Pharmacokinetic-Pharmacodynamic Modeling and Simulation

Pharmacokinetic-Pharmacodynamic Modeling and Simulation

Peter L. Bonate, PhD, FCP

Genzyme Corporation
San Antonio, TX USA

Springer

Peter Bonate
Genzyme Oncology
San Antonio, TX 78229
USA
peter.bonate@genzyme.com

Library of Congress Control Number: 2005928491

ISBN-10: 0-387-27197-X e-ISBN 0-387-27199-6
ISBN-13: 978-0387-27197-2

Printed on acid-free paper.

Printed in the United States of America. (SPI/MVY)

9 8 7 6 5 4 3 2 1

springeronline.com

TABLE OF CONTENTS

PREFACE

This book is written for the pharmacokineticist who performs pharmacokinetic-pharmacodynamic modeling and is occasionally asked to model data that may have nothing to do with pharmacokinetics, but may be important in other areas of drug development. The emphasis of this book is on modeling in drug development since that is my own area of expertise and because ultimately all pharmacokinetic-pharmacodynamic modeling is applied to the therapeutic use of drugs in clinical practice. Throughout this book, pharmacokinetic and pharmacodynamic models will be used without derivation and little in the way of explanation. It is expected the reader has basic knowledge of pharmacokinetics and simple pharmacodynamic models. If not, the reader is referred to Gibaldi and Perrier (1982), Wagner (1993), or Shargel and Yu (1999) for background material. The reader is also expected to have had a 1-year introductory course in statistics that covers basics of probability, regression, and analysis of variance. A 1-semester course in matrix algebra is desired but not needed.

The material in this text begins with a broad overview of modeling, which I call 'The Art of Modeling'. This chapter is meant to introduce some of the broad topics associated with modeling, such as model selection criterion, model validation, the importance of good communication, and ethics. The next chapter is linear regression, which is the foundation for most parametric modeling. From there nonlinear regression is covered, followed by variance models, weighting, and transformations. Lastly, case studies in linear and nonlinear models are presented to illustrate the theory that was presented in the previous chapters. In the material presented to this point, a key assumption is that each subject contributes a single observation to the data set. Next, the book moves to mixed effects models, which allow for multiple observations to be measured on the same individual. The next chapter is linear mixed effects models, which is meant as a brief introduction to the topic. Next is the theory of nonlinear mixed effects models, which form the foundation for population pharmacokinetic-pharmacodynamic modeling. This is followed by a chapter on practical issues in nonlinear mixed effects modeling, such as how weight, genetic, or racial information is incorporated into a model. The last chapter in this section presents some case studies on population pharmacokinetic-pharmacodynamic modeling. A key concept in this book is the inter-relatedness between the material. For example, nonlinear mixed effects models are simply extensions of linear mixed effects models, which are themselves extensions of linear models, etc. Thus, in order to understand the more complex chapters, it is necessary to understand the foundation material, e.g., what is a variance model and how are they used, how can a linear covariate model be built into a nonlinear mixed effects model, etc.

I wrote this book to be as reader-friendly as possible. Those parts of the book that are non-technical are written in an almost conversational tone with anecdotes and interesting quotes interspersed throughout. I love quotations and each chapter begins with one I thought especially poignant about the forthcoming material in the chapter. When mathematics are needed, I tried to

make those sections self-contained. Variables are defined in each chapter so the reader does not have to search for "now what is G again?"

John of Salisbury (1115–1180), a twelfth century English philosopher and historian, once wrote:

> *We are like dwarves sitting on the shoulders of giants. We see more, and things more distant than they did, not because our sight is superior or because we are taller than they, but because they raise us up, and by their great stature add to ours.*

I would like to thank the many giants that helped me understand things I was unclear about during the writing of this text and the reviewers that took the time to read the chapters and offer their opinions on how each could be improved. Without your help I would have been lost in many places. I would like to ask that if you do spot any mistakes or typographical errors to please contact me at peter.bonate@gmail.com.

I would also like to thank my wife, Diana, for her encouragement and my children, Robyn and Ryan, for reminding me that there is indeed more to life than writing "Daddy's Big Book of Science", which is what they called this while I was writing it.

Peter L. Bonate
Genzyme Corporation
San Antonio, Texas
June 2005

REVIEWERS

In a book such as this, it is impossible for one person to be an expert on everything. I could not have done this endeavor without the help of others. I would like to thank the following individuals who took the time to review the chapters.

Leon Aarons, PhD
Senior Lecture
School of Pharmacy and Pharmaceutical Sciences
University of Manchester
Manchester, UK

Jeff Barrett, PhD, FCP
Clinical Pharmacology & Therapeutics
Children's Hospital of Philadelphia
Philadelphia, PA USA

Seth Berry, PharmD
Associate Scientist
Quintiles Transnational Corp.
Kansas City, MO USA

Marie Davidian, PhD
Professor
Department of Statistics
North Carolina State University
Raleigh, NC USA

Johan Gabrielsson, PhD
Senior Principle Scientist
AstraZeneca
Sodertalje, Sweden

Danny R. Howard, PhD
Global Head, Pharmacokinetics
Novartis Pharmaceutical Corp.
East Hanover, NJ USA

Howard Lee, PhD, MD
Assistant Professor
Assoc. Director for the Clinical Investigation
Core Center for Clinical Pharmacology
Department of Medicine
University of Pittsburgh
Pittsburgh, PA USA

Bernd Meibohm, PhD, FCP
Associate Professor of Pharmaceutical Sciences
College of Pharmacy
University of Tennessee
Memphis, TN USA

Diane Mould, PhD
Projections Research, Inc.
Phoenixville, PA USA

David Ruppert, PhD
Professor of Engineering
School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY USA

Dan Weiner, PhD
Sr. Vice President
Business Development
Pharsight Corporation
Research Triangle Park, NC USA

Paolo Vicini, PhD
Associate Professor
Bioengineering
University of Washington
Seattle, WA USA

Jianjun Alan Xiao, PhD
Research Fellow
Merck & Co.
West Point, PA USA

Chapter 1

The Art of Modeling

Drawn by my eager wish, desirous of seeing the great confusion of the various strange forms created by ingenious nature, I wandered for some time among the shadowed cliffs, and came to the entrance of a great cavern. I remained before it for a while, stupefied, and ignorant of the existence of such a thing, with my back bent and my left hand resting on my knee, and shading my eyes with my right, with lids lowered and closed, and often bending this way and that to see whether I could discern anything within; but that was denied me by the great darkness inside. And after I stayed a while, suddenly there arose in me two things, fear and desire—fear because of the menacing dark cave, and desire to see whether there were any miraculous things within.

—Leonardo da Vinci (1452–1519), Renaissance scientist and philosopher

INTRODUCTION

The focus of this book is primarily on the development of pharmacokinetic and pharmacokinetic-pharmacodynamic models. Models that are reported in the literature are not picked out of thin air. Useful models take time and effort and what is rarely shown is the process that went into developing that model. The purpose of this chapter is to discuss model development, to explain the process, and to introduce concepts that will be used throughout this book. Those criteria used to select a model extend to whether the model is a linear model or a nonlinear mixed effects model and that is why this material is provided first. If the reader can understand what makes a good or validated model, then the particular type of model is irrelevant.

WHAT IS A MODEL AND WHY ARE THEY MADE?

A system is a collection of objects that interact to create a unified whole, such as a cell culture system, a rat, or a human. The type of models that are of interest in this book are mathematical models that represent the system of interest and "*can be used to explore the structure and behavior of the system*" (Wastney et al., 1997). A more simplistic definition might be that a mathematical model defines how you think your data were generated. Most famous mathematical models can be found in chemistry and physics, such as:

- Boyle's law, PV = constant, which states that for a given mass at fixed temperature the pressure (P) times the volume (V) of a gas is a constant;
- Newton's second law of motion, F = ma, which states that the force (F) acting on an object is equal to its mass (m) times its acceleration (a); and
- $E = mc^2$, perhaps the most famous equation of the last century, which most people believe has to do with Einstein's theory of relativity, but in actuality has nothing to do with it. This equation is founded on the basis that matter and energy are really different forms of the same thing and states that that the amount of energy (E) that could be produced is equal to the mass (m) of an atom times the speed of light (c) squared.

Mathematical models in biology tend to be more complex, but are all based on the same foundations used to develop models in the more physically oriented sciences.

In defining a mathematical model it is helpful to distinguish between the various components of the model. Models are built using experimentally derived data. This so-called data generating process is dependent on system inputs, system dynamics, and the device used to measure the output from a system (Fig. 1.1). But in addition to these systematic processes are the sources of error that confound our measurements. These errors may be measurement errors but also include process noise that is part of the system. One goal of mathematical modeling is to differentiate the "information" or systematic component in the system from the noise or random components in the system, i.e.,

$$\text{DATA} = \text{SYSTEMATIC COMPONENT} + \text{ERROR}.$$

Hence, models usually consist of a structural model or systematic component plus a statistical model that describes the error component of the model. Early in the modeling process the focus may lie with the systematic component and then move to a more holistic approach involving the error components. For example, the 1-compartment model after bolus administration is

$$C = \frac{D}{V}\exp\left(-\frac{CL}{V}t\right)+\varepsilon. \tag{1.1}$$

The first term on the right hand side of Eq. (1.1) is the structural model having two inputs (also called independent variables), D (dose) and t (time), and one output (also called the dependent variable), C (concentration). The variables V (volume of distribution) and CL (clearance) are referred to as model parameters which must be estimated from the observed concentration data. The second term in Eq. (1.1) is the error component (also called the variance model). ε represents the deviation between model predicted concentrations and observed concentrations.

Modeling is done for a number of reasons depending on the point of view. Scientifically, modeling "provides a systematic way of organizing data and observations of a system at the cell, tissue, organ, or whole animal (human) levels" and "affords the opportunity to better understand and predict physiological phenomena" (Epstein, 1994). Financially, companies utilize modeling as a way to better leverage business decisions and this has been shown to result in substantial cost savings over traditional experiments (Van Buskirk, 2000). And on a personal level, modelers model because it's fun and challenging.

Beyond characterizing data, once a model is developed, it can be used to answer "what if" questions—a process known as simulation. Hence, modeling and simulation (M&S) are often used in the same breath by modelers. But there are many important differences between modeling and simulation. A model looks back in time. Given a set of outputs (data), the model attempts to find a set of parameters that explain the data generating process. Simulation looks forward in time. Given a model and a set of parameters, what happens if the inputs are varied. In simulation, the model is fixed and the inputs are varied. In modeling, the inputs and outputs are fixed, but what happens in between is varied. More about the differences between M&S will become evident using examples throughout the book.

