoints is required,<sup>60</sup> family-wise  $\alpha$  level adjustment methods such as the Hochberg, Holm, or Bonferroni procedures are used.

### Sample size calculations

To enroll a very large number of subjects in a clinical trial is both inefficient and impractical for time, financial, and other reasons. Also, for example, in the case of rare diseases, it may not be possible to find a large number of subjects. In contrast, if too small a number of subjects is enrolled, it cannot be determined accurately if there is a meaningful difference in the primary end-point among treatment arms. A sample size calculation, required for phase III studies and most phase II studies, estimates the number of subjects needed to be able to detect a statistically significant difference (if one exists) in the primary endpoint among treatment arms, with certain assumptions regarding the type I error ( $\alpha$ ) and type II error ( $\beta$ ) (power is defined as  $1 - \beta$ ). These two types of error are crucial in determining the minimum required sample size. Generally, the type I error  $\alpha$  is set at 0.05, which implies that if a treatment difference is detected, there is a 5% probability that the observed difference will be due to chance. The type II error  $\beta$  is typically set at 0.2 or lower, and correspondingly the power is set at 0.8 (80%) or higher.

Specialized software such as nQuery and EAST are available to calculate the sample size for different study designs, endpoints (continuous, dichotomous, repeated measures, time to failure), effect size,  $\alpha$  and  $\beta$  (see [http://www.statsol.ie/html/nquery/nquery\\_home.html](http://www.statsol.ie/html/nquery/nquery_home.html), <http://www.cytel.com/Software/East.aspx>). The sample size calculation for a superiority trial, parallel arm design with two treatment arms, requires the following information: the assumed  $\alpha$  and  $\beta$ , the randomization ratio between the two arms, the standard deviation of the primary endpoint (continuous outcome) or the event rate (categorical outcome) in each arm, the clinically relevant difference between the two arms that needs to be detected, and the statistical test used to detect this difference, such as a two-sample *t*-test (continuous outcome) or Chi-squared test (categorical outcome) (see Appendix 1).<sup>40</sup> For a continuous outcome and two treatment arms, the sample size for arm 1 ( $n_1$ ) is:<sup>61</sup>

$$n_1 = \frac{(k\sigma_1^2 + \sigma_2^2)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{k\Delta^2}, \quad (23)$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are the variances of the primary endpoint of the two arms,<sup>e</sup>  $\Delta (= \mu_2 - \mu_1)$  is the effect size, ie, the difference in means of the primary endpoint of the two treatment arms,  $Z_{1-\alpha/2}$  is 1.96 for an  $\alpha$  of 0.05 and  $Z_{1-\beta}$  is 0.84 for a  $\beta$  of 0.2 or is 1.28 for a  $\beta$  of 0.1,  $k = n_2/n_1$  is the randomization ratio (hence  $n_2 = kn_1$ ) and a two-sided alternative is considered (see Appendix 1 for corresponding SAS code using Proc Power).

For a dichotomous outcome and two treatment arms, the sample size for arm 1 ( $n_1$ ) is:<sup>61</sup>

$$n_1 = \frac{(Z_{1-\alpha/2}\sqrt{(k+1)\bar{p}\bar{q}} + Z_{1-\beta}\sqrt{kp_1q_1 + p_2q_2})^2}{k\Delta^2}, \quad (24)$$

where  $p_1, p_2$  are the probabilities of success in the two arms,  $q_1 = 1 - p_1, q_2 = 1 - p_2, \Delta = p_2 - p_1, \bar{p} = (p_1 + kp_2)/(1 + k)$  and  $\bar{q} = 1 - \bar{p}$ , the sample size for arm 2 ( $n_2$ ) is  $n_2 = kn_1$  and a two-sided alternative is considered (see Appendix 1 for corresponding SAS code using Proc Power).

For studies with multiple treatment arms, in many cases, it may be possible to estimate the sample size based on two-arm designs. For example, sample size calculations may be done using say placebo versus dose 1, if powering for that case clearly yields enough power for placebo versus dose 2, assuming that the trial has three treatment arms.

Sample size calculations should also consider withdrawals or drop-outs from the study. One approach to deal with attrition is to over-enroll by say 10%–20%, which is estimated from withdrawal rates in pilot studies or prior experience with similar drug compounds or diseases. The additional subjects serve to protect the number of evaluable subjects and to afford the trial with acceptable statistical power. Notably, from an ethical perspective, it is unethical to conduct a clinical trial with inadequate statistical power to be able to observe an effect, if an effect exists. Similarly, from a cost perspective, it is a waste of resources to embark on an underpowered, and hence futile, study.

### Analysis samples for phase II and phase III clinical trials

Appropriate analysis samples need to be defined from the study subjects enrolled, to carry out the statistical analyses. All the samples defined below may not be applicable to

<sup>e</sup>The variability can be due to biological variability (intersubject variability) or temporal variability, measurement error (both intra-subject variability).<sup>53</sup> Biological and temporal variability can be greatly minimized by suitable randomization, study design, and statistical analyses, while measurement error can be minimized but not eliminated and is what drives the sample size; the higher the variability, the larger the sample size.<sup>53</sup>

(or need to be defined for) every study. For example, an open label study with one treatment arm may not require randomization and may define safety and other appropriate samples.

#### Randomized sample

This sample includes all subjects who are screened for the study, fulfill the eligibility criteria, and have been randomized to a treatment arm. The demographic and other baseline characteristics analyses are performed using this sample.

#### Safety sample

This is defined as subjects who are randomized to the study and take at least one dose of the study medication. Safety analyses are performed using this sample.