The implementation of mathematics into biology, physiology, pharmacology, and medicine is not new, but its use has grown in the last three decades as computer speeds have increased and scientists have begun to see the power of modeling to answer scientific questions. A conference was held in 1989 at the National Institutes of Health called "Modeling in Biomedical Research: An Assessment of Current and Potential Approaches." One conclusion from that conference was that "*biomedical research will be most effectively advanced by the continued application of a combination of models—mathematical, computer, physical, cell, tissue culture and animal—in a complementary and interactive manner*."

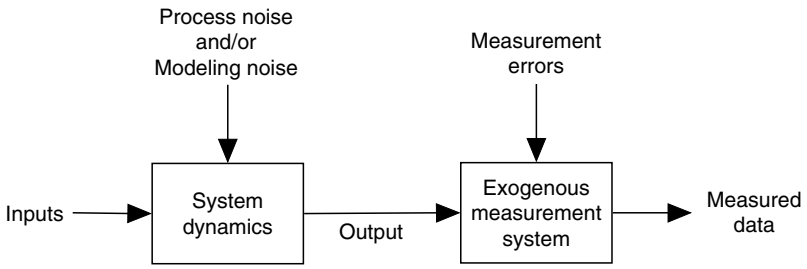

Figure 1.1 Diagram of the system under study. Redrawn with from DiStefano and Landaw (1984). Reprinted with permission from The American Physiological Society, Copyright 1984.

Today, the interplay between these different types of models has never been greater. As scientists become more ''mathletic'' and as the drive to decrease the use of living animals in medical research increases, mathematical models will play an increasingly important part of medical research.

The use of modeling in drug development, for which many of the examples in this book have been culled, is also becoming increasingly important. Aarons et al. (2001) and Balant and Gex-Fabry (2000) present comprehensive reviews on the applications and use of M&S in drug development. The cost to develop a new drug in 2003 was estimated at 802 million dollars (DiMasi, Hansen, and Grabowski, 2003). Clearly, drug companies must find ways to reduce costs and expedite getting a drug to market. Two recent papers written by financial advisors suggest that M&S will play a ''vital role'' in drug development by enabling scientists to predict how drugs will act in whole systems, organs, and at the sub-cellular level, to predict clinical trial outcomes before they are actually conducted, and to adapt clinical trials on the fly as patient data is accumulated without compromising its statistical validity, thereby lowering costs and potentially speeding development (IBM Business Consulting Services, 2003; PricewaterhouseCoopers, 1999).

These reports also criticize the pharmaceutical industry for slowly adopting M&S as a research and development tool. Perhaps drug companies have failed to routinely implement M&S as part of the development process because modelers have failed to show that the methodology can indeed lower the cost of drug development, expedite development time, or result in faster and more efficient approval times by regulatory agencies. Thankfully, regulatory agencies have issued recent guidances advocating a more integral role for M&S in the development process through the establishment of exposure-response relationships thereby forcing drug companies to increase the role of M&S in the development of drugs (United States Department of Health and Human Services et al., 2003). Currently, however, M&S groups within industry tend to be fringe or splinter groups, usually within the clinical pharmacology or clinical pharmacokinetics department, that operate sporadically on drug development projects or in cases where they are called in to ''save'' failed clinical trials. But, if the financial advisors are correct, then the role of M&S in drug development will only increase over time, to the benefit of those who love to model and to the company as well.

One theme that will be stressed throughout this book is the concept that there is and can never be a true model for a biological system. Biological systems are inherently nonlinear of potentially infinite dimension with feedback loops, possibly circadian variation, and are exceedingly complex with sometimes very tight control. It is folly to think that given the limited number of subjects and number of observations collected per subject in clinical or preclinical studies that the true model could be uncovered and its parameters estimated with any degree of precision. No modeler could ever develop a model with the degree of precision that explains such a system in the presence of the many uncontrolled variables that influence the data generating process. However, it may be possible that a reasonable approximation or simplification to the true data generating model could be developed.

Because the true model can never be identified, there can never be a ''right'' model. Box (1976) stated, in one of the most famous adages in pharmacokinetics, that ''*all models are wrong, some are useful*.'' This quote is made time and again, yet it is not uncommon to hear pharmacokineticists talk about the ''right model'' or even worse ''the wrong model.'' All models are wrong—there is no right model. Granted some models are better than others, but models are really in the eye of the beholder. A modeler may choose one model over another, especially when the model is complex, because along the model development process there are many forks in the road. One modeler may choose one path, whereas another may choose another path. At the end of the process, each modeler may have a model that is different from the other modeler, each with equal credibility. So which model is the right model? Well, neither is.

There is a famous film director from Japan named Kurosawa who directed a movie called Rashomon (1950). The story itself is violent, involving rape and murder, but is told in flashbacks from the point of view of each of the participants. With each narrator the characters are essentially the same, as are most of the details, but each person's story is different. In the end, Kurosawa never reveals what truly happened. The point is that reality is relative. Each modeler views a model from a different point of view, each of which may be a valid interpretation of the data, none of which may be correct. In Rashomon, all the presenters could be telling the truth or none of them could be telling the truth. This is called the Rashomon effect—there may be a multitude of models that describe a set of data giving the same degree of predictability and error (Breiman, 2002). The tendency in our profession to use the phrase ''the right model'' needs to be changed. Should the term ''better model'' be used? ''Is there a better model?'' ''Model A is better than Model B.'' Consider the case where a model is developed. New data is collected and the original model is revised to explain the new data. Does the development of the second

model mean that the first model was wrong? Certainly not, it means the second model is better.

It is unfortunate that as a profession we choose to sabotage ourselves with a poor choice of words to professionals outside our field. We need to move beyond using the phrase "the right model" since those outside our profession may not understand the nuances between "the right model," "the best model," "a better model," or "the wrong model." In doing so, we will avoid confusion and add credibility to our results.

MODELING AS PROBLEM SOLVING

Modeling is an exercise in problem solving and there are steps that can be taken to maximize the probability of solving the problem. The following outlines the steps for effective problem solving:

1. One of Steven Covey's "*7 Habits for Highly Effective People*" (1989) is to "*begin with the end in mind*." This is a good modeling advice. Recognize and define the problem. Define your destination before you go off on a modeling adventure. What use do you want your model to serve? If you are part of a team, such as a project team in a pharmaceutical company, get team members to agree to the problem and how you have defined it before you go off to solve it.

2. Analyze the problem. What data are available to solve the problem? Given the data available can the problem be solved? Has proper attention to study design and data collection been done to achieve the objective? Question whether a model is even necessary. Perhaps a noncompartmental analysis of the data will suffice instead. If the goal is to model multiple-dose data from single dose data, then something simple like the superposition principle may be useful.

3. Identify alternative solutions. Review past solutions for current problems. Perhaps something that you are trying to do has already been done and reported in the literature for a different drug. Sometimes it is possible to break a complex problem into a series of simpler problems. Sometimes you will need to be creative though and possibly need to brainstorm with others.

4. Evaluate the possible solutions. Define criteria for choosing a solution. Is time more important than cost? What is the most cost- and time-effective alternative to answer the question? Understand who will use the results and what will be the best way to communicate those results. Perhaps modeling is not the optimal solution.

5. Decide on a solution, keeping in mind that rarely will any single solution be perfect. Identify limitations of all proposed solutions.

6. Visualize the steps needed to get there. We have all heard stories of how great athletes visualize a race or a game beforehand to provide a competitive edge over their opponents. They visualize the steps leading to winning the event, such as starting from a runner's block, and then visualize the actual winning of the event, such as crossing the finish line. Modeling is no different. For example, suppose the goal is to identify those patient characteristics, like age, that might be predictive of exposure for a new drug. It will be useful to plan the steps needed to achieve the goal, such as collecting all the relevant data, developing a pharmacokinetic model, and then using linear regression to examine the relationship between area under the curve (a measure of exposure) and patient characteristics. Having a strategy before you start will always lead you to the finish line faster than starting without a strategy.

7. Implement the strategy by building a solution incrementally. Don't try to solve the problem all at once. If no solution can be found, try reformulating the problem. Examine the assumptions and look for hidden constraints. Perhaps some of the assumptions are unnecessary or are overly complex.

8. Remember that there are other alternatives, ones that you did not examine, so try to avoid totally focusing on the one solution you implemented. Take time to reflect at the completion of a project. What hurdles occurred during the process? What would you do differently next time if the same hurdle occurs? What would you do differently in general?

Despite the best plans, roadblocks are often encountered in developing complex models. For example, a modeler may envision what the model should look like but once the model is actually fit to the data, the parameter estimates may be poorly estimated or the goodness of fit of the model is poor, in which case, another model is often needed. It is not uncommon though for the modeler to be uncertain about what that next model should be. Hence, the modeler encounters a mental roadblock.

Getting past the roadblock is what separates a good modeler from a great modeler, an inexperienced modeler from an experienced modeler. Often the solution can be drawn from past experience from models or methods seen in the literature. Sometimes, though, creativity and insight are required, another aspect that makes modeling an art. Hewlett-Packard published a short series of on-line articles on how inventors invent. Thirteen of their best inventors were interviewed and asked how they overcome roadblocks and get creative. Most replied the same thing: switch gears and do something else. Many replied that their creative obstacles were

overcome, even inspired some said, while doing something mundane, like sleeping, showering, or simply walking down the hallway. Others thought that bouncing ideas off colleagues were useful.

Modeling can be very rewarding to the modeler, as is solving any complex problem. But it is easy to "spin your wheels" and lose focus or get caught in obstacles that seem insurmountable. Just remember, to remain focused on the end result, but stay flexible and open to new ideas that may develop during the process. As McCullough and Nelder (1989) put it—don't fall in love with your models. It is easy to get "locked into" a model and build a model out of pure nonsense.

TYPE OF MODELS

Models represent the system under study. But no system is measured without error. Humans have yet to create a device that measures something with absolute certainty (recall Fig. 1.1). Rescigno and Beck (1987) call the system to be studied the primary system, and what is used to study the primary system by the investigator the secondary system. Under these definitions, a model can be considered a type of secondary system used to test properties of the primary system. To form a model, a set of inputs and outputs must be available. Inputs perturb the system in some manner. For example, if the system under study were a human, then administering a dose of drug into the subject would represent the input. The blood samples used for pharmacokinetic analysis and any pharmacodynamic endpoints that are measured would then represent the set of outputs. Both the inputs and outputs cannot be measured perfectly and are subject to error, both systematic and random in nature. It is typically assumed that the input errors are negligible. When the input errors are not this gives rise to a special class of models called error-in-variables models, which will be discussed in the chapter on Linear Models and Regression.

Models can be classified into many different categories. Using the nomenclature of DiStefano and Landaw (1984), pharmacokinetic models can generally be broken down into two types: models of data and models of systems. Models of data, usually referred to as empirical models, require few assumptions about the data generating mechanism. Examples include allometric scaling and sum of exponentials used to characterize a concentration-time profile. Empirical models are useful when little is known about the underlying physical process from which the data are generated yet one still must make some conclusions regarding the data. While empirical models may be useful at prediction they should not be extrapolated.

Model of systems, or mechanistic models, are based on physical and physiological principles and should have as many features of the system incorporated into the model as the data allow (Thakur, 1991). Factors such as transport to tissues dependent on blood flow, kinetics of receptor binding, and intracellular diffusion processes may all play a role. These models usually take the form of differential equations or partial differential equations based on mass-balance, product-precursor, or mass-action principles. Examples of mechanistic models include physiological-based pharmacokinetic models where the transport into and out of tissues is modeled as a function of blood flow and permeability between the blood and tissue. While one places greater trust in mechanistic models because they are based on theory, an analyst should always ask the question "What if the theory is wrong?" for then a mechanistic model may not be representing the system. Some models may also be hybrid models, mechanistic in places where the physiology and pharmacology of the system are understood and empirical in places that are still black boxes.

Models of systems can also be categorized into various types based on the attributes of the system, including:

- Time-variant vs. time-invariant,
- Deterministic vs. stochastic,
- Static vs. dynamic,
- Lumped vs. distributed,
- Linear vs. nonlinear, and
- Continuous vs. discrete.

Each of these categories can then be combined into more descriptive categories, e.g., a nonlinear, time-variant system or a static, time-invariant discrete system.

Time-variant means that the parameters of the system change over time, such as an autoinduction process that increases a drug's hepatic clearance with repeated administration. Time-invariant or stationary parameters do not change over time. It is typically assumed that a drug's pharmacokinetics are stationary over time so that the principle of superposition[1] applies. With a static model, the output depends only on the input and does

[1] Superposition was developed in physics to explain the behavior of waves that pass simultaneously through the same region in space. In pharmacokinetics, superposition states that concentration-time profiles passing through the same relative region in time are additive. For example, if two doses are taken 24 hours apart and the concentration 6 hours and 30 hours after the first dose was 100 and 10 ng/mL, respectively, then the concentration 6 hours after the second dose (which is 30 hours after the first dose) would be equal to 110 ng/mL (100 ng/mL + 10 ng/mL). Thron (1974) presents a comprehensive review of linearity and the meaning of superposition.

not vary over time. In the analysis of kinetic systems, static models are restricted to steady-state conditions and one example is the physiological modeling of circulatory systems. In contrast, the output of a dynamic system changes over time. In a lumped system, the various organs are ''lumped'' into single groups. The classic example is a compartmental system. In a 1-compartment model the entire body is treated as a single compartment. In a 2-compartment model the richly perfused organs are treated as the central compartment with the slowly perfused, poorly distributed organs, like fat and skin, treated as the peripheral compartment. In a distributed system, the spatial aspects of a system are built into the model. Rarely are these seen in the pharmacokinetics field since their solution typically requires partial differential equations, which few pharmacokineticists are familiar with and few software packages are equipped to solve.

All biological systems are complex, nonlinear systems. Model or function nonlinearity is defined when the derivative of a model with respect to a model parameter depends on any parameter in the model, such as when clearance follows Michaelis–Menten kinetics, or when the derivative does not exist, such as a change-point model. Estimation of parameters in a nonlinear system is more difficult than a linear system and often involves numerical optimization techniques. System nonlinearity can arise when the rate of change in a component of the system depends on the state of another component in the system, such as might arise when a component of the system shows feedback. Even though nonlinearity applies to all physiological and pharmacokinetic systems, often a useful assumption is one of linearity. For example, the dose of a drug given may result in drug concentrations much less than the Michaelis constant for metabolism, in which case the system can be approximated by a linear one. Most drugs are assumed to have linear pharmacokinetics, although some drugs, like many anti-cancer agents, demonstrate nonlinear behavior. Sometimes a nonlinear equation can be transformed to a linear one, such as the Lineweaver–Burke transformation of the Hill equation in enzyme kinetics, although this is not recommended because the transformation often distorts the distribution of the random error component of the model (Garfinkel and Fegley, 1984).

Models are also classified into whether they are deterministic or stochastic. Stochastic (Greek for ''guess'') systems involve chance or probability, whereas a deterministic system does not. In a deterministic model no randomness is assumed to be present, an assumption that is clearly not realistic. Stochastic models assume random variability and take into account that variability. There is no such thing as a deterministic model—all measurements have some error associated with them and, as such, are stochastic models by definition. However, deterministic models are useful to understand the properties of a system in the absence of natural variation. Simulations may be done using the systematic component of a model to understand the behavior of the system under different conditions.

Two types of models are usually seen in pharmacokinetics: a pharmacokinetic model, which relates dose and dosing frequency to drug concentrations, and a pharmacodynamic model, which relates drug concentrations to an effect, such as change in stomach pH, a physiologic marker, such a glucose concentration, or an outcome, such as absence or presence of an adverse event. The pharmacokinetic model predicts the concentration-time profile of the drug in the sampled biological fluid, usually plasma or serum after the administered dose. The pharmacodynamic model predicts the observed effect given the concentration provided by the pharmacokinetic model. Derendorf and Meibohm (1999) review pharmacokinetic/pharmacodynamic relationships and provide some useful classifications for these models. The pharmacokinetic and pharmacodynamic model may be either mechanistic or empirical or both.

PROPERTIES OF A USEFUL MODEL

Models are either useful or less useful. So what makes one model more useful than another? Rescigno, Beck, and Thakur (1987) state that models should be judged by three points of view: retrodiction, prediction, and understanding. Retrodiction is simply the ability to recall what happened in an experiment—does the model conform to the original data from the primary system, i.e., is the model consistent with experimental knowledge. The model must also be predictive. What will happen in future experiments? Lastly, does the model increase our understanding of the primary system under study or does the model increase our understanding of the grand primary system. For example, suppose the model is one of renal transport kinetics. Does the model increase our understanding of the physiology of the kidney? If the model does not help us decide how to answer questions about the primary system or how it fits into the world, the model may be of little value.

These properties reported by Rescigno, Beck, and Thakur (1987) should, however, be treated as a minimal set of properties for a useful model. Table 1.1 presents some other properties of a useful model. Foremost is that the model is actually used and even more importantly, is used to make a decision. Modeling for the sake of modeling, while useful for educational purposes and

Table 1.1 Properties of a useful model.

- Ability characterize the observed data and to include the most important features of the data.
- Makes accurate and precise predictions.
- Increases understanding of the system.
- The model is actually used.
- The model is completed on time.
- Logically consistent, plausible.
- Validated by empirical observations.
- Robust to small changes in the data.
- Appropriate level of precision and detail.
- As simple as possible.
- Judged on what it is intended to do.
- Has flexibility.
- Is effective as a communication tool.
- Serves many different purposes.
- May allow for extrapolation outside the data range.

possible publication, is of no practical value if the model is not used by anyone. Second, the model should be logically consistent, which means that the model has mathematical and biological plausibility. For example, are the parameters related to clearance consistent with organ blood flow? Is the pharmacodynamic model consistent with the known biology of the system? Are the parameter estimates unique, consistent, and precisely estimated? If the parameters are not precisely defined, certain aspects of the model may be overparameterized or the data set itself may be insufficient to obtain precise parameter estimates, in which case more data must be collected.

Third, is the model validated? Model validation is a contentious topic and will be discussed in more detail throughout the book and later in the chapter. Some would argue that a model is only useful if it has been validated, while others may argue that there are situations where an exploratory model is very useful. Generally, the degree of validation is dependent on the field of interest (engineering may require more validation than biomedical sciences) and application of the model. Related to validation is flexibility, which is another property of a useful model. Can more data be added to the model without having to change its structure. If so, the model is more useful than one that changes every time more data is added.

The next two properties, appropriate level of detail and as simple as possible, are two sides of the same coin because model detail increases at the expense of simplicity. Modeler's refer to this aspect of model development as Occam's razor. Formulated by William of Occam in the late Middle ages in response to increasingly complex theories being developed without an increase in predictability, Occam's razor is considered today to be one of the fundamental philosophies of modeling—the so-called principle of parsimony (Domingos, 1999). As originally stated, "*Entities should not be multiplied beyond necessity*," the theory has mutated into many other familiar forms, such as Einstein's quote "*Everything should be made as simple as possible, but not simpler*." In basic terms, Occam's razor states that the simplest model should be chosen. By choosing the simpler model, those concepts or variables not needed to explain the data are eliminated, thereby reducing the chance for redundancies, inconsistencies, or ambiguities in the model. But what exactly is the simplest model is not entirely clear. One common interpretation is if two models fit a data set equally well, choose the model with the smaller number of estimable parameters. But modeling is not always that easy. Sometimes a more complex model will be chosen because it is more consistent with theory.

Models should be judged on what they were intended to do. For example, if a model was developed in young adults and has good accuracy and prediction under different dosing regimens, should the model be deemed inadequate when it is applied to geriatric patients and does not predict with any degree of accuracy? Similarly, if a model characterizes one aspect of the data well but fails to characterize another aspect of the data, is the model still a good model? These philosophical questions should probably be answered on a case by case basis and different individuals may answer them differently. In the latter example, it may be that the system is quite complex and the modeler is really interested only in making predictions about that part of the system. In this case, the model can still be of value. In the former example, the model is still a good model, just one that it is not very generalized.

For a model to be useful, it must be developed on-time. A great solution that is arrived at too late is of no value to anyone. It is better to have a model that can provide rough answers in a useable time frame than an elegant model that is done too late. Also, models that are completed late look unfavorably upon the modeler and next time may not be assigned a project or worse, the use of a model in the future to solve a problem may not be considered as an option by the project team. This latter consequence reflects badly on modelers all over, not just on the modeler who was late, because then project teams see modeling as taking too many man-hours and being unreliable.

A useful model is one that serves as an effective communication tool. Often pharmacokineticists are asked by project teams in the pharmaceutical industry to interpret safety or efficacy data in relation to drug concentrations. Is there a relationship between the two? A quantitative approach to the problem would be to develop a model relating drug concentrations to effect (exposure-response). The model can then be presented

to the team as evidence that indeed there is a predictable, controllable relationship between dose or concentration and outcome. Unfortunately many managers and people on project teams are neither math literate nor pharmacokinetic literate so the pharmacokineticist is caught between Scylla and Charybdis. A complex model, which may be needed for complex data, may be just as difficult to communicate as presenting the data itself. Still, a complex model presented effectively by a model communicator (pun intended) can be very compelling and aid in decision-making. More will be discussed on model communication later in the chapter.

Lastly, a useful model should serve many different purposes. The model may be used to characterize data for a report or publication, may be used to better understand the system under study, may be used to make predictions for future studies, or all of the above. The more purposes a model can serve the more useful it will become.