#### Intent-to-treat sample

This sample is defined to be the same as the randomized sample. Efficacy analyses are performed using this sample, and indicate how the drug would work in the target population. Modified intent-to-treat samples, such as randomized subjects who receive treatment and have at least one post-baseline efficacy measurement, may also be used to perform additional efficacy analyses.

#### Per-protocol sample

This sample is defined as all randomized subjects who continue to fulfill the eligibility criteria, follow the study protocol closely and have all efficacy measurements. Additional efficacy analyses may be performed using the per-protocol sample. Such analyses indicate how the drug would work in ideal conditions.

Note that missing data in a sample always present analytic problems. Two approaches, the last observation carried forward (LOCF)<sup>62,63</sup> and the multiple imputation technique,<sup>64–67</sup> are commonly used when some observations have missing data over the follow-up period. The LOCF method involves carrying forward the last nonempty observation of the subject to create a complete data set. For example, if the last observation for a subject is recorded at visit 5 and the study terminates at visit 8, the observation at visit 5 is used at visits 6, 7, and 8 for this subject. In the multiple imputation method, multiple complete data sets are generated using Markov chain Monte Carlo or regression methods to fill in the missing values, and parameter estimates from each complete dataset are pooled.

## Phase II clinical trial

A phase II trial can enroll a few hundred subjects and often focuses on statistically evaluating and comparing the safety

and efficacy endpoints between the experimental drug arm and control arm (assume a two treatment parallel arm study for simplicity). The control arm can be a placebo or the current standard treatment in the case of most cancer trials and trials of rare diseases. Phase II trials can be exploratory in terms of testing more than one dose level of the experimental drug to determine the optimal dose using the obtained dose-response efficacy curves and the corresponding safety results. The optimal dosing frequency for multiple doses may also be ascertained. Most phase II trials are randomized but may or may not be blinded. They can also be quite varied in terms of study design. In many cases, after completing a successful phase II trial, sponsors submit their protocol and SAP for a phase III trial to the FDA for review. The FDA often provides critical scientific suggestions on improving the study design and conducting a more efficient phase III trial that can also minimize patient risk.

### Case study from literature to determine the safety and efficacy of a drug

A single-arm, open-label phase II study is considered where the efficacy, safety and pharmacokinetics of trastuzumab (Herceptin<sup>®</sup>) monotherapy were evaluated.<sup>68</sup>

#### Experimental methods and analysis

HER2 positive breast cancer women were administered Herceptin<sup>®</sup> monotherapy as first-line treatment: a loading dose of 8 mg/kg IV and then a 6 mg/kg IV was administered once every three weeks until the disease progressed or the patient withdrew from the study. 105 women received a median of five cycles of Herceptin<sup>®</sup>.

#### Results

**Pharmacokinetics:** This study compared its PK results with the earlier study mentioned that had followed a weekly Herceptin<sup>®</sup> regimen<sup>29</sup> and concluded, not surprisingly, that the mean maximum and minimum Herceptin<sup>®</sup> concentrations were higher and lower respectively than those observed with the weekly regimen. The average Herceptin<sup>®</sup> exposure was similar to that in the weekly regimen.

**Safety:** Treatment emergent safety events included pyrexia, headache, nausea, and fatigue. Since decrease in cardiac function is a less common but serious concern with Herceptin<sup>®</sup> therapy, the left ventricular ejection fraction was monitored but was not found to change significantly during the study.

**Efficacy:** The study's efficacy endpoints were overall response rate, clinical benefit rate and time to disease progression. The overall response rate (ratio of subjects with

complete or partial response to the total number of subjects)<sup>f</sup> was 19% and the clinical benefit rate (complete and partial responses plus stable disease for at least 6 months) was 33%. The median time to disease progression was 3.4 months with a range of 0.6 to 23.6 months. The study concluded that the 3-weekly regimen with higher dosage did not affect the efficacy and safety of Herceptin<sup>®</sup>, and may be followed in place of the weekly regimen.

## Phase III clinical trial

A phase III trial can enroll hundreds or thousands of subjects and focuses on evaluating the safety of the experimental drug in this large group and testing or confirming the efficacy observed in the phase II trial. Most phase III trials are double blinded and randomized. In many cases where the optimal dosage has already been decided, a control arm and only one treatment arm of a specific dosage of the experimental drug are compared in a parallel arm study. Generally, the results of two successful, independent phase III trials need to be submitted to the FDA for drug approval (success typically implies a *P*-value less than 0.05 on both trials with regard to the efficacy endpoint assuming a superiority trial), although in certain cases one large study may be considered adequate. Common statistical tests and methods used to analyze phases II and III clinical trials data and present the results to the FDA are described in Appendix 1. Common computational tools used for these analyses are the statistical packages SAS (SAS Institute, Cary, NC) and R.

An independent, unblinded committee called the DMC is set up to ensure subject safety and protection from harm in phase II and phase III (and sometimes phase I) studies involving high risk subjects, high risk therapies or therapies of public health significance.<sup>69</sup> Notably, no IRB will approve the protocol of such studies unless they have a DMC. The DMC is ordinarily comprised of persons such as independent biostatistician(s) and clinicians who work in the disease sub-specialty area that the experimental drug treats, and its purpose is to monitor the data at regular intervals for safety issues. If the DMC finds serious adverse effects, potential harm to subjects or any other unacceptable safety issues of the experimental drug, it can recommend stopping the trial. The DMC also verifies the scientific validity of the data.

In phase II and phase III trials, interim analyses can be conducted to perform a partial statistical analysis of the

data collected at interim stages of the study, for example after 25%, 50%, and 75% of subjects have been enrolled.<sup>70</sup> Generally, the interim analyses are performed by independent biostatisticians using the appropriate statistical programs created by the sponsor; the relevant unblinded results obtained are presented to an independent committee such as the DMC. If the interim analysis results show a statistically significant difference in the primary efficacy endpoint among the treatment arms, the DMC can recommend that the trial be stopped, thus saving resources. However, in many cases, there is no clear evidence of efficacy, possibly because the number of subjects enrolled in the study at the particular interim analysis stage is not large enough to test the hypothesis; the trial is usually continued until the next interim analysis or until all the subjects are enrolled. Note that multiple testing introduced by interim analyses requires that the significance level be appropriately adjusted in order not to inflate the type I error,  $\alpha$ . The most common methods for adjusting for multiplicity view the nominal significance level as a spending function. Each interim analysis is considered to 'spend' a part of the nominal significance level  $\alpha$ , and at each level it is possible to calculate the 'spent' and 'unspent' level so that the overall nominal level is not overspent when the final analysis is conducted. Calculation of the spending function such as the O'Brien–Fleming approach<sup>71–74</sup> and the Pocock approach<sup>75–77</sup> is implemented using sophisticated statistical software.<sup>78</sup> Clinical trial stopping rules, including stopping a trial for futility, are currently an area of active statistical research.

## After a phase III clinical trial

After phase III trials of the experimental drug, the sponsor files a New Drug Application (NDA) with the FDA in the United States or follows a similar procedure elsewhere. The detailed application describes the composition of the drug and its chemical properties, its manufacturing details, its PK/PD properties from preclinical studies, and its safety and efficacy statistical analyses and result interpretations from the human phase I to III trials. The FDA evaluates the new drug with input from a panel of qualified physicians, chemists, pharmacologists, biostatisticians and others and decides whether to approve the drug or not. The decision of approval or nonapproval is a dichotomous one, and is mainly based on the safety and efficacy results provided and the risk to benefit ratio of the drug; in this context, also note that from the definitions of the type I and II error given, the risk to the public is essentially the type I error, and the risk to the sponsor is

<sup>f</sup>To better understand complete response, partial response, stable disease and disease progression categorization, consult the RECIST criteria for solid tumors.



the type II error. In addition to the drug's safety and efficacy, the FDA examines the drug manufacturing methods to ensure its quality and purity, and the accuracy of the drug labeling provided (as an example of labeling, an important component of Herceptin®'s label is its warning on cardiotoxicity, since clinical trials have shown heart damage and impairment in heart function to be a less common but serious side effect of Herceptin®).

Once the drug is approved, it can be marketed. In the case of some drugs, the sponsor conducts a phase IV study (also called a postmarketing or pharmacovigilance study). A phase IV study is a long-term follow-up study, carried out postapproval on a large number of patients to assess the long-term effects, safety, and effectiveness of the drug.

### Other contemporary clinical trial designs: adaptive designs

Adaptive design trials are the clinical trials of the future, designed with the aim of increasing the efficiency of the drug-testing process and or reducing costs. These trials allow changes in the middle such as re-estimating sample size (and the option to enroll more patients at the time: the randomization ratio between treatment arms may be changed and or the number of sites may be increased at this time), and early stopping for efficacy and or futility.<sup>79</sup> One example of an adaptive design trial is a study where the phases II and III are combined and one treatment out of the treatments being tested is selected in the interim.<sup>79</sup> In summary, adaptive design trials allow one to learn from the interim data (such as obtaining better estimates of the effect size and variance, and then re-estimating the sample size) and to make appropriate trial design changes in the middle. However, trial modifications cannot be done ad hoc and need to be prespecified in the trial design and in the protocol, to maintain the trial's integrity and obtain valid results.

### Conclusion

Clinical research and clinical trials have gained greater importance in recent years, with several new clinical research organizations opening up in various parts of the world and conducting a large number of trials related to a wide range of diseases. Clinical trials afford subjects the opportunity to gain access to a new, more efficacious and safer medication to alleviate or cure their disease than the standard treatment available. There are potential risks and side effects of every new drug. However, such trials and studies are crucial in making progress in this field and in putting out rigorously tested drugs that are safe and effective in treating the related disease.

Every new drug undergoes comprehensive testing in various phases before it is put out in the market. In the pre-clinical phase, *in vitro* and animal studies of the drug are conducted and the obtained data are analyzed by creating PK/PD models that help determine the absorption, distribution, metabolism, excretion properties, and the effects of the drug on the body. Another important aspect that is investigated is whether the drug displays any toxic effects or carcinogenic properties, and if so at what dosing level. Thus, preclinical studies also provide estimates for an initial safe dosage that can be given to humans and the dosing frequency of the drug. After the preclinical studies are completed, the drug undergoes thorough testing in humans in (generally) three phases before it is approved. Phase I studies are carried out on a small number of subjects to assess the tolerability and safety of the drug in humans. Phase II studies typically involve a few hundred subjects and analyze the safety and efficacy properties of the drug in treating the condition in humans. Phase III studies can involve thousands of subjects and analyze the safety and efficacy properties of the drug in this larger group. Phase IV studies, which are long-term pharmacovigilant studies after the drug is on the market, are carried out in the case of some drugs. Presently, clinical trials are moving towards adaptive designs that often use Bayesian approaches and allow one to learn from the interim data and make appropriate changes in the conduct of the trial in the middle. By learning from mathematical modeling of preclinical data and from the interim analyses of clinical trial data, the ultimate aim is to design streamlined, efficient trials that reduce the number of subjects enrolled, as well as trial time and costs.

The entire drug-testing process involves detailed medical studies, mathematical modeling of the preclinical data, statistical analysis of the clinical trials data and requires the expertise of biochemists, molecular biologists, pharmacologists, physicians, mathematicians, statisticians and others. The testing requires collaboration, and interpretation of results among experts from various areas working together on the same drug. This holistic approach will provide keener insights into how the particular drug works. Such a unified approach has not always been the case in the past and the different sub-specialties of clinical research have been somewhat compartmentalized. However, the present trend in clinical research is a unified approach and it is hoped that even more integrated studies of new drugs will continue to be carried out, leading to some novel drugs that are highly effective in curing the associated condition.