As an example, consider the data plotted in Fig. 1.2. The observed data were simulated using the model

$$C = 8\exp(-0.3t) + 2\exp(-0.01t) + \varepsilon \qquad (1.2)$$

where ε is normally distributed random error with mean zero and variance 4. As already stated, there may be a multitude of models that describe a set of data each with equal predictability and error (Rashomon effect). Four models were fit to the data

$$C = \beta_1 + \beta_2\frac{1}{t} + \beta_3\frac{1}{t^2} \qquad (1.3)$$

$$C = \beta_1 + \beta_2\frac{1}{t} + \beta_3\frac{1}{t^2} + \beta_4\frac{1}{t^3} \qquad (1.4)$$

$$C = \beta_1 + \beta_2\frac{1}{t} + \beta_3\frac{1}{t^2} + \beta_4\frac{1}{t^3} + \beta_5\frac{1}{t^4} \qquad (1.5)$$

$$C = A\exp(-\alpha t) + B\exp(-\beta t). \qquad (1.6)$$

Equations (1.3)–(1.5) are inverse polynomials up to fourth degree and have estimable parameters β. Equation (1.6) is the model used to generate the data and has estimable parameters $\{A, \alpha, B, \beta\}$.

Which model is more useful? Table 1.2 shows the residual sum of squares (which will be discussed in greater detail later) for each model. For now, the smaller the residual sum of squares, the "better" the model. In theory, one would expect that the form of the equation

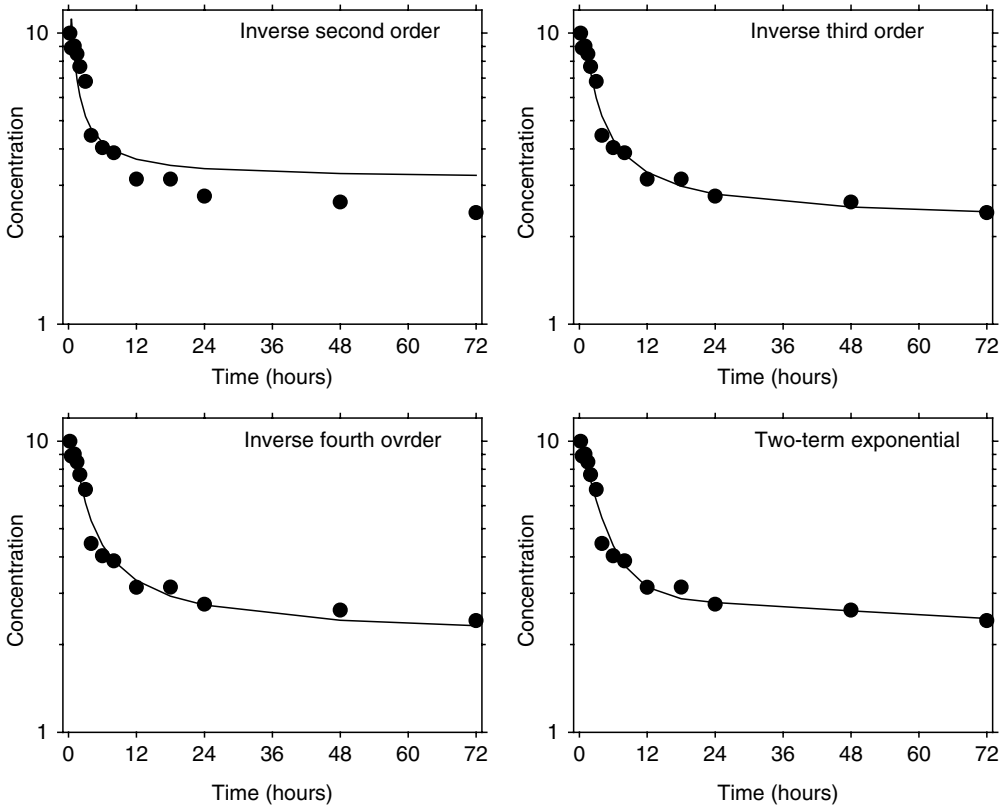

Figure 1.2 Scatter plots of simulated concentration-time data (•) and fitted models (solid lines). Data were simulated using a 2-term polyexponential model and were fit with inverse order polynomials up to degree four and to a 2-term exponential equation. The residual sum of squares for each model is shown in Table 2.

Table 1.2 Summary statistics for models fit to the data shown in Fig. 1.2.

Model	Number of estimable parameters	Residual sum of squares
Second order inverse polynomial	3	15.7
Third order inverse polynomial	4	1.80
Fourth order inverse polynomial	5	1.61
Two-term exponential model	4	2.47

used to generate the data would result in the best model. In other words, one would expect Eq. (1.6) to be the superior model to the other models since it is of the same form as data generating model. However, both Eqs. (1.4) and (1.5) resulted in smaller residual sum of squares than Eq. (1.6) which means these inverse polynomial models better predicted the observed concentrations than the exponential model. Occam's razor would lead us to choose Equation (1.4) over Eq. (1.5) since the former has fewer estimable parameters. This example illustrates why there is no such thing as a right model. Equation (1.4) is clearly the wrong model, but it in fact has better predictive properties than the equation used to generate the data. If interest were solely in being able to make predictions about the data at some point in time, say at 36 hours post-dose when no samples were collected, an inverse polynomial may be more useful as it will be more accurate than an exponential model.

So why aren't inverse polynomials used more frequently if they can have better predictive properties than exponential equations? The polyexponential equation is consistent with the theory for an n-compartmental system, which is one of the properties of a useful model. In this particular case, a two-term exponential equation is consistent with a 2-compartment model following bolus intravenous administration. The model parameters from the two-term exponential equation also directly translate to pharmacokinetic parameters, such as volume of distribution. There is no similar theory for inverse polynomials—they are strictly empirical equations. The parameters of an inverse polynomial have no physiological meaning. A useful model may also allow for extrapolations outside the range of data measured. For example, given a two-term exponential model [Eq. (1.6)] the limits for such a model are A+B when time equals zero and zero when time goes to infinity. This is what one would expect. Following bolus administration, concentrations are at their maximal and finite in value at time equal zero. As time goes towards infinity, all the drug in the body is eventually removed and concentrations approach zero. But for an inverse polynomial, at time equal zero, the dependent variable is undefined because inverse time (i.e., 1/0) does not exist. Taking the limit as time approaches zero, the dependent variable blows up towards infinity, which clearly is not possible as drug concentrations in the body must be finite since a finite amount of drug is given. At the other end of the time scale, when time approaches infinity, the dependent variable approaches the intercept term (β_1) because all the inverse time terms approach zero. So, the inverse polynomial model predicts concentrations to remain in the body infinitely equal in concentration to the model intercept. These concepts are illustrated in Fig 1.3. This example illustrates the hazards of extrapolating an empirical model outside the data range. So, despite better predictive properties, pharmacokineticists rely on models with pharmacokinetic interpretations that are consistent with theory.

In summary, a useful model, like the concept of a good model, is in the eye of the beholder. The model may fail at predicting certain aspects of the data, but if the modeler is not concerned with that portion of the data that is unexplainable, the model may still have value. Another modeler may argue, however, that the model is not useful since it fails to explain all aspects of the data. In another case, a modeler may also be quite satisfied at developing a model given the data on hand, but a project team may find the model to be useless if it cannot be used to help guide clinical development. Modelers must ever strive to make their models useful.

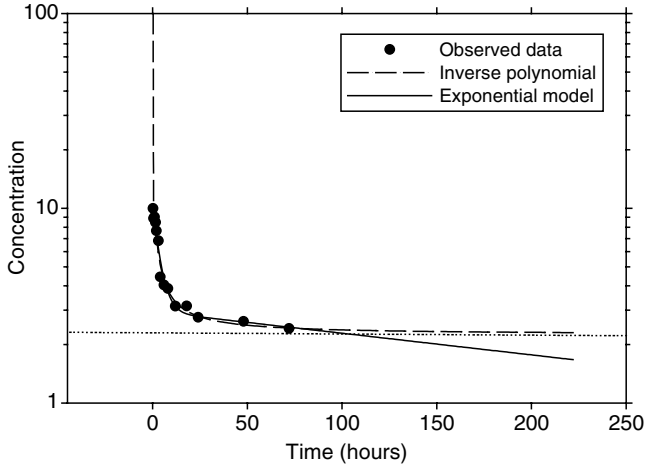

Figure 1.3 Extrapolation of the cubic inverse polynomial model and exponential model as time goes to zero and infinity. Observed data are denoted by ●. The solid line denotes the two-term exponential model, while the dashed line indicates the inverse polynomial of degree three. The dotted line denotes the intercept of the inverse polynomial model. In terms of residual sum of squares, the inverse polynomial model is the superior model but does not allow for extrapolation beyond the observed data range nor does the inverse polynomial model terms have any physiological meaning.

THE MODEL DEVELOPMENT PROCESS

Mesterton-Gibbons (1989) describes the modeling process as being as simple as ABC. Clearly that's not true, but it does make for a nice, catchy mnemonic. 'A' is for assume. Often there is inadequate information at the outset to solve a problem, except for the most simplest cases, so assumptions are needed right from the beginning. These assumptions may be in the form of parameter values, or model structure, or distributional assumptions, like the distribution of the model residuals. A model that has poor predictability may not be a poor model at all. Indeed, the problem may simply be that the assumptions are wrong. Next, 'B' is for borrow. Few models are developed in a vacuum. Most models are based on other models. Hence, knowledge is borrowed from the literature, from previous experience, or from colleagues and then a starting model is built to evaluate. 'C' is then to criticize the model and the assumptions the model was predicated upon. Modeling is iterative. If the model does not meet our needs then we go back to 'A,' modify our assumptions, and then start over again, hopefully learning from what we have just done.

While clever, most books on modeling don't use this simple mnemonic and instead present a more formal process initially proposed by Box and Hill (1967). They stated that the process of model-building can be thought to involve three stages:

1. Identification of the model,
2. Fitting the model, and
3. Diagnostically check the adequacy of the fit.

Although these stages have been repeatedly reported throughout the statistical literature, they are really only the middle part of the process. Chatfield (1988) expanded the number of stages to five, which include the original three by Box and Hill:

1. Look at the data,
2. Formulate a sensible model,
3. Fit the model to the data,
4. Diagnostically check the adequacy of the fit, and
5. Present the results and conclusions.

Models are not static—they change over time as more data and experience with the drug are accumulated. Basic assumptions made about a model may later be shown to be inaccurate. Hence, a more comprehensive model development process is:

1. Analyze the problem,
2. Identify relevant variables to collect,
3. Perform the experiment and collect data,
4. Look at, clean the data, and format for modeling,
5. Formulate a model,
6. Fit the model to the data,
7. Diagnostically check the adequacy of the fit,
8. Validate the model,
9. Update the model as appropriate (go back to Step 5),
10. Interpret the results, and
11. Communicate the results.

Figure 1.4 illustrates this process graphically.