## Disclosures

The authors report no conflicts of interest in this work.

## References

- Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and practice*. Philadelphia, PA: John Wiley & Sons, Inc.; 2002.
- Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. 3rd edition. Chicago, IL: Springer; 1999.
- Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ*. 2000;321(7259):504.
- Kolahi J, Bang H, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. *Community Dent Oral Epidemiol*. 2009;37(6):477–484.
- Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Control Clin Trials*. 2004;25(2):143–156.
- Bernstein D, Simmons J. Navigating differences between FDA and EMEA for regulatory compliance during drug development. *Bio Pharm Int*. 2006;19:16–23.
- Ono S, Yoshioka C, Asaka O, et al. New drug approval times and clinical evidence in Japan. *Contemp Clin Trials*. 2005;26(6):660–672.
- Malinowski HJ, Westelink A, Sato J, et al. Same drug, different dosing: differences in dosing for drugs approved in the United States, Europe, and Japan. *J Clin Pharmacol*. 2008;48(8):900–908.
- Available from: <http://www.ich.org/LOB/media/MEDIA486.pdf>. Accessed on February 10, 2010.
- Smith-Tyler J. Informed consent, confidentiality, and subject rights in clinical trials. *Proc Am Thorac Soc*. 2007;4(2):189–193.
- Gabrielsson JL. Utilization of physiologically based models in extrapolating pharmacokinetic data among species. *Fundam Appl Toxicol*. 1991;16(2):230–232.
- Dixit R, Boelsterli UA. Healthy animals and animal models of human disease(s) in safety assessment of human pharmaceuticals, including therapeutic antibodies. *Drug Discov Today*. 2007;12(7–8):336–342.
- Miller K, Louie A, Baltch AL, et al. Pharmacokinetics of pentoxifylline and its metabolites in healthy mice and in mice infected with *Candida albicans*. *Antimicrob Agents Chemother*. 1998;42(9):2405–2409.
- Simmons D. The use of animal models in studying genetic disease: Transgenesis and induced mutation. *Nature Educ*. 2008;1(1).
- Hart BA, Bajramovic JJ. Non-human primate models of multiple sclerosis. *Drug Discovery Today: Disease Models*. 2008;5(2):97–104.
- Coppola BA, Omens JH. Use of larger species such as dog and pig as model systems to study cardiac disease. *Drug Discovery Today: Disease Models*. 2008;5(3):195–200.
- Wang L, Martin DR, Baker HJ, et al. Neural progenitor cell transplantation and imaging in a large animal model. *Neurosci Res*. 2007;59(3):327–340.
- Scheerlinck JY, Snibson KJ, Bowles VM, et al. Biomedical applications of sheep models: from asthma to vaccines. *Trends Biotechnol*. 2008;26(5):259–266.
- Gabrielsson J, Weiner D. *Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications*. 4th Edition. Stockholm, Sweden: Swedish Pharmaceutical Press; 2007.
- Anderson BJ, McKee D, Holford NHG. Size, myths and the clinical pharmacokinetics of analgesia in paediatric patients. *Clin Pharmacokinet*. 1997;33:313–327.
- Holford NHG. A size standard for pharmacokinetics. *Clin Pharmacokinet*. 1996;30:329–332.
- West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science*. 1997;276:122–126.
- The University of Auckland School of Medical Sciences. 2009. MEDSCI 722 workshop timetable. Available from: <http://www.fmhs.auckland.ac.nz/sms/pharmacology/holford/teaching/medsci722/timetable.aspx>. Accessed on Dec 10, 2009.
- Bourne D. Biopharmaceutics and pharmacokinetics. 2009. Available from: <http://www.boomer.org/c/p4/>. Accessed on January 10, 2010.
- Beal SL. On the Solution to the Michaelis-Menten Equation. *J Pharmacokinet Biopharm*. 1982;10(1):109–119.
- Boeckmann AJ, Sheiner LB, Beal SL. *NONMEM Users Guide – Part V Introductory Guide*. San Francisco, CA: NONMEM Project Group, University of California; 1994.
- Bonate P. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. 1st edition. Chicago, IL: Springer; 2005.
- Luo FR, Yang Z, Camuso A, et al. Dasatinib (BMS-354825) pharmacokinetics and pharmacodynamic biomarkers in animal models predict optimal clinical exposure. *Clin Cancer Res*. 2006;12(23):7180–7186.
- Bruno R, Washington CB, Lu JF, et al. Population pharmacokinetics of trastuzumab in patients with HER2+ metastatic breast cancer. *Cancer Chemother Pharmacol*. 2005;56(4):361–369.
- Marchetti S, Schellens JH. The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials. *Br J Cancer*. 2007;97(5):577–581.
- Whitehead J, Zhou Y, Hampson L, et al. A Bayesian approach for dose-escalation in a Phase I clinical trial incorporating pharmacodynamic endpoints. *J Biopharm Stat*. 2007;17(6):1117–1129.
- Wang K, Ivanova A. Two-dimensional dose finding in discrete dose space. *Biometrics*. 2005;61(1):217–222.
- Normolle D, Lawrence T. Designing dose-escalation trials with late-onset toxicities using the time-to-event continual reassessment method. *J Clin Oncol*. 2006;24(27):4426–4433.
- Ivanova A, Wang K. A non-parametric approach to the design and analysis of two-dimensional dose-finding trials. *Stat Med*. 2004;23(12):1861–1870.
- Splawinski J, Kuñiar J. Clinical trials: active control vs placebo – what is ethical? *Sci Eng Ethics*. 2004;10(1):73–79.
- Terézhalmy GT, Bartizek RD, Biesbrock AR. Plaque-removal efficacy of four types of dental floss. *J Periodontol*. 2008;79(2):245–251.
- Singh D, Brooks J, Hagan G, et al. Superiority of “triple” therapy with salmeterol/fluticasone propionate and tiotropium bromide versus individual components in moderate to severe COPD. *Thorax*. 2008;63(7):592–598.
- Hirsch FR, Hansen HH, Hansen M, et al. The superiority of combination chemotherapy including etoposide based on in vivo cell cycle analysis in the treatment of extensive small-cell lung cancer: a randomized trial of 288 consecutive patients. *J Clin Oncol*. 1987;5(4):585–591.
- Snapinn S. Noninferiority trials. *Curr Control Trials Cardiovasc Med*. 2000;1(1):19–21.
- Kay R. *Statistical Thinking for Non-Statisticians in Drug Regulation*. Philadelphia, PA: Wiley-Interscience; 2007.
- Kieser M, Friede T. Planning and analysis of three-arm non-inferiority trials with binary endpoints. *Stat Med*. 2006;26(2):253–273.
- Munk A, Mielke M, Skipka G, et al. Testing noninferiority in three-armed clinical trials based on likelihood ratio statistics. *Can J Stat*. 2007;35(3):413–431.
- Blackwelder WC. Current issues in clinical equivalence trials. *J Dent Res*. 2004;83 Spec No C:C113–C115.
- Wiens BL, Iglewicz B. Design and analysis of three treatment equivalence trials. *Control Clin Trials*. 2000;21(2):127–137.
- Beller EM, Gebski V, Keech AC. Randomisation in clinical trials. *Med J Aust*. 2002;177(10):565–567.
- Torgerson DJ, Campbell MK. Use of unequal randomisation to aid the economic efficiency of clinical trials. *BMJ*. 2000;321(7263):759.
- Dumville JC, Hahn S, Miles JN, et al. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials*. 2006;27(1):1–12.
- Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Train*. 2008;43(2):215–221.