The first step of the process should be to identify the problem, which has already been extensively discussed. The next step is to identify the relevant variables to collect. Data are usually not cheap to collect. There is ordinarily a fine line between money available to perform an experiment and the cost of collecting data. We want to collect as much data as possible, but if a variable is not needed then perhaps it should not be collected. Once the variables to be collected are identified, the accuracy and bias of the measurement methods should be examined because collected data are of no value if it is biased or inaccurate. Sometimes, however, the modeler is not involved in choosing which variables to collect and is brought in to analyze data after an experiment is already completed. It may be that the data needed to solve the problem was not collected or that only some of the data were collected, in which case some creative thinking may be needed to obtain a solution.

The next step is to perform the experiment and collect the data. Whether the data is a small scale animal study or a large scale clinical trial, the basics of data collection are the same. The validity of the results of any study are dependent on the quality of data collected. Therefore, data collection, whether stored electronically or on paper, must be designed to ensure high quality. Two keys to good quality are randomization and blinding, although in practice sometimes neither of these can be done. Document everything from a regulatory point of view; if it isn't documented, it never happened.

Rarely, however, will the data be in a format suitable for analysis. The more complex the data, the more pre-processing will be needed to put the data in a format suitable for analysis. The next step then is to look at the data and clean as needed. Check the quality of the data. Have the data been entered to suitable precision, for instance, two places behind the decimal? Perform descriptive statistics and look at histograms to examine for discordant results. It is not uncommon in large clinical multi-national clinical trials for clinical chemistry data to be of different units between the United States

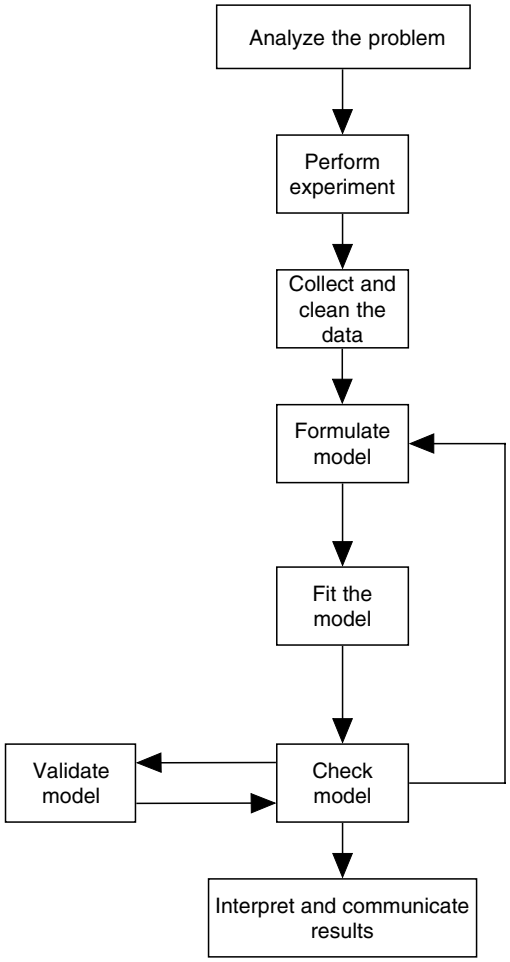

Figure 1.4 The model development process.

and Europe. Merging of data must then be transformed to a common unit before analysis can proceed. Data may have been wrongly entered into the computer. It is not uncommon to clean a data set and do an initial model fitting to the data only to find that something was missed and more data cleaning is needed. Whatever cleaning is done to the data should be documented in whatever final report is written on the analysis. Also, what quality control checks were made on the data should be documented. If data need to be transformed prior to analysis, such as a log-transformation on a dependent variable, then this too should be clearly documented in the report.

Once the data is cleaned and placed in a suitable format for analysis, herein begins the typical stages of model development reported by most books. Model development tends to be iterative in that a base model is chosen and evaluated. If the model is rejected, a new model is generated and evaluated. Once a suitable model is chosen, the model is validated to examine its generalizability. When a final model is found, the results are interpreted and presented, either in written or oral format. This process of model development is an empirical one, dependent on the data set used to build the model. Sometimes this is referred to as exploratory data analysis, data mining, or data dredging. Rare is the model in drug development built on theory, and then tested using experimental data—the so-called confirmatory model.

For particularly complex models, such as multicompartment, multiple input-output experiments, one trick to modeling such data is to break the model down into subsystems. So for instance, suppose concentration-time data for parent drug and metabolite are available and it is known the metabolite is formed by irreversible metabolism of parent drug. One way to model the data is to first define and fit a model for parent drug, ignoring the metabolite. Then once that model is identified, analyze the metabolite data using a forcing function based on the parent concentration-time profile as the input to the metabolite model. Once both models are identified, combine them together and then re-fit the joint model. A similar approach can be made for pharmacokinetic-pharmacodynamic models where the pharmacodynamic model is a simple function of the concentration data. First, find the pharmacokinetic model and fix the model parameters. Then find the pharmacodynamic model keeping the pharmacokinetic model fixed. Once both models are fitted, combine the models and fit the joint model simultaneously.

Within the empirical model development framework, model development iterates until a suitable model is chosen. But model development may also occur globally across researchers. An excellent example of between-scientist model development is with the models used to characterize the pharmacokinetics of paclitaxel (Taxol®), an agent that is used in the treatment of various cancers. Paclitaxel is a poorly soluble drug given by infusion and formulated in a mixture of alcohol and a polyoxyethylated castor oil called Cremophor EL (50/50, v/v). Early studies reported that paclitaxel pharmacokinetics could be characterized by a 2-compartment model with first-order elimination (Brown et al., 1991; Longnecker et al., 1987; Wierkin et al., 1987). Large between-subject variability was observed in the parameter estimates, e.g., clearance ranged from 53 to 1260 $mL/min/m^2$ (Longnecker et al., 1987). Model predicted concentrations were judged to be reasonably close to observed concentrations and later studies, using noncompartmental analysis, produced similar pharmacokinetic estimates (Grem et al., 1987; Wiernik et al., 1987). But these early studies used long infusion times, 6 to 26 hours in length.

When paclitaxel was given as short infusion, hypersensitivity reactions typically occurred. Hence, paclitaxel was typically given by prolonged infusion. It was specu-

lated that decreasing the length of infusion and premedicating with a corticosteroid and anti-histamine would decrease the occurrence of hypersensitivity reactions, and be more convenient for the patient. In a clinical study testing this hypothesis, the shorter infusion with premedication did indeed result in a lower incidence of hypersensitivity reactions and was also shown to be less neutropenic (Eisenhauer et al., 1994). Unfortunately, paclitaxel concentrations were not determined in this study. In a similar repeated study that did measure paclitaxel plasma concentrations, when paclitaxel was given as either a 3 or 24 hour infusion, systemic clearance estimates after the 24 hour infusion were greater than after the 3 hour infusion (Huizing et al., 1993). A 3-compartment model was now more consistent with the data, which the authors attributed to having a more sensitive analytical assay thereby detecting the presence of another phase in the concentration-time profile. The authors also speculated that saturable pharmacokinetics was occurring but did not attempt to include this phenomenon in their model.

Dose-dependent clearance and distribution was then later observed in a Phase 1 study in children with solid tumors (Sonnichsen et al., 1994). In a study in adults with ovarian cancer, Gianni et al. (1995) used a 3-compartment model with saturable intercompartmental clearance into Compartment 2 and saturable, Michaelis–Menten elimination kinetics from the central compartment to describe the kinetics after 3 hour and 24 hour infusion. Now at this point one would typically assume that the mechanism for nonlinear elimination from the central compartment is either saturable protein binding or saturable metabolism. But the story is not that simple. Sparreboom et al. (1996a) speculated that since Cremophor EL is known to form micelles in aqueous solution, even many hours after dilution below the critical micellular concentration, and can modulate P-glycoprotein efflux, that the nonlinearity in pharmacokinetics was not due to paclitaxel, but due to the vehicle, Cremophor EL. This hypothesis was later confirmed in a study in mice (Sparreboom et al., 1996b).

An *in vitro* study was then conducted with human red blood cells (RBCs) which showed that the blood to plasma ratio in the absence of Cremophor EL was 1.07, but after the addition of Cremophor EL giving concentrations similar to those seen at the end of a 3 hour infusion of 175 mg/m^2 paclitaxel, the blood to plasma ratio decreased to 0.69 (Sparreboom et al., 1999). Hence, Cremophor EL decreased the unbound fraction of paclitaxel available for distribution into tissues in a concentration-dependent manner, which explains the saturable tissue distribution phenomenon in multi-compartmental models. The presence of many compartments within the blood that paclitaxel may distribute into (unbound, plasma protein bound, RBCs, and Cremophor EL-derived micelles) also explains the nonlinear elimination kinetics from the central compartment. Current models now measure paclitaxel in each of these blood compartments and use a 3-compartment model with saturable elimination and saturable tissue distribution due to saturable transport, which is quite different than the first model developed for paclitaxel (Henningsson et al., 2001; van Zuylen et al., 2001). But the story is not over. Karlsson et al. (1997) have argued that, using plasma paclitaxel concentrations as the dependent variable, the current model for saturable tissue distribution due to saturable transport cannot be kinetically distinguished from a model with linear transport processes but with saturable, noninstantaneous tissue binding. So the modeling process continues. This example illustrates how since 1987 the pharmacokinetic models for paclitaxel have changed from simple linear pharmacokinetic models to complex nonlinear ones. Is there a universal model for paclitaxel pharmacokinetics? Yes. Will it ever be found? Maybe. Meanwhile, science and modeling progress.

GOODNESS OF FIT CRITERIA

Once a model is developed, the next step is to either assess how "good" the model is or to compare the model to alternative models in order to determine which model is "better." The words "good" and "better" are used because they are meant to represent semi-quantitative terms that intuitively one has a feeling for, but cannot really be defined. Goodness of fit criteria are either graphical in nature or are presented as some metric, like the coefficient of determination (R^2, which will be discussed later). Metric-like criteria have an advantage in that they are quantifiable. For example, a model with an R^2 of 0.9 can be judged superior to a model with an R^2 of 0.3, all other things being equal. However, few things beat a good graphical analysis to demonstrate the validity of a model and with today's software packages one would be remiss if these graphics were not examined on a routine basis.

Residuals and Residual Analysis

If Y is the observed data vector and $\hat{\mathbf{Y}}$ is the model predicted data vector, ordinary residuals are the difference between observed and model predicted values

$$e_i = Y_i - \hat{Y}_i \tag{1.7}$$

where $i = 1, 2, \ldots, n$. Positive residuals indicate that the model underpredicts the observation, whereas

negative residuals indicate that model overpredicts the observation. Residuals are usually assumed to be independent, normally distributed with mean zero and variance σ^2 if the model is appropriate. The examination of residuals as part of the model evaluation process is referred to as residual analysis, which is useful because it can aid in isolating outliers or erroneous data points, i.e., observations that are discordant from the others, can aid in determining if the model assumptions are wrong or whether a different structural model should be used, and can aid in detecting observations that exert undue influence on the reported model parameters. Other types of residuals exist, such as Studentized or weighted residuals (which will be discussed later in the chapter).

An unbiased model should have residuals whose mean value is near zero. For a linear model the residuals always sum to zero, but for a nonlinear model this is not always the case. Conceptually one metric of goodness of fit is the squared difference between observed and predicted values, which has many different names, including the squared residuals, the sum of squares error (SSE), the residual sum of squares, or error sum of squares

$$\mathrm{SSE} = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 = \sum_{i=1}^{n} e_i^2. \tag{1.8}$$

The problem with SSE is that SSE decreases as the number of model parameters increases. Alternatively one could calculate the variance of the residuals called the mean square error (MSE)

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n - p} \tag{1.9}$$

where p is the total number of estimable parameters in the model and the denominator is collectively referred to as the *degrees of freedom*[2]. MSE is an unbiased estimate of the error variance term σ^2 if the model is appropriate.

The residuals themselves also contain important information on the quality of the model and a large part of model evaluation consists of residual analysis. Most residual analyses involve graphical examination of systematic trends or departures from the expected values. The following plots are often created and examined after model creation:

1. Scatter plot of predicted value (ordinate) versus residual (abscissa). No systematic trend in the residuals should be observed with the data appearing as a shotgun blast. Systematic trends are indicative of model misspecification. See Fig. 1.5 for an example.

2. Plot of absolute or squared residuals versus predicted value. Again, no systematic trend in the residuals should be observed and the plot should appear as a shotgun blast. Heteroscedasticity or misspecification of the variance model is evident if a positive trend in squared or absolute residuals with increasing predicted

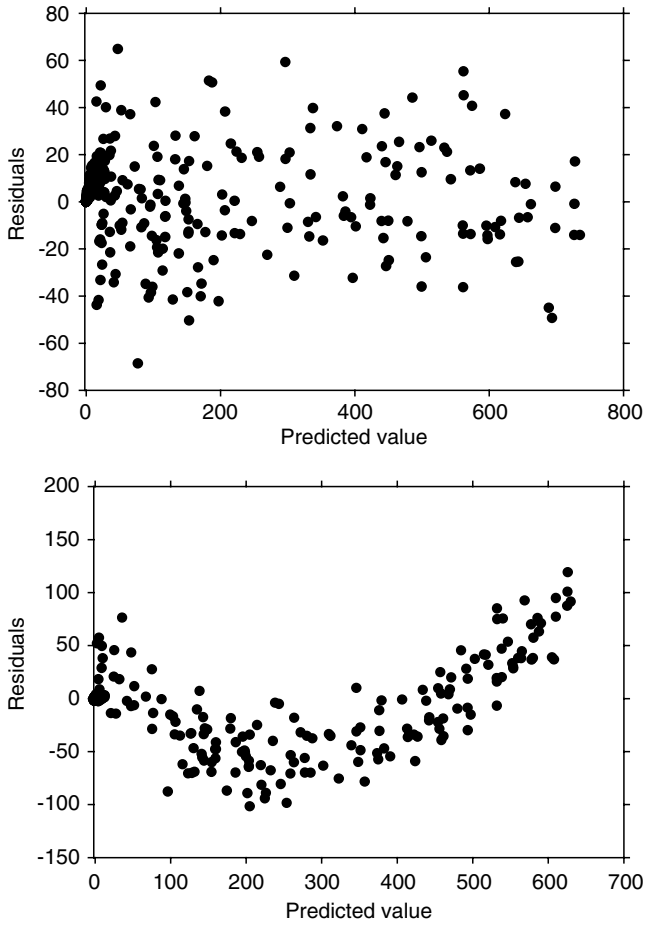

Figure 1.5 Sample residual plot. Paired (x, Y) data were simulated using the model $Y = 13 + 1.25x + 0.265x^2$. To each Y value was added random error from a normal distribution with mean zero and standard deviation 25. The top plot is a plot ordinary residuals versus predicted values when the fitted model was a second-order polynomial, the same model as the data-generating model. The bottom plot is the same plot when the fitted model was linear model (no quadratic term). Residual plots should appear as a shotgun blast (like the top plot) with no systematic trend (like the bottom plot).

[2] Defining degrees of freedom in a simple manner is difficult. First consider that there are n "pieces of information" contained within a data set having n observations. From these n pieces of information, either a parameter or variability can be estimated with each item being estimated decreasing the information in the data set by one degree of freedom. The degrees of freedom then is the number of pieces of information less all the estimated items. For example, given a data set in which the mean was estimated, the degrees of freedom then is $n - 1$. With a model having p-estimable parameters, the degrees of freedom is $n - p$.

values is observed. More will be discussed on this plot in the chapter on Variance Models, Weighting, and Transformations.

3. The residuals should also be uncorrelated so that if the residuals, e_1, e_2, e_3, etc., are lagged and plotted, i.e., e_1 vs. e_2, e_2 vs. e_3, etc., there should be no trend in the plot. When the residuals are correlated, such a process is termed 'autocorrelation' and unfortunately, this plot is rarely examined in pharmacokinetic/pharmacodynamic models.

4. A histogram of residuals. The histogram should show approximate normality with the center of mass located near zero (Fig. 1.6). Histograms, while easy to generate, do not easily detect subtle deviations from normality. More on histogram analysis is presented elsewhere in this chapter.

5. Normal probability plots or, half-normal plots are recommended instead of histograms for detecting deviations from normality. For a normally distributed random variable X with mean 0 and variance σ^2, a good approximation to the expected value of the ith observation is

$$E(X_i) = \sqrt{MSE}\left[\Phi^{-1}\left(\frac{i - 0.375}{n + 0.25}\right)\right] \tag{1.10}$$

where $\Phi^{-1}(f)$ denotes the fth percentile of the standard normal distribution. A standard normal plot plots the expected value of the residual against the value of the residual itself (ordered from smallest to largest). One criticism of the normal plot is what has been called 'supernormality' (Atkinson, 1985) in that residuals from non-normal distributions will tend to appear more normal than they truly are. Thus an adequate normal plot in and of itself is not confirmatory for a normal distribution. A modification of the normal plot, used to combat supernormality, is the half-normal plot where

$$\sqrt{MSE}\left[\Phi^{-1}\left(\frac{n + i + 0.5}{2n + 9/8}\right)\right] \tag{1.11}$$

is plotted against the absolute value of the ith residual, again sorted from smallest to largest. Half-normal plots tend to show more sensitivity to kurtosis at the expense of not showing skewness but are more sensitive at detecting outliers and influential observations than normal plots.

6. Another type of plot, called a QQ (quantile-quantile) plot, plots the residuals ordered from smallest to largest against

$$\Phi^{-1}\left(\frac{i - 0.5}{n}\right). \tag{1.12}$$

Normal, half-normal, and QQ plots that are linear are consistent with normality, whereas non-normal distributions tend to have systematic deviations from linearity. It should be noted that some software packages omit the $\sqrt{MSE}$ term in Eq. (1.10) since this omission has no impact on the nature of the plot, simply a change in intercept. Atkinson (1981) suggests plotting a "envelope" around the plot using simulation to aid in the interpretation of the plot but no software packages do this and so the reader is referred there for more details.

Figure 1.7 presents an example of the normal, half-normal, and QQ plot for simulated data from a normal, chi-squared distribution with four degrees of freedom, and Student's T-distribution with four degrees of freedom. It must be stressed that these plots are not always conclusive and that with small sample sizes normality can easily be mistaken for non-normality.

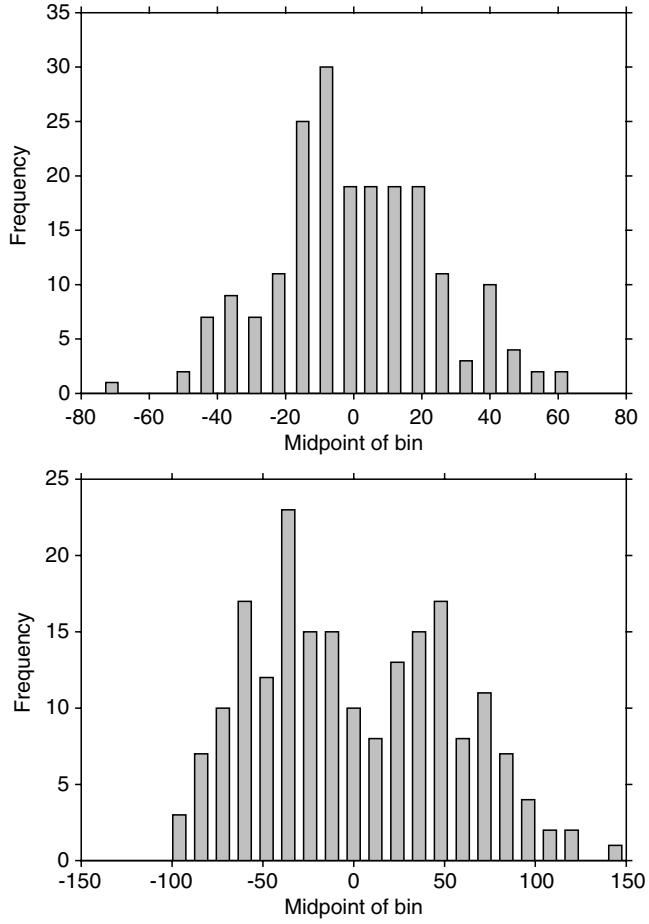

Figure 1.6 Histogram of residuals from data in Fig. 1.5. Top plot is quadratic model. Bottom plot is plot of linear model.