49. McEntegart DJ. Pursuit of balance using stratified and dynamic randomization techniques: an overview. *Drug Inf J*. 2003;37(3): 293–308.
50. Signorini DF, Leung O, Simes RJ, et al. Dynamic balanced randomization for clinical trials. *Stat Med*. 1993;12(24):2343–2350.
51. GebSKI VI, Beller EM, Keech AC. Randomised controlled trials: elements of a good study. *Med J Aust*. 2001;175(5):272–274.
52. Sica GT. Bias in research studies. *Radiology*. 2006;238(3): 780–789.
53. Chow S, Liu J. *Design and Analysis of Clinical Trials: Concepts and methodologies*. 2nd Edition. Philadelphia, PA: John Wiley and Sons, Inc.; 2004.
54. Gluud LL. Bias in clinical intervention research. *Am J Epidemiol*. 2006;163(6):493–501.
55. Thall PF. A review of phase 2–3 clinical trial designs. *Lifetime Data Anal*. 2008;14(1):37–53.
56. Appel LJ. A primer on the design, conduct, and interpretation of clinical trials. *Clin J Am Soc Nephrol*. 2006;1(6):1360–1367.
57. Fitzpatrick S. *Clinical Trial Design*. : The Institute of Clinical Research; 2006.
58. Green S, Liu PY, O’Sullivan J. Factorial design considerations. *J Clin Oncol*. 2002;20(16):3424–3430.
59. Lathyris DN, Trikalinos TA, Ioannidis JP. Evidence from crossover trials: empirical evaluation and comparison against parallel arm trials. *Int J Epidemiol*. 2007;36(2):422–430.
60. Quan H, Luo X, Capizzi T. Multiplicity adjustment for multiple endpoints in clinical trials with multiple doses of an active treatment. *Stat Med*. 2005;24(14):2151–2170.
61. Rosner B. *Fundamental of Biostatistics*. 5th Edition. Stamford, CT: Duxbury Thomson Learning; 2000.
62. Streiner DL. Missing data and the trouble with LOCF. *Evid Based Ment Health*. 2008;11(1):3–5.
63. Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry*. 2002;47(1): 68–75.
64. Taylor L, Zhou XH. Multiple imputation methods for treatment non-compliance and nonresponse in randomized clinical trials. *Biometrics*. 2009;65(1):88–95.
65. Yang X, Li J, Shoptaw S. Imputation-based strategies for clinical trial longitudinal data with nonignorable missing values. *Stat Med*. 2008;27(15):2826–2849.
66. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res*. 2007;16(3):199–218.
67. Shaffer ML, Chinchilli VM. Including multiple imputation in a sensitivity analysis for clinical trials with treatment failures. *Contemp Clin Trials*. 2007;28(2):130–137.
68. Baselga J, Carbonell X, Castañeda-Soto NJ, et al. Phase II study of efficacy, safety, and pharmacokinetics of trastuzumab monotherapy administered on a 3-weekly schedule. *J Clin Oncol*. 2005;23(10):2162–2171.
69. Fleming TR. Data monitoring committees and capturing relevant information of high quality. *Stat Med*. 1993;12(5–6):565–570; discussion 571–573.
70. Grant AM, Altman DG, Babiker AB, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess*. 2005;9(7): 1–238, iii–iv.
71. O’Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–556.
72. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*. 1983;70:659–663.
73. DeMets DL, Lan G. The alpha spending function approach to interim data analyses. *Cancer Treat Res*. 1995;75:1–27.
74. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13(13–14):1341–1352; discussion 1353–1356.
75. Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials*. 1989; 10(4 Suppl):209S–221S.
76. Hughes MD, Pocock SJ. Stopping rules and estimation problems in clinical trials. *Stat Med*. 1988;7(12):1231–1242.
77. Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*. 1987;43(1):213–223.
78. Jennison C, Turnbull BW. *Group Sequential Trials: Applications to Clinical Trials*. Baton Rouge, LA: Chapman and Hall/CRC; 2000.
79. Bretz F, Koenig F, Brannath W, et al. Adaptive designs for confirmatory clinical trials. *Stat Med*. 2009;28(10):1445–1463.
80. Walker GA. *Common Statistical Methods for Clinical Research with SAS Examples*. 2nd edition. Cary, NC: SAS Institute Inc.; 2006.
81. Berry DA. Bayesian clinical trials. *Nature Rev Drug Discov*. 2006;5: 27–36.
82. Biswas S, Liu DD, Lee JJ, Berry DA. Bayesian clinical trials at the University of Texas M.D. Anderson Cancer Center. *Clin Trials*. 2009;6(3):205–216.
83. Schoenwald RD. *Pharmacokinetic Principles of Dosing Adjustments: Understanding the Basics*. Baton Rouge, LA: CRC Press LLC.; 2001.
84. Puri PK, Tyrer PJ. *Sciences Basic to Psychiatry*. Philadelphia, PA: Elsevier Health Sciences; 1998.