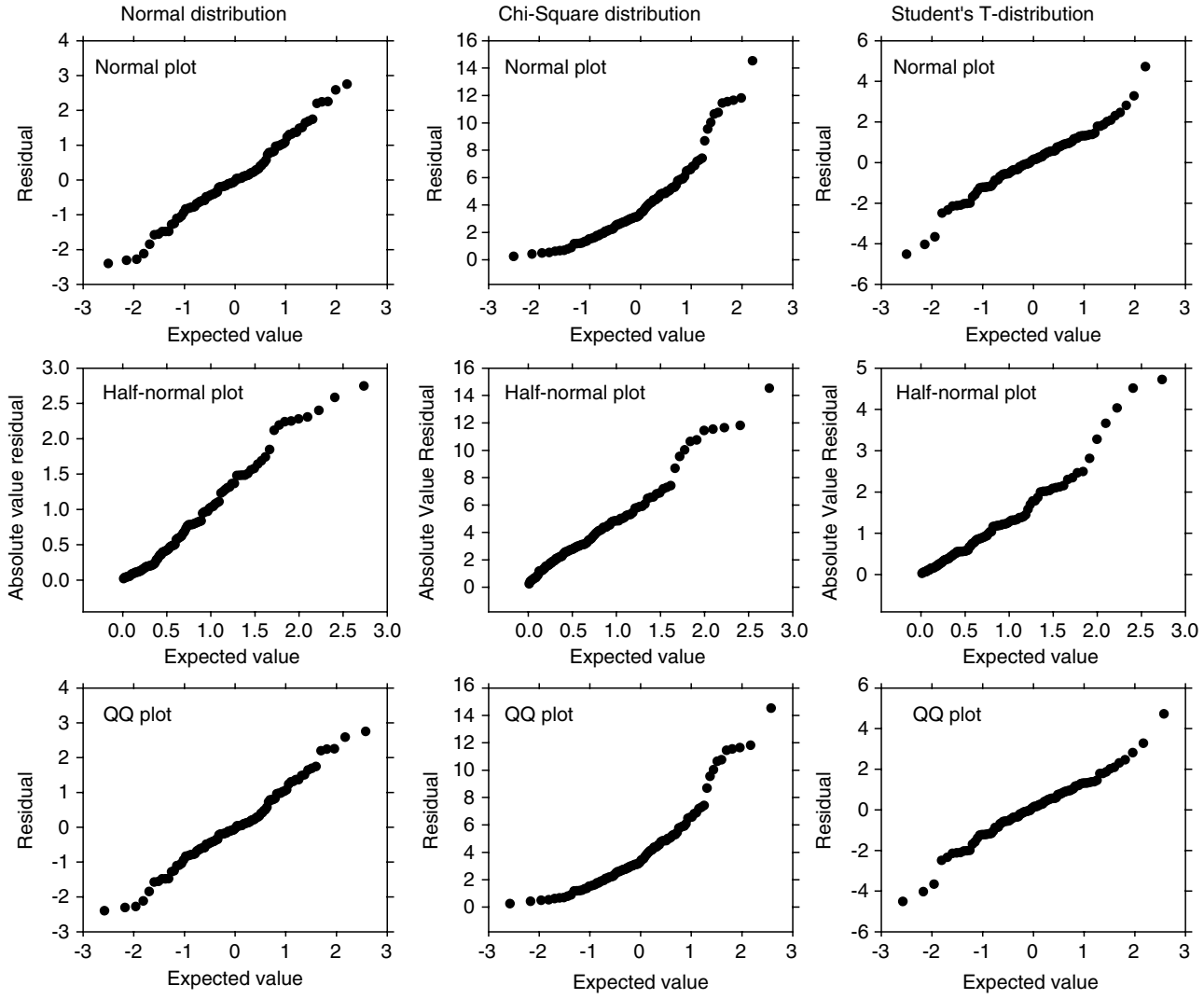

Figure 1.7 Normal, half-normal, and QQ plots for 100 simulated observations from a normal distribution (left), chi-squared distribution with four degrees of freedom (middle), and student's T-distribution with four degrees of freedom (right). If the data are consistent with a normal distribution, the resulting plots should all show approximate linearity with no curvatures. The normal plot and QQ plot are usually indistinguishable. The half-normal plot is usually more sensitive at detecting departures from normality than the normal or QQ plot.

7. A plot of residuals against explanatory variables not included in the model is useful to detect whether the explanatory variable should be included in the model. The plot should show no systematic trend if the explanatory variable is not predictive, but if the plot shows a systematic trend then this is evidence that perhaps the variable should be included in the model.

8. If one of the independent variables is time, a scatter plot of residuals versus time is useful. The plot should show random variation centered around zero. Systematic trends in the plot indicate the model does not predict the time trend accurately. More formally a runs test or Durbin–Watson test can be performed to test for lack of randomness.

9. If multiple observations are available on each sampling unit, such as a subject in a clinical trial, a plot of residuals versus subject number may be informative at detecting systematic deviations between subjects. Each subject's residuals should be centered around zero with approximately the same variance. Subjects that show systematic deviance from the model will tend to have all residuals above or below the zero line. This plot becomes more useful as the number of observations per subject increases because with a small number of

observations per subject, the sensitivity of the plot at detecting deviance from the model decreases.

Many of the plots just suggested are not limited to ordinary residuals. Weighted residuals, partial residuals, studentized residuals, and others can all be used to aid in model diagnostics. Beyond residual plots, other plots are also informative and can help in detecting model inadequacies. One notable plot is a scatter plot of observed versus predicted values usually with the line of unity overlaid on the plot (Fig. 1.8). The model should show random variation around the line of unity. Systematic deviations from the line indicate model misspecification whereas if the variance of the predicted values increases as the observed values increase then the variance model may be inappropriate.

While informal, an informative graph can be helpful in detecting gross violations of the model assumptions. A graph is also an effective means to communicate how well a model performs. Subtle differences between two competing models, however, usually cannot be differentiated by the basis of graphics unless the two models are highly dissimilar. It should also be noted that data transformations (which will be discussed in later chapters) may also affect these plots considerably. For further details on residual analysis the reader is referred to Atkinson (1985) or Cook and Weisberg (1982).

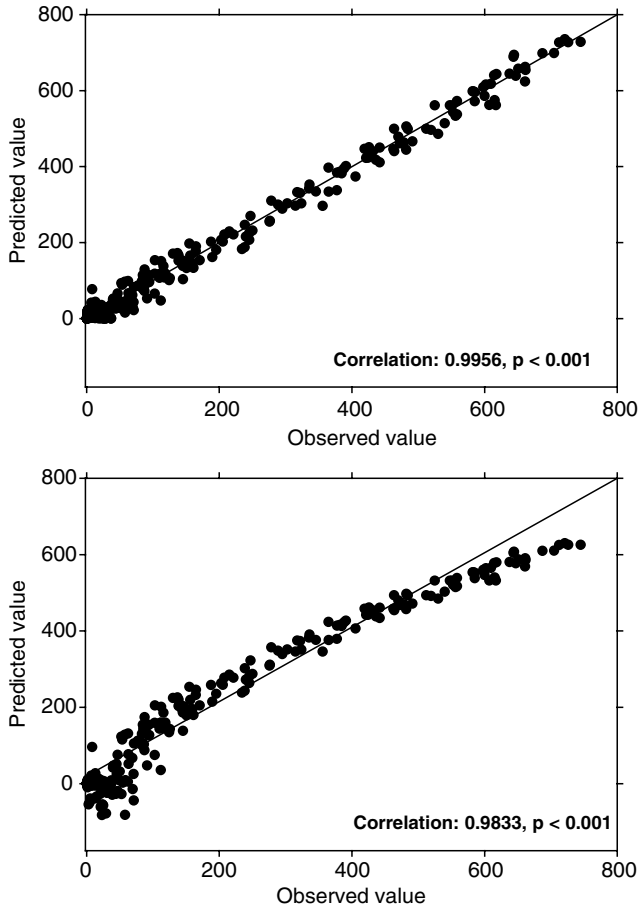

Figure 1.8 Plot of observed versus predicted values for the data in Fig. 1.5. The top plot shows reasonable goodness of fit while the bottom plot shows a systematic trend suggesting the model is not a reasonable approximation to the data. Solid line is line of unity. Note that both plots have very high correlation coefficients, despite one having a systematic trend from the line of unity.

Goodness of Fit Metrics

Along with graphical assessment one may present metrics, actual numbers, that attempt to quantify the goodness of fit of the model. Two such metrics were presented in the previous section, SSE and MSE, and in this section other metrics will be presented. Formal hypothesis tests may be done on these metrics, such as comparing the metric from one model against another. However, many test statistics based on these metrics tend to be sensitive to the assumption of the underlying distribution, e.g., normally distributed, such that the results from these tests should be treated with skepticism (Cook and Weisberg, 1982).

Common modifications to SSE and MSE lead to a class of metrics called discrimination functions. These functions, like the Akaike Information Criteria (AIC), are then used to choose between competing models. One problem with functions like the AIC and MSE is that the actual value of the function is impossible to interpret without some frame of reference. For instance, how can one interpret a MSE or an AIC of 45? Is that a good or bad? Further, some discrimination functions are designed to be maximized whereas others are designed to be minimized. In this book, the model with the smallest discrimination function is superior to all other models having the same number of estimable parameters, unless otherwise noted. This class of functions will be discussed in greater detail in the section on Model Selection Criteria.

Three goodness of fit metrics bear particular attention: the coefficient of determination, the correlation coefficient, and the concordance coefficient. The coefficient of determination (R^2) is simply

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}. \qquad (1.13)$$

where $\overline{Y}$ is the mean of the observed Y values. The reason R^2 is so often used is its ease of interpretation—it explains the proportion of variation "explained" by

the model and ranges from 0 to 1 with 1 being a perfect fit to the data. Still what constitutes a good R^2 is debatable and depends on what is being measured. The R^2 for an analytical assay should be very high, probably greater than 0.98, while an R^2 greater than 0.4 may be acceptable some cases, like correlating apparent oral clearance to creatinine clearance.

R^2 is not perfect and has many flaws. Many cautionary notes have been written on the misuse of R^2 (Healy, 1984; Kvalseth, 1985). For example, with linear models having no-intercept or with nonlinear models, it may be possible for SSE to be larger than SST leading R^2 to be negative. R^2 is also influenced by the range of the observations with the wider the range of the independent variable, the larger R^2 tends to be (Helland, 1987). Further, when additional terms are added to a model, R^2 will always increase because SSE will always decrease. Since R^2 can be artificially increased due to additional model parameters, an adjusted R^2 is often used that adjusts R^2 for the additional degrees of freedom

$$R^2_{adj} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SST}. \tag{1.14}$$

Thus, the adjusted R^2 may decrease when additional terms are added to the model and they do not contribute to the goodness of fit.

Related to the coefficient of determination is the correlation coefficient, ρ, which is almost exclusively used in the association between two variables, X and Y. In relation to goodness of fit, X is the observed dependent variable, e.g., plasma drug concentrations, and Y is the model predicted dependent variable, e.g., predicted plasma drug concentrations, such as the plot shown in Figure 8. In the case of two variables ρ has maximum likelihood estimator

$$\hat{\rho} = r = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}. \tag{1.15}$$

r is a biased estimator for ρ, but becomes less biased as n goes to infinity. Since the covariance between two random variables is bounded by the product of the individual standard deviations, ρ is bounded by ± 1. r is also called Pearson's correlation coefficient for his pioneering work in this area, although the symbol 'r' was first noted by Sir Frances Galton for his work on regression towards the mean (which he called reversion) (Pearson, 1920; Zar, 1984).

Pearson's product-moment correlation coefficient, often simply referred to as the correlation coefficient, r, has two interesting properties. First,

$$\sqrt{R^2} = \pm r. \tag{1.16}$$

Similarly, the maximum likelihood estimate for ρ^2 is R^2, which is why the coefficient of determination is usually mentioned in the same breath as the correlation coefficient. Second, r is scale invariant. If X and/or Y is multiplied by 10 or 1000, r does not change.

The correlation coefficient, r, is probably the most misused statistic in science. Much has been written criticizing the reporting of correlation coefficients. Foremost is the argument that X and Y must be *bivariate normal* for correlation results to be valid (Analytical Methods Committee and Royal Society of Chemistry, 1988), which is not exactly correct. The interpretation of the correlation coefficient depends on whether X and Y is random or fixed. If X and Y are random, then indeed they should be bivariate normal and r is an estimate of ρ, the population correlation coefficient. However, if x is fixed and Y is a function of x, then r is interpreted as the square root of the coefficient of determination. In both instances, the correlation coefficients are identical. However, the interpretation is subtly different.

Second, reporting of correlation coefficients without a graphical depiction of the data can be exceedingly misleading. Figure 1.9 shows four plots with misleading correlation coefficients [these examples were suggested by Harmatz and Greenblatt (1992)]. In all four cases, the correlation coefficients were highly significant, but clearly there is something going on in the underlying structure of the data. The bottom left plot also shows how a single data point can produce a highly significant correlation coefficient.

Third, correlation coefficients have been criticized for their use in assessing goodness of fit (Harmatz and Greenblatt, 1992). Correlation coefficients are a measure of association, not a measure of goodness of fit. A model may have a high correlation but still have poor predictive qualities. One particular area where correlation coefficients have been abused is in the assessment of linearity of an analytical method for measuring drug concentrations in biological fluids. The correlation coefficient does not indicate linearity or lack thereof. Figure 1.10 shows an example of a scatter plot where nonlinearity is occurring at high concentrations, but still leads to an exceedingly significant correlation coefficient. Another area where correlations are inappropriately used is in the assessment of dose proportionality in clinical studies of new chemical entities.

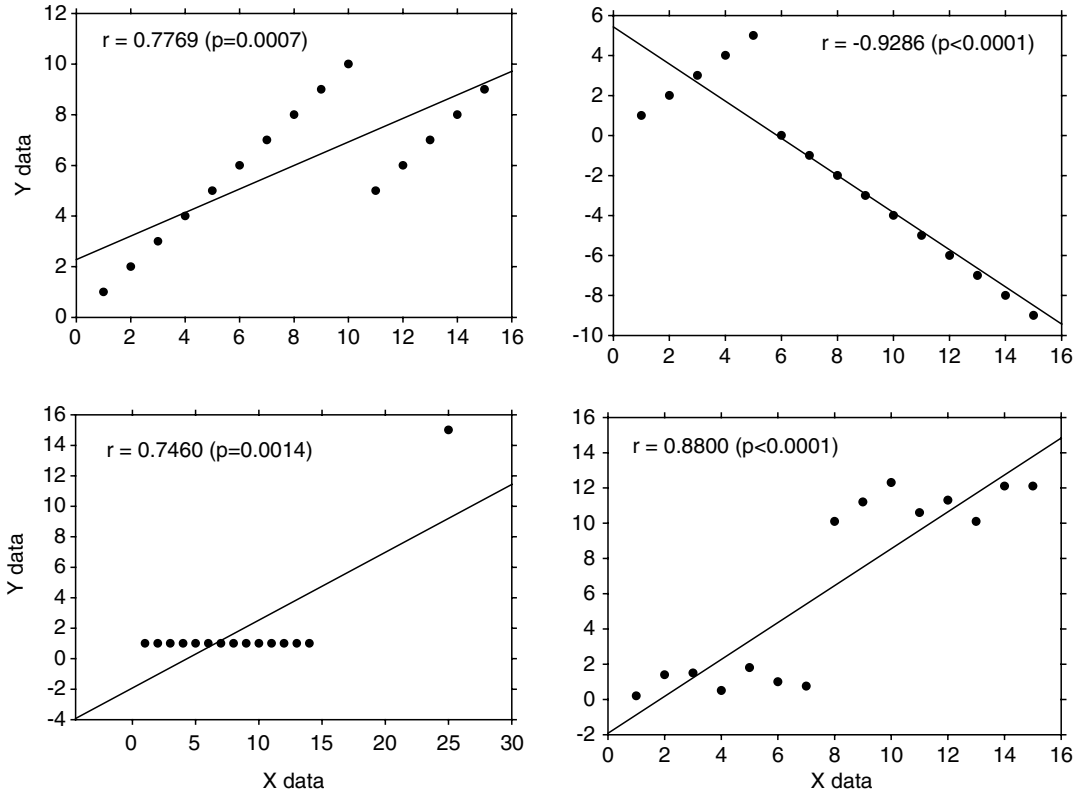

Figure 1.9 Example plots of misleading correlation coefficients suggested by Harmatz and Greenblatt (1992). Note that, in contrast to R^2, the correlation coefficient is independent of the scale of the X- and Y-axes.

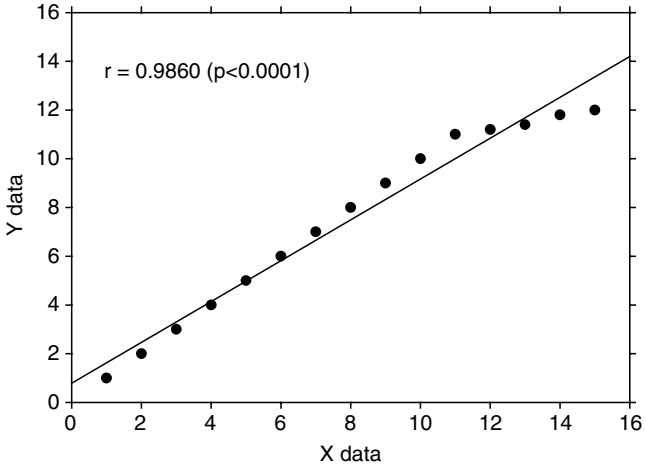

Figure 1.10 Example of a misleading correlation coefficient used in assessing linearity. The correlation coefficient for the simple linear model is quite high ($r = 0.9860$) but notice that the model systematically overpredicts higher concentrations, while underpredicts concentrations in the middle of the data range. This model shows systematic bias in its prediction despite having an excellent correlation.

Fourth, the correlation coefficient is reported not for their magnitude, but for their significance, e.g., $p < 0.01$. Tests of this nature are simply testing whether the correlation coefficient equals zero or not. Since the significance of a correlation coefficient is dependent on the sample size, large sample sizes easily can lead to significant correlations. For example, a correlation coefficient of 0.20, which under a linear model indicates that 4% of the variance is explained by the predictor variables, becomes significant at $p \leq 0.05$ when the sample size is more than 100. A more relevant hypothesis test is the one-sided null hypothesis that the correlation is greater than some value deemed by the analyst to have value, such as $\rho > 0.95$.

One other point needs mentioning in regard to correlation, and that is **correlation does not imply causality**. These last words are highlighted to stress their importance. Table 1.3 presents the nine tenets for causality as presented by Hill (1965). Correlation by its nature implies that X and Y can be reversed without loss of generality. Causality implies cause and effect. Just because a significant correlation has been detected between X and Y, it is entirely possible that the relationship is not

Table 1.3 Basic tenets of causality as proposed by Hill (1965).

Tenet	Meaning
• Strength of association	A high correlation between the causal variable and outcome is needed.
• Consistency	The results should be repeatable and consistent across studies.
• Specificity	The outcome is specific for the causal variable.
• Temporality	Changes in the causal variable should lead to changes in the outcome variable.
• Biological gradient	The more intense the causal variable the more intense the outcome variable.
• Biologic plausibility	There should be biological basis for the cause and effect.
• Biologic coherence	Implies a cause-and-effect interpretation.
• Experimental evidence	Experimental evidence supports the theory and is consistent with causality.
• Analogy	Similar to other cause-and-effect outcomes.

Note: Hill originally applied these criteria to the causality between risk factors and disease in epidemiology. These criteria have been modified to reflect causality between a causal variable and outcome variable in general.

due to an unknown, unobservable variable Z. For example, if X acts on Z and Z acts on Y, but Z is unobservable, an artifactual relationship between X and Y may exist. Still, it is easy to lose sight of the fact that just because a variable is correlated with another does not mean that a causal relationship exists.

Sheiner and Beal (1981) have pointed out the errors involved in using the correlation coefficient to assess the goodness of fit (GOF) in pharmacokinetic models. Pearson's correlation coefficient overestimates the predictability of the model because it represents the "best" linear line between two variables. A more appropriate estimator would be a measure of the deviation from the line of unity because if a model perfectly predicts the observed data then all the predicted values should be equal to all the observed values and a scatter plot of observed vs. predicted values should form a straight line whose origin is at the point (0,0) and whose slope is equal to a 45° line. Any deviation from this line represents both random and systemic error.

At the time the Sheiner and Beal (1981) paper was published, there were no good measures to assess the deviation from a 45° line. Lin (1989) developed the concordance coefficient to assess the reproducibility of two assay methods and was designed to correct some of the problems associated with the correlation coefficient, thereby measuring what Sheiner and Beal (1981) proposed almost 10 years earlier. The concordance coefficient (ρ_c) between X and Y measures the degree of agreement between X and Y by assessing the degree to which data pairs fall on the 45° line through the origin and can be estimated by

$$\hat{\rho}_c = \frac{2S_{xy}}{S_x^2 + S_y^2 + (\overline{X} - \overline{Y})^2}. \tag{1.17}$$

where

$$S_x^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}, \tag{1.18}$$

$$S_Y^2 = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n}, \text{ and} \tag{1.19}$$

$$S_{xy} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n}. \tag{1.20}$$

The concordance coefficient has the following properties:

1. $-1 \le |\rho| \le \rho_c \le |\rho| \le 1$,
2. $\rho_c = 0$ if and only if $\rho = 0$,
3. $\rho_c = \rho$ if and only if $\mu_x = \mu_y$ and $\sigma_x = \sigma_y$,
4. $\rho_c = \pm 1$ if and only if readings are in perfect agreement or perfect reversal, and
5. $|\rho_c|$ can be < 1 even if $|\rho| = 1$.

The reader is referred to Lin (1981) for details on calculating the variance of the concordance coefficient, which is not easily done. One disadvantage to the concordance coefficient is that, like the correlation coefficient and coefficient of determination, there are no guidelines as to what constitutes good agreement between observed and predicted values. This is left to the analyst to decide. Similarly the concordance coefficient can be just as misleading as a correlation coefficient and all the caveats regarding use of the correlation coefficient apply to the concordance coefficient as well. Vonesh, Chinchilli, and Pu (1996) present the use of concordance coefficient in population models.

In summary, the measures presented in this section represent metrics that assess the goodness of fit in a