## Appendix I

### NONMEM example program for analyzing PK data involving drug absorption (linear single compartment model)

This example is written by the authors using the example in Chapter 2 of the NONMEM Users Guide Part 5<sup>26</sup> to solve another similar simple absorption problem given on pages 115 and 116 of the book by Bonate.<sup>27</sup> The problem solved there using SAS has been verified here using the NONMEM code given below. The program outputs estimates for the absorption constant (KA), elimination constant (K), the volume of distribution (= V/1000).

```

$PROBLEM SINGLE COMPARTMENT ABSORPTION
MODEL
$INPUT AMT TIME DV
$DATA DATAFOR SICOABMO
$SUBROUTINE ADVAN2
$PK
KA = THETA(1)
K = THETA(2)
V = THETA(3)
; THE FOLLOWING ARE REQUIRED BY PREDPP
S2 = V/1000
$ERROR
Y = F+ERR(1)
$THETA (0,0.5) (0,0.01) (0,100)
$OMEGA 1225
$ESTIMATION PRINT = 5
$COVARIANCE
$TABLE AMT TIME
$COVARIANCE
$SCATTER PRED VS DV UNIT

```

An example of a SAS program to analyze population pharmacokinetic data is given in [http://support.sas.com/documentation/cdl/en/statug/59654/HTML/default/statug\\_nlmixed\\_sect039.htm](http://support.sas.com/documentation/cdl/en/statug/59654/HTML/default/statug_nlmixed_sect039.htm)

### Sample SAS codes for sample size calculations

A SAS program to calculate sample size (ntotal) assuming two independent treatment arms (randomization ratio 1:2) and a continuous outcome follows:

**proc power;**

```

twosamplemeans alpha = 0.05
groupmeans = (2, 4.8)

```

```

test = diff_satt
sides = 2
power = 0.90
ntotal = .
groupweights = (1,2)
groupstddevs = (0.1 0.15);

```

**run;**

A SAS program to calculate sample size (ntotal) assuming two independent treatment arms (randomization ratio 1:2) and a dichotomous outcome follows:

**proc power;**

```

twosamplefreq alpha = 0.05
groupproportions = (0.3, 0.45)
test = pchi
sides = 2
power = 0.90
ntotal = .
groupweights = (1,2);

```

**run;**

## Common statistical tests and methods for the analysis of clinical trial data

Data from a clinical trial can be continuous such as laboratory results and vital signs measurements or categorical such as demographics (sex, race, education) or a physical examination result (normal or abnormal). Categorical data can be further classified as nominal, where the categories are not ordered (such as sex, race), and ordinal, where they are ordered (such as educational level).

In general, continuous data are summarized in terms of mean, standard deviation, median, minimum and maximum, while categorical data are summarized in terms of the number of events and percentages. The following are the common statistical tests and methods used to analyze continuous and categorical outcome variables.<sup>80</sup>

### Statistical tests

#### Continuous data

##### One-sample *t*-test or paired *t*-test

Difference in the means of the continuous outcome variable between two arms which are not independent, such as the average change in blood pressure from baseline on treatment with a single hypertension drug, where the baseline and post-treatment measurements are performed for each subject. The null hypothesis is that the drug does not produce an average change in blood pressure that is statistically significantly different from zero.

### Two-sample *t*-test

Difference in the means of the continuous outcome variable between two arms which are independent, such as the average change in blood pressure from baseline on treatment with two different hypertension drugs. The null hypothesis is that there is no statistically significant difference in the average change in blood pressure on treatment, between the two independent treatment arms.

### Analysis of variance (ANOVA)

Extension of the two-sample *t*-test to more than two arms, and is used in clinical trials data analyses when there are three treatment arms. Adjusting the significance level to account for multiple comparisons of means can be done using procedures such as Tukey, Scheffé, and Bonferroni.

### Linear regression

The relation between a number of independent continuous variables and one or more dependent continuous variables is modeled, assuming that the outcome variables depend linearly on the independent variables.

### Analysis of covariance (ANCOVA)

A combination of ANOVA and linear regression. Using ANCOVA, the difference in the continuous outcome variable among arms can be studied, adjusting for categorical or continuous covariates such as study center, gender and age. It can help determine the relative importance of various covariates, and if there is interaction (covariate which is an effect modifier) and or confounding (association between covariates, which can occur, for example, in an unbalanced [unstratified] trial). Similar issues can also be investigated using ANOVA but where all covariates are categorical.

### Wilcoxon signed-rank test

Nonparametric analog of the parametric one-sample *t*-test (Nonparametric statistics is used when the data do not follow a Gaussian distribution).

### Wilcoxon rank-sum test

Nonparametric analog of the parametric two-sample *t*-test.

### Kruskal–Wallis test

Nonparametric analog of the parametric one-way ANOVA.

### Categorical data

#### Chi-squared test

Used to analyze a (say 2\*2) contingency table where typically the rows are the two treatment arms and the two columns are

the responders and nonresponders to the treatment, and the results of the test determine whether there is an association between the treatment arms and response. The Chi-squared test is also used to analyze  $g \times r$  tables (where  $g$  is the number of arms and  $r$  is the number of responses), but the interpretation of results from such tables can be more complex than that from a 2\*2 table.

#### Fisher's exact test

Used for the analysis of contingency tables and is very similar to the Chi-squared test. However, as the name implies, the results of the Fisher's exact test are exact and valid for any number of subjects, ie, even if the cell counts of the table are small, while the Chi-squared test is valid only for large cell counts. Both the Chi-squared test and the Fisher's exact test are nonparametric tests.

#### Cochran–Mantel–Haenszel test

Used to adjust for categorical covariates while analyzing a (say 2\*2) contingency table. Hence, the test helps determine, for example, whether there is an association between treatment arms and response, after adjusting for a categorical covariate(s). It can be a parametric or nonparametric test depending on the number of covariates adjusted for and whether nonparametric scores are specified.

#### McNemar's test

Used to analyze matched data from the same sample of subjects and determines whether there is a significant difference between the two correlated proportions in a 2\*2 contingency table. In clinical research, it is used to determine what the effect of a drug is on a certain disease by considering baseline and postbaseline measurements of the same subjects. It is a nonparametric test.

## Statistical methods

### Logistic regression

This method is very similar to ANCOVA but is used when the outcome variable is a binary variable or an ordinal variable; the independent variables can be continuous or binary. When both the outcome variable and the independent variable (assume only one covariate) are binary, the results from a logistic regression model are the same as those from a 2\*2 contingency table with a Chi-squared test. Similarly, the results of a Cochran–Mantel–Haenszel test and logistic regression model will be the same for a simple case, such as adjusting a 2\*2 contingency table for one binary covariate. However, when there are multiple covariates which can be

continuous or categorical variables, logistic regression is a powerful tool to analyze the data. Logistic regression can also be used in clinical studies to determine if there is a linear trend in the drug dosage versus observed response (example no improvement, some improvement, complete recovery).

### Survival analysis (Log-rank test, generalized Wilcoxon test and Cox proportional hazards model)

Survival analysis is used to analyze the difference in survival times among arms when some observations are censored. Censoring occurs when an outcome is not observed for reasons ranging from loss of subjects (to follow up) to no observed event until the time the study ends. The nonparametric Log-rank test is used to compare time to event, such as death or recurrence of disease, among treatment arms. The Log-rank test assigns all time points the same weight, whereas the generalized Wilcoxon test (another nonparametric test) assigns earlier time points a higher weight. This test has a greater type II error or a large number of censored subjects. In both the Log-rank and generalized Wilcoxon tests, the analysis can be stratified by factors such as gender.

The semi-parametric Cox proportional hazards model also compares survival times among treatment arms, while adjusting for censoring and any covariates, including a specialized procedure in which time-dependent covariates are included in the model. An underlying assumption of the proportional hazards model is that the hazard (risk) ratio between the (say two) treatment arms remains constant over time.

The Log-rank test (using `proc lifetest` with `strata` and `group` options in SAS) and the Cox proportional hazards model (using `phreg` in SAS) yield similar results when stratified analyses are performed to compare treatments while controlling for the effect of other variables (say gender).

### Longitudinal data analysis

Longitudinal data analysis or repeated measures analysis is often performed when there is correlated data for each subject, such as the size measurement of each subject's tumor at regular intervals. In such a case, while analyzing the difference among treatment arms, the correlated nature of each subject's data is incorporated (such as in a parametric repeated measures ANOVA). Failure to account for the correlation results in an incorrect estimation of standard errors and hence incorrect results of the hypothesis testing.

### Meta analysis

Meta analysis is a useful technique, employed to pool and analyze data from related but independent, separate

clinical trials (investigating the same outcome) and obtain quantitative combined efficacy and or safety results. The pooled meta analysis results are only as good as the quality of the studies included. Also to increase the validity of the results, the issues of heterogeneity, publication bias, confounding need to be examined.

### Bayesian methods in clinical trials

Bayesian methods used in clinical trials often focus on making an educated guess for a prior distribution of the outcome of interest, and using it along with the accumulated trial data to estimate a posterior distribution for the outcome. Thus, Bayesian methods are often ideal in adaptive design trials, which use accrued trial information.<sup>81,82</sup> However it should be noted that the reliability of the posterior and predictive distributions depends heavily on constructing a meaningful and agreed upon prior distribution.

## Appendix 2 glossary

### 3 + 3 design for dose escalation

Three subjects are first enrolled at a particular dose level. a) If zero out of the three subjects experience a dose limiting toxicity (DLT) (usually grade 3/4 adverse event), then the subjects are administered the next higher dose level or stay at the same dose, if it is the highest dose level. b) If one of the three subjects enrolled experiences a DLT, three additional subjects are enrolled at the same dose level. If none of the three additional subjects experience a DLT, then the subjects are administered the next higher dose level or otherwise the dose is declared as an MTD. If one or more of the three additional subjects enrolled experience a DLT, all subjects are administered the next lower dose – the next lower dose is declared as MTD if six subjects have already been enrolled (it is not permitted to re-escalate). c) If two or more of the subjects enrolled experience a DLT, the subjects are administered the next lower dose – the next lower dose is declared as MTD if six subjects have already been enrolled (it is not permitted to re-escalate).

### Absorption

When the drug is administered by an enteral route (drug given via the digestive tract) such as oral dosing, it needs to be absorbed by the body before it is eliminated. Factors that affect oral absorption, for example, include size of particles in the tablet, drug solubility, other medications being taken simultaneously, and food being taken.

### Allometric scaling

Allometric scaling can be used to extrapolate PK results from animal studies to humans. The volume of distribution and

clearance for individuals with different body size are believed to follow the allometric scaling laws given below:<sup>20-22</sup>

$$V_d = V_{std} * (W_i/W_{std}),$$

$$CL = CL_{std} * (W_i/W_{std})^{0.75},$$

where  $W_i$  is the weight of the individual and  $W_{std}$  is the weight of a reference individual (for example,  $W_{std}$  can be ~70 kg) and  $V_{std}$  and  $CL_{std}$  are the population volume of distribution and clearance standardized to the reference individual with weight  $W_{std}$ .

The allometric scaling laws work for extrapolation of results among various species (ie, of mammals) in many cases because the average number of heartbeats and breaths per lifetime are roughly the same.<sup>19</sup> This implies that a long-living species such as man has a lower metabolism rate than most other species.

## Area under the concentration (versus time) curve (AUC)

The area under the drug concentration versus time curve is calculated by integrating the concentration from time zero to infinity. An equation that is valid for oral dosing is  $Dose_1/Dose_2 = AUC_1/AUC_2$ . If the AUC values of two different products of the same drug and of the same dose are compared in the case of oral dosing, it indicates if the two products are equivalent.<sup>83</sup>

## AUMC

The AUMC is the area under the concentration multiplied by time versus time curve.

## Bioavailability

Bioavailability, a factor whose value lies between zero and one, quantifies how much of the dose administered orally or by any other enteral route can be absorbed across the gut wall and is subsequently released into the blood stream. The bioavailability of a drug administered by IV or any other parenteral route (ie, administered by means other than through the digestive tract) is one, since the drug enters the blood stream directly.

## Biomarker

Sometimes a biomarker or surrogate marker is used as or instead of an endpoint, when the actual endpoint may be difficult, time consuming or impractical to measure. A surrogate marker is usually directly correlated with the endpoint or clinical outcome and is an indicator of the drug effect on the condition. The CD4 cell count as a measure of the treatment effect for AIDS or the C-reactive protein (CRP) as a marker for inflammation associated with cardiovascular disease are common examples of biomarkers.

## Bolus dose

A bolus dose is a single concentrated or large dose of a drug; it can be administered intravenously or orally.

## Clearance

Clearance is a measure of how quickly the body is able to eliminate the drug, and its units are liters/hour. Drug clearance from the body occurs by the combined activities of the liver and kidneys. Clearance of the drug by the kidneys is completed through urine excretion. The liver metabolizes the drug, which is then excreted by the kidneys. One way to estimate clearance is by obtaining the ratio of the total drug dosage administered to the AUC value.

## Distribution

The drug that enters the blood stream is distributed to various parts of the body such as the heart, brain, lungs, liver, kidneys, bones and different tissues. The distribution is governed by factors such as the rate of blood flow to various organs, solubility of the drug in water and fat, binding of the drug to proteins, and pH of the drug and that of the region(s) to which it is transported.

## Excretion

The elimination of the drug from the body is generally due to metabolism and excretion. Most drug excretion occurs via urine that is produced by the functioning of the kidneys. Excretion can also occur via bile and to a much smaller extent via sweat, tears, and exhalation. Some factors that affect excretion are drug solubility in water, binding of the drug to protein, blood flow to the kidneys, age, and renal disease.

## Loading dose

A loading dose is an initial higher dose given to a subject at the beginning of the treatment.

$$\text{loading dose} = \text{target concentration} * \text{volume of distribution/bioavailability.}$$

## Maintenance dose

A maintenance dose is a lower dose given to a subject after the initial loading dose to maintain the therapeutic level of the drug in the body.

$$\text{maintenance dose rate} = \text{target concentration} * \text{clearance}$$

## Metabolism

Drug metabolism occurs mainly due to the action of the liver. It involves the conversion of the drug to another (more water soluble) substance that can be excreted by the kidneys or

bile. Metabolism occurs in the case of most drugs, with few drugs being eliminated directly. It generally takes place in two stages called phase I (which typically involves oxidation of the drug but can also involve hydrolysis or reduction of the drug) and phase II (where an ionized group added to the drug makes it more water soluble). Age, hepatic and other diseases, food consumed, and smoking are factors that affect drug metabolism.

## Volume of distribution

The volume of distribution is a measure of how the drug is distributed in the body, and is an apparent or effective

volume that does not necessarily equal the actual anatomical or physiological volume due to various reasons such as the binding of the drug to protein, the binding of the drug to tissue, and the hydrophilic properties of the drug.<sup>84</sup> It is the ratio of the drug dose administered to the measured initial plasma concentration of drug.

### Open Access Journal of Clinical Trials

#### Publish your work in this journal

The Open Access Journal of Clinical Trials is an international, peer-reviewed, open access journal publishing original research, reports, editorials, reviews and commentaries on all aspects of clinical trial design, management, legal, ethical and regulatory issues, case record form design, data collection, quality assurance and data auditing

Submit your manuscript here: <http://www.dovepress.com/open-access-journal-of-clinical-trials-journal>

methodologies. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress