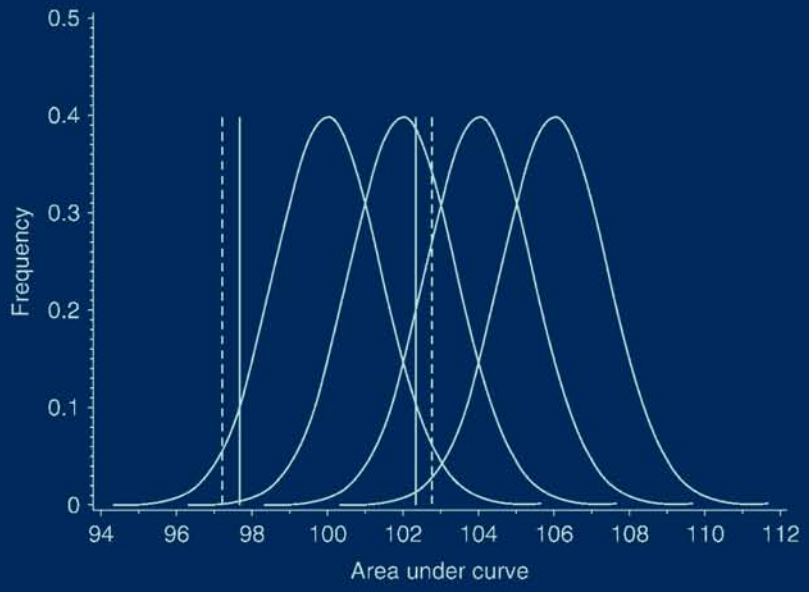
Chapman & Hall/CRC Biostatistics Series

Design and Analysis of Bioavailability and Bioequivalence Studies

Third Edition



Shein-Chung Chow

Jen-Pei Liu

CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Design and Analysis of Bioavailability and Bioequivalence Studies

Third Edition

Chapman & Hall/CRC Biostatistics Series

Chapman & Hall/CRC Biostatistics Series

Published Titles

1. *Design and Analysis of Animal Studies in Pharmaceutical Development,* Shein-Chung Chow and Jen-pei Liu
2. *Basic Statistics and Pharmaceutical Statistical Applications,* James E. De Muth
3. *Design and Analysis of Bioavailability and Bioequivalence Studies, Second Edition, Revised and Expanded,* Shein-Chung Chow and Jen-pei Liu
4. *Meta-Analysis in Medicine and Health Policy,* Dalene K. Stangl and Donald A. Berry
5. *Generalized Linear Models: A Bayesian Perspective,* Dipak K. Dey, Sujit K. Ghosh, and Bani K. Mallick
6. *Difference Equations with Public Health Applications,* Lemuel A. Moyé and Asha Seth Kapadia
7. *Medical Biostatistics,* Abhaya Indrayan and Sanjeev B. Sarmukaddam
8. *Statistical Methods for Clinical Trials,* Mark X. Norleans
9. *Causal Analysis in Biomedicine and Epidemiology: Based on Minimal Sufficient Causation,* Mikel Aickin
10. *Statistics in Drug Research: Methodologies and Recent Developments,* Shein-Chung Chow and Jun Shao
11. *Sample Size Calculations in Clinical Research,* Shein-Chung Chow, Jun Shao, and Hansheng Wang
12. *Applied Statistical Design for the Researcher,* Daryl S. Paulson
13. *Advances in Clinical Trial Biostatistics,* Nancy L. Geller
14. *Statistics in the Pharmaceutical Industry, Third Edition,* Ralph Buncher and Jia-Yeong Tsay
15. *DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments,* David B. Allsion, Grier P. Page, T. Mark Beasley, and Jode W. Edwards
16. *Basic Statistics and Pharmaceutical Statistical Applications, Second Edition,* James E. De Muth
17. *Adaptive Design Methods in Clinical Trials,* Shein-Chung Chow and Mark Chang
18. *Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries,* Daryl S. Paulson
19. *Statistical Design and Analysis of Stability Studies,* Shein-Chung Chow
20. *Sample Size Calculations in Clinical Research, Second Edition,* Shein-Chung Chow, Jun Shao, and Hansheng Wang
21. *Elementary Bayesian Biostatistics,* Lemuel A. Moyé
22. *Adaptive Design Theory and Implementation Using SAS and R,* Mark Chang
23. *Computational Pharmacokinetics,* Anders Källén
24. *Computational Methods in Biomedical Research,* Ravindra Khattree and Dayanand N. Naik
25. *Medical Biostatistics, Second Edition,* A. Indrayan
26. *DNA Methylation Microarrays: Experimental Design and Statistical Analysis,* Sun-Chong Wang and Arturas Petronis
27. *Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition,* Shein-Chung Chow and Jen-pei Liu

Chapman & Hall/CRC Biostatistics Series

Design and Analysis of Bioavailability and Bioequivalence Studies

Third Edition

Shein-Chung Chow

Jen-pei Liu

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-58488-668-6 (Hardcover)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Preface

As the first decade of the twenty-first century draws to an end, the arena of bioavailability and bioequivalence has generated a lot of scientific, statistical, and regulatory activities and issues from the pharmaceutical industry, health authorities, as well as academia, since the publication of the second edition of our book in 2000. In particular, a series of regulatory guidelines or guidances were issued by different health authorities in the world. In January 2001, the U.S. Food and Drug Administration (FDA) issued the guidance on *Statistical Approaches to Establishing Bioequivalence*. Six months later, in July 2001, the European Agency for the Evaluation of Medicinal Products (EMEA) issued the *Note for Guidance on the Investigation of Bioavailability and Bioequivalence*. In March 2003, the U.S. FDA released the guidance on *Bioavailability and Bioequivalence Studies for Orally Administrated Drug Products—General Considerations*. Later, the World Health Organization, in 2005, issued the draft revision of the guidelines on *Multisource* (*Generic*) *Pharmaceutical Products*: *Registration Requirements to Establish Interchangeability*. On the other hand, tremendous opportunities as well as challenges still lie ahead for bioavailability and bioequivalence in the twenty-first century because of breakthroughs in biotechnology and methodological research in medicine, pharmacokinetics, and statistics. In response to the challenges, upon the invitation of Professor R.B.D'Agostino, one of the co-editors of *Statistics in Medicine*, we were invited as guest editors for a special issue of 13 papers on individual bioequivalence that was published on October 30, 2000. In addition, the U.S. FDA issued a document on *Critical Path Opportunities for Generic Products* on May 1, 2007 to address emerging challenges and opportunities for generic drug products. The issues on the regulations and scientific issues of biosimilar products or follow-on biologics still remain unresolved. Consequently, there is an urgent need for the third edition of this book to provide a complete and overall presentation of the latest development of activities and results in bioavailability and bioequivalence on regulatory requirements, scientific and practical issues, and statistical methodology.

The third edition is different from the first and second editions in four aspects. First, we have revised and updated each section to reflect recent developments in statistical methodology in the design and analysis of bioavailability and bioequivalence studies. For example, the third edition provides a complete update of the status of regulations on bioavailability and bioequivalence, especially, the guidelines issued by the U.S. FDA, EMEA, and WHO. Second, the third edition is expanded to 20 chapters, 4 chapters more than the second edition and 8 chapters more than

the first. The third edition includes four new chapters as well as some new sections to present a complete account of the new developments in bioavailability and bioequivalence studies. The four new chapters include ''Population Pharmacokinetics'' (Chapter 17), ''Other Pharmacokinetic Studies'' (Chapter 18), ''Review of Regulatory Guidances on Bioequivalence'' (Chapter 19), and ''Frequently Asked Questions and Future Challenges'' (Chapter 20). Third, to deliver an effective presentation of the material, we modified the configurations of the 20 chapters into 5 parts: ''Preliminaries,'' ''Average Bioequivalence,'' ''Population and Individual Bioequivalence,'' ''*In Vitro* and Alternative Evaluation of Bioequivalence,'' and ''Other Bioequivalence Studies.'' Part I, ''Preliminaries'', describes the regulatory history of bioavailability and bioequivalence, design of bioavailability studies, and statistical inference for the standard 2×2 crossover design. Part II, ''Average Bioequivalence,'' reviews the methods for evaluation of average bioequivalence, power and sample size determination, transformation and assessment of intra- and inter-subject variabilities, outlier detection, and higher-order designs for evaluation of average bioequivalence. Part III, ''Population and Individual Bioequivalence,'' gives an update of the methods for the design and analysis of population and individual bioequivalence. Part IV, ''*In Vitro* and Alternative Evaluation of Bioequivalence,'' includes assessment of average bioequivalence with negligible plasma levels, *in vitro* bioequivalence studies, and *in vitro* dissolution profile comparison. Part V, ''Other Bioequivalence Studies,'' consists of meta-analysis for bioequivalence review, population pharmacokinetics, other pharmacokinetic studies, review of regulatory guidance, and future challenges. Finally, the third edition has 120 new references from the bioavailability and bioequivalence literature.

Similar to the first two editions, the third edition is also entirely devoted to the design and analysis of bioavailability and bioequivalence studies. It covers all of the statistical issues that may occur in the various stages of design and data analysis in bioavailability and bioequivalence studies. We strongly believe that this new, updated and much expanded third edition not only is an extremely useful reference book for pharmaceutical scientists and researchers, regulatory reviewers, clinicians, and biostatisticians in the academia, regulatory agencies, and pharmaceutical industry but also serves as an advanced textbook for graduate courses dealing with the topics of bioavailability and bioequivalence in the areas of pharmacokinetics, clinical pharmacology, and biostatistics. It is also our intent that this book will serve as a bridge among the pharmaceutical industry, government regulatory agencies, and academia.

Although the information, material, and presentation configuration of the third edition are different from the first two editions, the third edition still focuses on concepts rather than technical details. The mathematics and statistics dealt in the book are still fundamental. We have received many positive and constructive feedbacks and comments from scientists and researchers in academia, regulatory agencies, including the FDA, and pharmaceutical industry. Therefore, we have maintained our intuitive writing style as well as the emphasis on concepts through numerous examples and illustrations.

We would like to thank Jessica Vakili and David Grubbs of Taylor & Francis for their administrative assistance and support. We are deeply indebted to the Duke

University School of Medicine (especially, Rob Califf, MD, Robert Harrington, MD, Ralph Corey, MD, John McHutchison, MD, and Wesley Burks, MD) and the National Taiwan University for their encouragement and support. We also want to express our sincere gratitude to many pharmaceutical scientists, researchers, and biostatisticians for their feedbacks, support, and encouragement. Chow wishes to thank his fiancée Annpey Pong, PhD, for her constant encouragement and support during the preparation of this edition. Liu wishes to express his appreciation to his wife, Professor Wei-Chu Chie, MD, PhD, and daughter, Angela, for their patience, endurance, understanding, and support during the preparation of this edition.

Finally, we are fully responsible for any errors remaining in the book. The views expressed in this book are those of the authors and are not necessarily those of the Duke University School of Medicine, the National Taiwan University, or the National Health Research Institutes of Taiwan.

Shein-Chung Chow
Jen-pei Liu

Authors

Shein-Chung Chow, PhD is currently a professor in the Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina. Before joining Duke University, he was the director of the Taiwan Cooperative Oncology Group Statistical Center and the executive director of National Clinical Trial Network Coordination Center of Taiwan. Dr. Chow has also held various positions in the pharmaceutical industry including vice president. He has worked in the areas of biostatistics and data management, and as a medical writer at Millennium Pharmaceuticals, Inc., Cambridge, Massachusetts. He had also been executive director, statistics and clinical programming, at Covance, Inc. Princeton, New Jersey; director and department head at Bristol-Myers Squibb, Plainsboro, New Jersey; senior statistician at Parke-Davis Pharmaceutical Division, Warner-Lambert Company, Ann Arbor, Michigan; and research statistician at Wyeth-Ayerst Laboratories, Rouses Point, New York. Through these positions, Dr. Chow provided technical supervision and guidance to project teams on statistical issues and made presentations for partners, regulatory agencies, or scientific bodies, defending the appropriateness of statistical methods used in clinical trial design, data analyses, and the validity of reported statistical inferences. Dr. Chow has identified best statistical and data management practices, organized and led working parties for the development of statistical design, analyses, and presentation applications, and has participated on many data safety monitoring boards.

Dr. Chow's professional activities include playing key roles in many professional organizations such as officer, board of directors member, advisory committee member, and executive committee member. He has served as program chair, session chair/moderator, panelist, and instructor/faculty at many professional conferences, symposia, workshops, tutorials, and short courses. He is the editor-in-chief of the *Journal of Biopharmaceutical Statistics*. Dr. Chow is also the editor-in-chief of the CRC Press biostatistics series. He was elected fellow of the American Statistical Association in 1995 and was elected member of the International Statistical Institute in 1999. He was the recipient of the DIA Outstanding Service Award (1996), the ICSA Extraordinary Achievement Award (1996), and the Chapter Service Recognition Award of the American Statistical Association (1998). Dr. Chow was scientific advisor to the Department of Health, Taiwan, Republic of China during 1999–2001 and from 2006 till date. Dr. Chow was president of the International Chinese Statistical Association, chair of the advisory committee on Chinese pharmaceutical affairs, and a member of the advisory committee on statistics of the DIA.

Dr. Chow has authored/co-authored over 170 methodology papers and 14 books, which include *Advanced Linear Models*, *Design and Analysis of Bioavailability and Bioequivalence Studies* (first and second editions), *Statistical Design and Analysis in Pharmaceutical Science*, *Design and Analysis of Clinical Trials* (first and second editions), *Design and Analysis of Animal Studies in Pharmaceutical Development*, *Encyclopedia of Biopharmaceutical Statistics* (first and second editions), *Sample Size Calculations in Clinical Research*, *Adaptive Design Methods in Clinical Trials*, and *Statistical Design and Analysis of Stability Studies*.

Dr. Chow received a BS in mathematics from the National Taiwan University, Taipei, Taiwan and a PhD in Statistics from the University of Wisconsin, Madison.

Jen-pei Liu, PhD is currently the director of Statistical Education Center, Nationa Taiwan University, the director of Consulting Center for Statistics and Bioinformatics, National Taiwan University, and a Professor of Statistics, Division of Biometry, Department of Agronomy, National Taiwan University, Taipei, Taiwan. Professor Liu is a clinical biostatistician with more than 20 years experience, and has been involved with all phases of clinical trials in a wide spectrum of therapeutic areas. He also has extensive publications. His publication includes five reference books in statistical methods for the biopharmaceutical industry (Marcel Dekker) and *Design and Analysis of Clinical Trials* (co-authored with S.C. Chow; John Wiley & Sons), over 120 articles in peer-reviewed statistical journals, and over 20 invited book chapters in the *Encyclopedia of Biostatistics* and *Encyclopedia of Biopharmaceutical Statistics*. His papers were cited in the 2001 FDA guidance on *Statistical Approach to Establishing Bioequivalence* and in the 2005 WHO guideline on multisource drug products. He is an associate editor for the *Journal of Biopharmaceutical Statistics*, an associate editor for the *Taiwan Journal of Public Health* (since January 2002), and a member of the advisory committee on *Statistics in Biosciences* (since January 2008). He is one of the editors for the CRC Press biostatistics series. He serves as a referee for many international statistics journals. Currently, he is on the Drug Consultation Committee, the Committee for Medical Devices, and the Committee for the Chinese Drug and Pharmacy, all for the Department of Health, Taiwan. Dr. Liu received his BS in agronomy and MS in biometry from the National Taiwan University, Taipei, Taiwan and his MS and PhD in Statistics from the University of Kentucky, Lexington.

Part I

Preliminaries

Chapter 1

Introduction

1.1 History of Bioavailability Studies

The term *bioavailability* is a contraction for biological availability (Metzler and Huang, 1983). The definition of bioavailability has evolved over time with different meanings by different individuals and organizations. For example, differences are evident in the definitions by Academy of Pharmaceutical Sciences, the Office of Technology Assessment (OTA, 1974) of the Congress of the United States (1974), Wagner (1975), and the 1984 Drug Price Competition and Patent Restoration amendments to the Food, Drug, and Cosmetic Act. Throughout this book, however, the definitions and some related terms regarding bioavailability provided in the Act, which are adopted by the United States Food and Drug Administration (FDA), will be used (21 CFR, Part 320.1, 1983).

The bioavailability of a drug is defined as the rate and extent to which the active drug ingredient or active moiety from a drug product is absorbed and becomes available at the site of drug action. For drug products that are not intended to be absorbed into bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety is absorbed and becomes available at the site of action. A comparative bioavailability study refers to the comparison of bioavailabilities of different formulations of the same drug or different drug products. When two formulations of the same drug or two drug products are claimed *bioequivalent*, it is assumed that they will provide the same therapeutic effect or that they are therapeutically equivalent and they can be used interchangeably. Two drug products are considered pharmaceutical equivalents if they contain identical amounts of the same active ingredient. Two drugs are identified as pharmaceutical alternatives to each other if both contain an identical therapeutic moiety, but not necessarily in the same amount or dosage form or as the same salt or ester. Two drug products are said to be bioequivalent if they are pharmaceutical equivalents (i.e., similar dosage forms made, perhaps, by different manufacturers) or pharmaceutical alternatives (i.e., different dosage forms) and if their rates and extents of absorption do not show a significant difference to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives become available at the site of action when administered at the same

molar dose under similar conditions in an appropriately designed study. For more discussion regarding the definition of bioavailability, see Balant (1991) and Chen et al. (2001a,b).

The study of absorption of an exogenously administered compound (sodium iodide) can be traced back to 1912 (Wagner, 1971). The concept of bioavailability, however, was not introduced until some 30 years later. Oser et al. (1945) studied the relative absorption of vitamins from pharmaceutical products and referred to such relative absorption as physiological bioavailability. In recent years, generic drug products, which are those manufactured by generic drug companies or the innovator companies themselves, have become very popular. Bioavailability/bioequivalence studies are of particular interest to the innovator and the generic drug companies in the following ways. First, for the approval of a generic drug product, the FDA usually does not require a regular new drug application (NDA) submission, which demonstrates the efficacy, safety, and benefit–risk of the drug product, if the generic drug companies can provide the evidence of bioequivalence between the generic drug products and the innovator drug product through bioavailability and bioequivalence studies in a so-called abbreviated new drug application (ANDA). Second, when a new formulation of a drug product is developed, the FDA requires that a bioavailability study be conducted to assess its bioequivalence to the standard (or reference) marketed formulation of the drug product. Thus, bioavailability studies are important because an NDA submission includes the results from phases 1–3 clinical trials, which are very time consuming and costly to obtain. Finally, under the Food and Drug Administration Modernization Act (FDAMA) passed by the U.S. Congress in 1997, after the approval, depending on the magnitudes of changes in components and composition or method of manufacture, the FDA may require the evidence of bioequivalence between the pre- and postchange products under NDA or postchange generic product with the reference list product under ANDA. For details, see the FDA guidance on scale-up and postapproval changes (FDA, 1995a; FDA, 1997a).

The concept of bioavailability and bioequivalence became a public issue in the late 1960s because of the concern that a generic drug product might not be as bioavailable as that manufactured by the innovator. These concerns rose from clinical observations in humans together with the ability to quantify minute quantities of drug in biological fluids. This initiated not only a period of four decades of extremely active scientific research and development in bioavailability and bioequivalence, but also started the process and formulation of the current regulatory requirements for approval of generic drug products. Spanning from the early 1970s to date, the research and development of bioavailability and bioequivalence can be roughly divided into four phases. The first phase is from early 1970s to 1984 when the U.S. Congress passed the Drug Price Competition and Patent Term Restoration Act that authorized the to approve generic drug products through bioavailability and bioequivalence studies. The second phase begins from 1984 to 1992 after the issue of the U.S. FDA guidance entitled *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design in 1992*, which provides the sponsors a guidance as to how the data

should be analyzed and presented in an ANDA submission for bioequivalence review. The concept of population and individual bioequivalence for addressing drug interchangeability in terms of drug prescribability and drug switchability and their corresponding statistical methods has been discussed in the third phase since 1992. The fourth phase starts at the dawn of the twenty-first century when based on the fruit of research conducted in the last 30 years of the twentieth century, the FDA issued and implemented the new guidance on general considerations and statistical approaches to bioavailability and bioequivalence studies.

In 1970, the FDA began to ask for evidence of biological availability in applications submitted for approval of certain new drugs. In 1971, a drug bioequivalence study panel was formed by the OTA to examine the relationship between the chemical and therapeutic equivalence of drug products. On the basis of the recommendations in the OTA report, the FDA published a set of regulations for the submission of bioavailability data in certain new drug applications. These regulations became effective on July 1, 1977 and are currently codified in 21 CFR, Part 320. In 1971, by the time the FDA began to require evidence of bioavailability for NDA of some drug products, the Biopharmaceutical Subsection of the American Statistical Association simultaneously formed a Bioavailability Committee to investigate the statistical components for the assessment of bioequivalence. Metzler (1974) summarized the efforts by the Committee and addressed several concerns about some statistical issues in bioavailability studies. During the decade of the early 1980s, the search for statistical methods for the assessment of bioequivalence received tremendous attention. Several methods that met the FDA requirements for statistical evidence of bioequivalence were proposed. These methods included an a posteriori power approach, reformation of bioequivalence hypotheses (Schuirmann, 1981; Anderson and Hauck, 1983), a confidence interval approach (Westlake, 1972, 1976, 1979; Metzler, 1974), and a Bayesian approach (Rodda and Davis, 1980; Mandallaz and Mau, 1981). A detailed discussion of these statistical developments during this period can be found in Metzler and Huang (1983).

In 1984, the FDA was authorized to approve generic drug products under the Drug Price Competition and Patent Term Restoration Act. However, as more generic products become available, the following concerns were raised:

1. Whether generic drug products are comparable in quality to the innovator drug product.

2. Whether the generic copies of innovator drug products have comparable therapeutic effect.

To address these concerns, a hearing on bioequivalence of solid oral dosage forms was conducted by the FDA during September 29–October 1, 1986 in Washington, DC. As a consequence of the hearing, a bioequivalence task force was formed to examine the current procedures adapted by the FDA for the assessment of bioequivalence between immediate solid oral dosage forms. Some efforts were also directed at investigating the statistical issues that often occur in various stages of design and data analysis in bioavailability and bioequivalence studies. A report from the

bioequivalence task force was released in January 1988. Several statistical issues related to the assessment of bioequivalence are summarized below:

1. Lot-to-lot uniformity
2. Alternative statistical designs for intra-subject variability
3. Statistical methodology in decisional criteria for bioequivalence
4. Product-to-product variability
5. Detection and treatment of outlying data.

Relative to these issues, several statistical methods have been developed that provide some answers to the above statistical questions. For example, for the evaluation of lot-to-lot (or batch-to-batch) uniformity, Chow and Shao (1989) proposed several statistical tests for batch-to-batch variability. To account for the heterogeneity of intra-subject variability, an estimation procedure for the assessment of intra-subject variability, assuming that the coefficient of variation (CV) is the same from subject to subject, was proposed by Chow and Tse (1990b) under a conditional random effects model. Chow (1989) and Chow and Shao (1991) compared the decision rules under lognormality assumption. Chow and Shao (1990) also proposed an alternative approach for assessing bioequivalence using the idea of a confidence region. The proposed procedure was shown to rigorously meet the FDA's requirements for average bioequivalence. For outlier detection, Chow and Tse (1990a) proposed two tests using the idea of likelihood distance and estimates distance for detection of a possible outlying subject. The same problem was also examined by Liu and Weng (1991).

In June, 1992, the first edition of this book was published, which provides a comprehensive and unified summarization of literature on statistical design and analysis of bioavailability and bioequivalence up to 1991. Two weeks after the first edition of this book was published, the FDA guidance on *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design* was issued. Chow and Liu were invited by the Division of Biometrics and Division of Bioequivalence of the FDA to give a presentation on the review of the guidance from the pharmaceutical perspectives in April, 1983 (Chow and Liu, 1994a,b). As a follow-up, Chow and Liu also organized a special invited paper session on the FDA guidance at the 1993 American Statistical Association (ASA) Joint Statistical Meetings held in San Francisco, California on August 9, 1993. At the invited paper session, various issues concerning the 1992 guidance were discussed. Other details on the discussion of the issue of bioequivalence during the second phase can be found in a supplement entitled *Bioequivalence Assessment: Methods and Applications of* the *International Journal of Clinical Pharmacology Therapy and Toxicology* edited by V.W. Steinijans and H.U. Shulz (Vol. 30, Suppl. 1, 1992).

As the century closes to the end, generic drug products have played more important role in health care than before because of the necessity for reduction of health cost by all countries. As a result, in the third phase after 1992, different concepts of bioequivalence, such as drug interchangeability including drug prescribability

and drug switchability (Hauck and Anderson, 1992; Chow and Liu, 1995a), have evolved, with suggestions of different requirements for approval of generic drugs: for example, population bioequivalence (Liu and Chow, 1992b) and individual bioequivalence (Hauck and Anderson, 1992). At the same time, several important symposia and workshops were held to discuss and exchange ideas and views in definition and procedures of population and individual bioequivalence between drug products. To enhance communications and to exchange information of the state-of-the-art scientific developments and advancements among regulatory agencies, academia, and the pharmaceutical industry, on September 19–20, 1994, the authors of this book, along with the experts from the FDA organized a major symposium on Bioavailability and Bioequivalence, which was sponsored by the Drug Information Association (DIA) held in Rockville, Maryland. The papers presented in this symposium were published in a special issue of the *Drug Information Journal*, edited by S.C. Chow (Vol. 29, No. 3) (see, Chow, 1995). In addition, L. Endrenyi of the University of Toronto, J. Mau of the Heinrich Heine University, Düsseldorf Medical Institutions, and R. Williams of the FDA held an international workshop on statistical and regulatory issues on the assessment of bioequivalence in Düsseldorf, Germany, October 19–20, 1995, to address current regulatory viewpoints and unsolved scientific issues on bioequivalence. The papers presented at this workshop were published in a special issue of the *Journal of Biopharmaceutical Statistics*, edited by S.C. Chow (Vol. 7, No. 1). Furthermore, Fédération Internationale Pharmaceutique (FIP) held its Bio-International'96 Conference in Tokyo, Japan, April 22–24, 1996 to address various issues of bioequivalence including highly variable drug products, individual bioequivalence, alternative metrics and approaches, and the role of *in vitro* dissolution test. The papers presented at this conference were published in the proceedings of FIP Bio-International, 1996 *Bioavailability, Bioequivalence and Pharmacokinetics Studies*, edited by K.K. Midha and T. Nagai, Tokyo, Japan.

In late October 1997, the FDA circulated a draft guidance entitled *In Vivo Bioequivalence Studies Based on Population and Individual Bioequivalence Approaches* for comments. According to the FDA, the draft guidance is not for implementation at that time. However, when it is finalized, it will replace the 1992 guidance for bioequivalence assessment. This new draft guidance requires the sponsors to provide evidence of individual bioequivalence for approval of generic drugs as well as innovator drug products for which postapproval changes are required in bioequivalence testing as specified in the scale-up and postapproval change (SUPAC) guidelines.

Note that since this new draft guidance will have a great influence on the design and analysis of bioequivalence studies, several professional (statistical) meetings were organized to (1) review the guidance, (2) evaluate the feasibility and scientific merits of individual bioequivalence for approval of generic products, (3) exchange ideas for improvement of the recommended statistical procedures, and (4) discuss the strategy for future implementation of the guidance from the perspectives of the academia, the pharmaceutical industry, and regulatory agencies around the world. Just to name a few, professional meetings held in 1998 including Midwest Biopharmaceutical Statistics Workshop (MBSW), Muncie, Indiana; Drug Information Association (DIA) annual meeting, Boston, Massachusetts; the American Statistical

Association (ASA) annual Joint Statistical Meetings, Dallas, Texas; Statisticians in the Pharmaceutical Industry (PSI) annual meeting, Horragate, the United Kingdom; and the International Biometric Conference (IBC), Cape Town, South Africa. Most comments from these professional meetings indicated that the draft guidance in its current content and format lacks clinical and statistical considerations (Chow, 1999). In addition, since many important scientific and methodological issues still remain unresolved, Chow and Liu assembled a special issue on individual bioequivalence for *Statistics in Medicine*, which was published on October 30, 2000. In this special issue, 13 papers by the authors from academia, government as well as industry were published for various opinions and comments and different scientific approaches to individual bioequivalence. On the basis of this issue and other research on the merits, feasibility, methodology of population and individual bioequivalence, on February 2, 2001, the FDA issued the guidance entitled *Statistical Approaches to Establishing Bioequivalence*. Six months later, the European Agency for the Evaluation of Medicinal Products (EMEA) issued the *Note for Guidance on the Investigation of Bioavailability and Bioequivalence*. On March 19, 2003, the FDA issued the current guidance entitled *Bioavailability and Bioequivalence Studies for Orally Administrated Drug Product–General Considerations*. Recently, the World Health Organization (WHO) recognizes that the quality and supply of generic drugs is a critical issue of global health and are vital to developing countries. As a result, after a series of meetings among the members of FIP/WHO BCS Task, for multisource (generic) pharmaceutical products, in 2005, the WHO issued the draft revision of the guidelines on *Registration Requirements to Establish Interchangeability*. Research on methodological development for bioequivalence assessment is still very active in the twenty-first century. More details will be provided in later chapters of this book.

1.2 Formulations and Routes of Administration

When a drug is administered to a human subject, the drug generally passes through an absorption phase, distribution phase, metabolism phase, and finally an elimination phase within the body. As mentioned in Section 1.1, the bioavailability of a drug is defined as the rate and extent to which the active ingredient of the drug is absorbed and becomes available to the body. Because clinical effects may be associated with blood or plasma levels of the drug, the information of bioavailability is useful for the assessment of a drug's efficacy and safety. Bioavailability is usually determined by some pharmacokinetic measurements that can be estimated from the blood or plasma concentration–time curve obtained following drug administration. The blood or plasma concentration–time curve, however, is dependent, in part, on the dosage form and the route of administration.

In the pharmaceutical industry, when a new drug is discovered, it is important to design an appropriate dosage form for the drug so that it can be delivered to the body efficiently for optimal therapeutic effect. The dosage form, however, should also account for the acceptability to the patients. Dosage forms, such as tablet, capsule,

solution, powder, and liquid suspension, are usually considered. For a given drug product, several dosage forms may be designed for different purposes. For example, solution and liquid suspension dosage forms may be more appropriate than solid dosage forms for children and elderly patients. However, in practice, most drugs are taken orally in solid dosage forms (e.g., tablet and capsule). Generally, solid dosage forms have to dissolve to be absorbed. The dissolution of the drug depends on the particle size. The reduction of particle size may increase the bioavailability of the drug. Examples of drugs for which bioavailability has been increased as a result of particle size reduction are aspirin and estradiol (Dare, 1964).

The route of administration can certainly affect the bioavailability of a drug. Different routes of administration may result in a significant difference in bioavailability. For example, a study of kanamycin (Kunin, 1966) demonstrated that the oral administration has extremely low bioavailability (about 0.7%). In contrast, the bioavailability of intramuscularly administered kanamycin is much greater (about 40%–80%). Basically, there are several routes by which drugs are commonly administered. These routes may be classified as either intravascular or extravascular. Intravascular administration refers to giving the drug directly into the blood, either intravenously or intra-arterially. Extravascular administration includes the oral, intramuscular, subcutaneous, sublingual, buccal, pulmonary, rectal, vaginal, and transdermal routes. Drugs administered extravascularly must be absorbed to enter the blood.

Because different dosage forms may affect the bioavailability of the drug, they may exhibit market variability in their absorption. Thus, before a drug can be released for medical use, the FDA requires that the drug be tested *in vitro* in compliance with United States Pharmacopeia and National Formulary (USP/NF) specifications to ensure that the drug contains the labeled active ingredient within an acceptable variation. The USP/NF standards for the evaluation of the drug include potency testing, content uniformity testing, dissolution testing, disintegration testing, and weight variation testing (Chow and Liu, 1995b). In addition, a bioavailability study is also required by the FDA. The assay method used for the active ingredients to quantify the drug must be validated in terms of the closeness of the test results obtained from the assay method to the true values (accuracy) and the degree of closeness of the test results to the true values (precision).

Note that since different dosage forms or routes of administration may affect the bioavailability of the drug, a comparative bioavailability (bioequivalence) study may involve the comparison of different dosage forms (or formulations) of the same drug, generic drug product, and the marketed (innovator) drug product of the same active ingredient, and different routes of administration.

1.3 Pharmacokinetic Parameters

In a comparative bioavailability study in humans, following the administration of a drug, the blood, serum, or plasma concentration–time curve is often used to study the rate of absorption and elimination of the drug which can be characterized by

taking blood samples immediately before and at various time points after drug administration. However, instead of direct and indirect pharmacokinetic measures, the 2003 FDA guidance recommends that reliance on systemic exposure measures reflects comparable rate and extent of absorption. These exposure measures are defined relative to early, peak, and total portion of the plasma, serum, or blood concentration–time curve. The pharmacokinetic parameters representing different exposure measures involve the area under the plasma or blood concentration–time curve (AUC) for total exposure, partial AUC for early exposure, maximum or peak concentration ($C_{\max}$), and time to achieve maximum concentration ($t_{\max}$) for peak exposure, respectively. The measurements of these pharmacokinetic parameters can be derived either directly from the observed blood or plasma concentration–time curve, which is independent of a model, or is obtained by fitting the observed concentrations to a one- or a multicompartment pharmacokinetic model. In the following case, the determination of some pharmacokinetic parameters assumes first-order absorption and elimination.

One of the primary pharmacokinetic parameters for total exposure in a bioavailability study is the AUC. The AUC is often used to measure the extent of absorption or total amount of drug absorbed in the body. Several methods exist for estimating the AUC from zero time until time t, at which the last blood sample is taken. These methods include the interpolation using the trapezoidal rule, the Lagrange and spline methods, the use of a planimeter, the use of digital computers, and the physical method that compares the weight of a paper corresponding to the area under the experimental curve to the weight of a paper of known area. Among these methods, the method of interpolation appears to be the one most commonly used. Yeh and Kwan (1978) discussed the advantages and disadvantages of using the Lagrange and spline methods relative to the trapezoidal rule in the method of interpolation. For simplicity, we introduce only the method of linear interpolation using the trapezoidal rule. Let $C_0, C_1, \ldots, C_k$ be the plasma or blood concentrations obtained at time $0, t_1, \ldots, t_k$, respectively. The AUC from zero to t_k, denoted by $\mathrm{AUC}(0 - t_k)$, is obtained by

$$\mathrm{AUC}(0 - t_k) = \sum_{i=2}^{k} \left(\frac{C_{i-1} + C_i}{2} \right) (t_i - t_{i-1}). \tag{1.3.1}$$

The AUC, however, should be calculated from zero to infinity, not just to the time of the last blood sample, as is so often done. The portion of the remaining area from t_k to infinity could be large if the blood level at t_k is substantial (Martinez and Jackson, 1991). The AUC from zero to infinity, denoted by $\mathrm{AUC}(0 - \infty)$, can be estimated as follows (Rowland and Tozer, 1980):

$$\mathrm{AUC}(0 - \infty) = \mathrm{AUC}(0 - t_k) + \frac{C_k}{\lambda}, \tag{1.3.2}$$

where

C_k is the concentration at the last measured sample after drug administration

λ is the terminal or elimination rate constant, which can be estimated as the slope of the terminal portion of the log concentration–time curve multiplied by -2.303

The FDA regulation requires that sampling be continued through at least three more terminal half-lives of the active drug ingredient or therapeutic moiety, or its metabolites, measured in the blood or the decay of the acute pharmacological effect so that the elimination would have been completed and any remaining area beyond time t_k is negligible. Therefore, the FDA recommends that at least three to four samples should be obtained during the terminal log–linear phase to get an accurate estimate of λ from linear regression. Note that a few missing values or unexpected observations in the plasma concentration–time curve within (t_1, t_k) will generally have little effect on the calculations of AUC$(0 - t_k)$ and AUC$(0 - \infty)$. However, if there are many missing values or unexpected observations in the plasma concentration–time curve, especially at endpoints (i.e., t_1 and t_k), the bias of the estimate of AUC could be substantial.

In addition of the AUC, the absorption rate constant is usually studied during the absorption phase. Under the single-compartment model, the absorption rate constant can be estimated based on the following equation using the method of residuals (Gibaldi and Perrier, 1982).

$$C_t = \frac{k_a F D_0}{V(k_a - k_e)}(e^{-k_e t} - e^{-k_a t}), \tag{1.3.3}$$

where

k_a and k_e are the absorption and elimination rate constants, respectively
D_0 is the dose administered
V is the volume of distribution
F is the fraction of the dose that reaches the systemic circulation

Given Equation 1.3.3, C_{max} and t_{max} can similarly be obtained as follows:

$$t_{max} = \frac{2.303}{k_a - k_e} \log\left(\frac{k_a}{k_e}\right), \tag{1.3.4}$$

and

$$C_{max} = \frac{k_a F D_0}{V(k_a - k_e)}(e^{k_e t_{max}} - e^{k_a t_{max}}). \tag{1.3.5}$$

In practice, however, the estimates from a pharmacokinetic model usually are not used for the comparison of formulations. Thus, C_{max} is estimated directly from the observed concentrations. That is, $C_{max} = \max\{C_0, C_1, \ldots, C_k\}$. Similarly, t_{max} is estimated as the corresponding time point at which the C_{max} occurs. Because the partial AUC is an early exposure measure, the FDA suggests that the partial AUC be truncated at the population median of t_{max} and at least two quantifiable samples be collected before the expected C_{max} to allow estimation of the partial AUC.

During the elimination phase, the pharmacokinetic parameters that are often studied are the elimination half-life ($t_{1/2}$) and rate constant (k_e) (Chen and Pelsor, 1991). The plasma elimination half-life is the time taken for the plasma concentration

to fall by one-half. Assume that the decline in plasma concentration is of first order, the $t_{1/2}$ can be obtained by considering

$$\log D = \log D_0 - \frac{k_e t}{2.303}, \tag{1.3.6}$$

where D is the amount of drug in the body. Thus, at $D = D_0/2$ (i.e., $t = t_{1/2}$) we have

$$\log\left(\frac{1}{2}\right) = -\frac{k_e t_{1/2}}{2.303}.$$

Hence,

$$t_{1/2} = \frac{0.693}{k_e},$$

where k_e is given by

$$k_e = (-2.303)\left(\frac{d \log D}{dt}\right).$$

The first order elimination half-life is independent of the amount of drug in the body. In practice, all the drug may be regarded as having been eliminated (about 97%) by five half-lives.

The above pharmacokinetic parameters are usually considered in a single dose trial. In practice, drugs are most commonly prescribed to be taken on fixed time interval basis (i.e., multiple doses such as b.i.d., t.i.d., or q.i.d.). Dosing a drug several times a day can result in a different drug concentration profile than that produced by a single dose. If the dosing interval is less than the time required to eliminate the entire dose, the peak plasma level following the second and succeeding doses of a drug is always higher than the peak level after the first dose. This leads to drug accumulation in the body relative to the initial dose. For a multiple dose regimen, the amount of drug in the body is said to have reached a steady-state level if the amount or average concentration of the drug in the body remains stable. The following pharmacokinetic parameters at steady are usually studied:

$$\begin{aligned}
C_{max} &= \frac{D_0}{1 - \left(\frac{1}{2}\right)^{\varepsilon}}, \\
C_{min} &= \frac{D_0}{1 - \left(\frac{1}{2}\right)^{\varepsilon}} \left(\frac{1}{2}\right)^{\varepsilon}, \\
C_{av} &= \frac{C_{max} - C_{min}}{\log\left(\frac{C_{max}}{C_{min}}\right)}, \\
\text{Percent fluctuation} &= \left(\frac{C_{min}}{C_{max}}\right) \times 100\%,
\end{aligned} \tag{1.3.7}$$

where ε is the dosing interval τ divided by elimination half-lives. Note that τC_{av} is the area under the curve within a dosing interval at steady state, which is equal to that following a single dose. In a multiple dose study, however, how to choose or combine the information of several pairs of C_{max} and C_{min} from a subject is an interesting question. This certainly has some influence on statistical analysis of the data. Wang et al. (1996) studied patient compliance and fluctuation of the serum drug concentration.

Example 1.3.1

To illustrate how to estimate AUC, C_{max}, t_{max}, $t_{1/2}$, and k_e from the observed concentrations, it is helpful to consider the following example. Table 1.3.1 lists the primidone concentrations (μg/mL) versus time points (hours) from a subject over a 32 hours period after administered a 250-mg tablet of a drug. The blood samples were drawn immediately before and at time points 0.5, 1.0, 2, 3, 4, 6, 8, 12, 16, 24, and 32 hours. The plot of primidone concentration–time curve for the subject is exhibited in Figure 1.3.1. From Table 1.3.1, AUC$(0-32)$ and C_{max} can be obtained as follows.

$$\begin{aligned}
\text{AUC}(0-32) &= \sum_{i=2}^{12}\left[\frac{C_{i-1}+C_i}{2}\right](t_i - t_{i-1}) \\
&= \frac{(0+0)}{2}(0.5-0) + \frac{(0+2.8)}{2}(1-0.5) + \cdots \\
&\quad + \frac{(2+1.6)}{2}(32-24) \\
&= 85.95 \ (\mu\text{g} \times \text{h/mL}), \\
C_{max} &= \max(0,0,2.8,\ldots,1.6) = 4.7 \ \mu\text{g/mL}.
\end{aligned}$$

TABLE 1.3.1: Calculation of AUC using the trapezoidal rule.

Blood Sample (i)	t_i	C_i	$(C_i + C_{i-1})/2$	$t_i - t_{i-1}$	$(C_i + C_{i-1})(t_i - t_{i-1})/2$
1	0.0	0.0	—	—	—
2	0.5	0.0	0.00	0.5	0.00
3	1.0	2.8	1.40	0.5	0.70
4	1.5	4.4	3.60	0.5	1.80
5	2.0	4.4	4.40	0.5	2.20
6	3.0	4.7	4.55	1.0	4.55
7	4.0	4.1	4.40	1.0	4.40
8	6.0	4.0	4.05	2.0	8.10
9	8.0	3.6	3.80	2.0	7.60
10	12.0	3.0	3.30	4.0	13.20
11	16.0	2.5	2.75	4.0	11.00
12	24.0	2.0	2.25	8.0	18.00
13	32.0	1.6	1.80	8.0	14.40

Note: AUC$(0-30) = 85.95$.

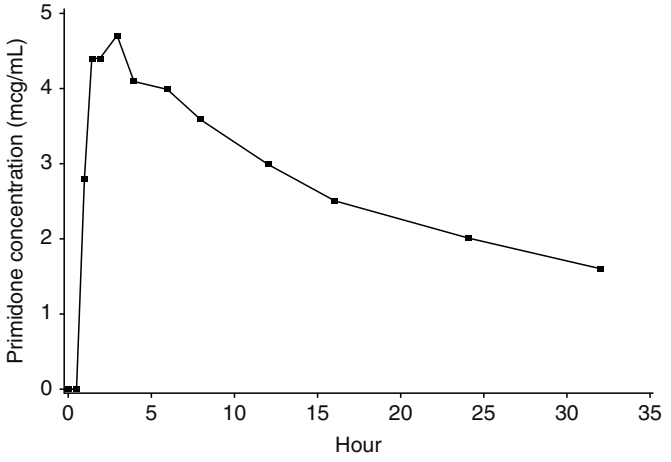


FIGURE 1.3.1: Primidone concentration–time curve.

t_{max} is estimated as the corresponding time point at which C_{max} was achieved. Thus, $t_{max} = 3.0$ hours. For the estimation of the elimination rate k_e, the last seven concentrations during the elimination phase were used to fit a linear regression based on the log concentrations with a base 10 using the least-squares method (Draper and Smith, 1981). The resultant regression line is given by

$$\log_{10}(C_i) = 0.6713 - 0.01518t_i.$$

Thus, the elimination rate is

$$k_e = (-2.303)\ (-0.01518) = 0.03496\ (\text{h}^{-1}).$$

Consequently, the elimination half-life is

$$t_{1/2} = \frac{0.693}{0.03496} = 19.8\ (\text{h}).$$

The AUC$(0-\infty)$ can be obtained as

$$\begin{aligned}\text{AUC}(0-\infty) &= \text{AUC}(0-32) + C_{32}/0.03496\\ &= 85.95 + 1.6/0.03496\\ &= 131.72\ (\mu\text{g}\cdot\text{h/mL}).\end{aligned}$$

In the above example, we selected the last seven concentrations during the elimination phase to calculate the elimination rate. In practice, the number of concentrations used may depend on the plasma concentration–time curve for each subject. This is an interesting statistical question that needs further attention.

1.4 Clinically Important Differences

The definition of a clinically significant difference is important for the assessment of therapeutic equivalence in terms of efficacy, safety, and benefit–risk ratio. In bioavailability and bioequivalence studies, it is our intention to consider bioequivalence in terms of therapeutic equivalence. However, this ultimate assumption of bioequivalence can be verified only through rigorous prospective clinical trials that may relate bioavailability parameters with clinical endpoints through the data from blood concentrations and clinical efficacy and safety evaluations. In practice, such clinical trials are rarely carried out owing to the following difficulties:

1. Unlike healthy subjects who are often used for bioavailability–bioequivalence studies, patients cannot be well controlled.

2. Patients are more heterogeneous in a wide variety of characteristics.

However, the ultimate obstacle lies in the estimation and translation of the differences in bioavailability into the therapeutic differences of interest.

Westlake (1979) pointed out that a statistically significant difference in the comparison of bioavailability between drug products does not necessarily imply that there is a clinically significant difference between drug products. For example, the AUC for the test product may exhibit an 80% bioavailability compared with the reference product. The 20% difference in AUC, which may be statistically significant, however, may not be of clinical significance in terms of therapeutic effect. In other words, although there is a 20% difference, both test and reference products can still reach the same therapeutic effect. Thus, they should be considered therapeutically equivalent. Generally, a set of bioequivalence limits, say (a, b), is given for the evaluation of clinical difference. If the difference (usually in percentage) in AUC between the test and reference products is within the limits, then there is no clinical difference, or they are considered to be therapeutically equivalent. Bioequivalent limits for therapeutic equivalence generally depend on the nature of the drug, targeted patient population, and clinical endpoints (efficacy and safety parameters) for the assessment of therapeutic effect. For example, for some drugs, such as topical antifungals or vaginal antifungals, that may not be absorbed in blood (Huque and Dubey, 1990), the FDA proposed some equivalent limits for some clinical endpoints (binary responses), such as cure rate as in Table 14.4.1. This table indicates that if the cure rate for the reference drug is greater than 95%, then a difference in cure rate within 5% is not considered a clinically important difference (see Table 1.4.1).

TABLE 1.4.1: Equivalence limits for binary responses.

Equivalence Limits (%)	Response Rate for the Reference Drug (%)
±20	50–80
±15	80–90
±10	90–95
±5	>95

1.5 Assessment of Bioequivalence

The assessment of bioequivalence for different drug products is based on the following fundamental bioequivalence assumption: When two drug products are equivalent in the rate and extent to which the active drug ingredient or therapeutic moiety is absorbed and becomes available at the site of drug action, it is assumed that they will be therapeutically equivalent and can be used interchangeably.

Given the fundamental bioequivalence assumption, bioequivalence studies are, therefore, the surrogates for clinical trials for assessment of therapeutic equivalence in efficacy and safety between drug products. This is the reason why the title of the WHO guidelines is on the requirements for establishment of interchangeability of multisource pharmaceutical products. The purpose of bioequivalence trials, hence, is to identify pharmaceutical equivalents or pharmaceutical alternatives that are intended to be used interchangeably for the same therapeutic effect (21 CFR, 320.50). Thus, bioequivalent drug products are therapeutic equivalents and can be used interchangeably. As a result, US FDA was authorized to ask the sponsors, through an ANDA, to provide the evidence of bioequivalence for approval of generic copies of an innovator drug product after the patent has expired under the Drug Price Competition and Patent Term Restoration Act passed by the U.S. Congress in 1984.

As indicated in Hauck and Anderson (1992) and Chow and Liu (1995a), drug interchangeability can be classified as either drug *prescribability* or drug *switchability*. Drug prescribability is referred to as the physician's choice for prescribing an appropriate drug product for his or her new patients among an innovator drug product and a number of its generic copies that have been shown to be bioequivalent to the innovator drug product. Drug prescribability is usually assessed by population bioequivalence (Chow and Liu, 1992). On the other hand, drug switchability (Anderson, 1993; Liu and Chow, 1995) is related to the switch from an innovator drug product to a generic product within the same subject whose concentration of the active ingredients has been titrated to a steady, efficacious, and safe level. To assure drug switchability, it is recommended that bioequivalence be assessed within individual subjects.

Once the fundamental assumption and the purpose of bioequivalence trials are clearly defined and understood, the next question is what and how to assess bioequivalence. The essential pharmacokinetic parameters for systematic exposure in the FDA regulations for an *in vivo* bioavailability study are $\mathrm{AUC}(0-\infty)$, $C_{\max}$, λ, and $t_{1/2}$ of the therapeutic moiety. As discussed in Section 1.3, these pharmacokinetic parameters can be derived either directly from the observe blood or plasma concentration–time curve or obtained by fitting the observed concentrations to a one- or multicompartment pharmacokinetic model. In general, the use of the observed $\mathrm{AUC}(0-\infty)$, $C_{\max}$, or $t_{\max}$ from the blood or plasma concentration–time curve is preferred, for they provide the essential information about the pharmacokinetic characteristics in assessment of bioequivalence, and are model-independent and easy to calculate. However, there are some drawbacks in these estimates. For example, the predetermined sampling time points are often too few to have reliable estimates on $C_{\max}$ and $t_{\max}$ in most bioavailability studies. Consequently, the distribution of the estimated $t_{\max}$ is not continuous,

but rather, discrete. On the other hand, when a pharmacokinetic model is considered, the goodness of fit of the model should be performed by examining the residuals. In practice, it is almost impossible to fit the same theoretical model for each subject in the study. Moreover the sampling time points are too few to provide reliable estimates for the pharmacokinetic parameters under the model, even though, theoretically, the assumed pharmacokinetic model may adequately describe the observed blood or plasma concentration–time curve. Therefore, the 2003 FDA guidance of general considerations for bioequivalence studies suggests that 12 to 18 samples, including a predose sample, be collected per subject per dose. This sampling can continue for at least three or more terminal half-lives of the drug. In addition, the sampling time should be spaced in such a way that $C_{\max}$ and λ can be estimated accurately.

The statistical concept for evaluation of bioequivalence lies with investigation of the closeness between the marginal distributions of pharmacokinetic responses of interest from the two drug products. As a result, there are three types of bioequivalence, namely, *average bioequivalence* (ABE), *population bioequivalence* (PBE), and *individual bioequivalence* (IBE). On the basis of the fact that the distribution of some random variables (e.g., normal random variable) is uniquely determined by its moments, the equivalence between two distributions can be assessed through the moments of the marginal distributions of the test and reference formulations. The first two moments of the distribution reflect the average and the variability of the distribution. The comparison of the first moments of the distributions of the pharmacokinetic parameters [say, $\mathrm{AUC}(0-\infty)$] for the two drug products refers to the comparison of *average bioavailability*, whereas the comparison between the second moments refers to the *variability of bioavailability*. To provide a better understanding of average bioavailability and variability in bioavailability, equivalence in averages and variabilities are illustrated in Figures 1.5.1 through 1.5.3. For example,

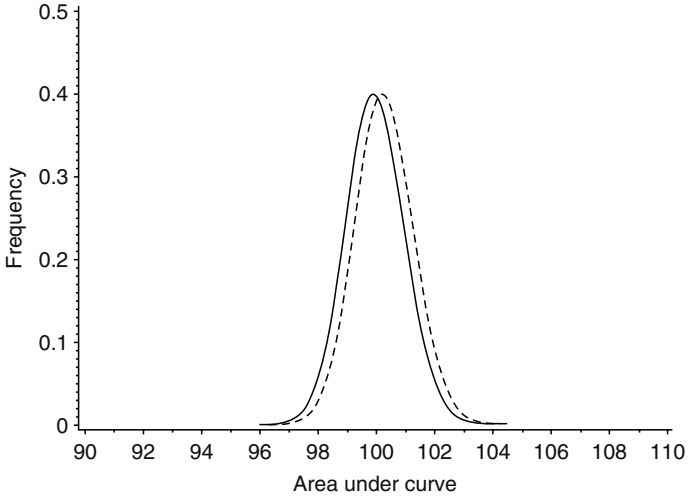


FIGURE 1.5.1: Equivalence in both means and variabilities.

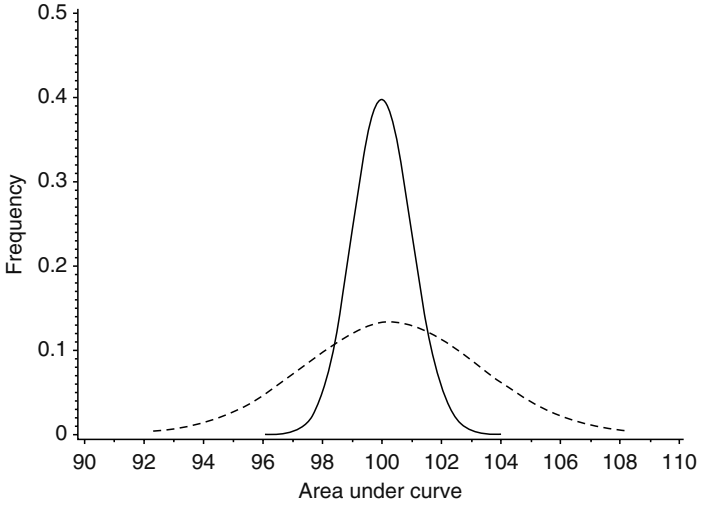


FIGURE 1.5.2: Equivalence in means, but not in variabilities.

if the distribution for AUC$(0-\infty)$ is normal and if the AUC$(0-\infty)$ of two products are equivalent in both averages and variabilities, then the two drug products are bioequivalent. As a result, to ensure drug prescribility, it is required to establish population bioequivalence which, in turn, dictates bioequivalence in both average and variability. However, in general, equivalence in the first two moments does not guarantee equivalence between formulations. Average bioequivalence, a part of population bioequivalence is referred to as equivalence in averages of the marginal distributions of the bioavailabilities between drug products. Currently, the regulations of

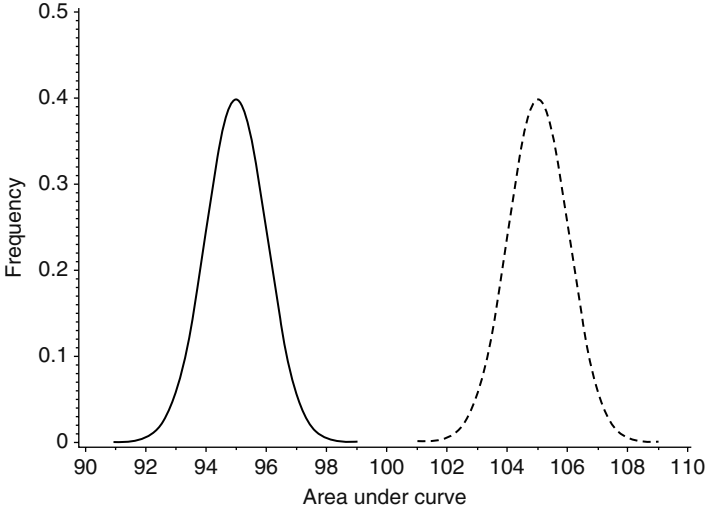


FIGURE 1.5.3: Equivalence in variabilities, but not in means.

most countries including the United States, European Union (EU), and Japan require only that evidence of average bioavailability be provided for approval of generic drug products (FDA, 2003b). However, in practice, this regulation on average bioavailability does not guarantee that two drug products can be used interchangeably in terms of drug efficacy and safety, especially the interchangeability among generic copies of the same innovator drug product. Some discussions can be found in Cornell (1980), Metzler and Huang (1983), Liu (1991), Liu and Chow (1992b), Chow and Liu (1995a), and Chow and Shao (1999).

It has been recognized that drug switchability requires individual bioequivalence (Hauck and Anderson, 1992; Anderson, 1993; Chow and Liu, 1995a). For a given individual, the statistical concept of individual bioequivalence is to examine the closeness between the two marginal distributions of the pharmacokinetic responses that are obtained under the repeated administrations of the test and reference formulations from the same subject. Under the normality assumption, it is then necessary to establish equivalence in average and variability of the two marginal distributions for a given individual. Results of comparison between the test and reference formulations for each individual over a population of subjects can then be assembled for evaluation between the test and reference formulations, as illustrated in Figure 1.5.4. Because assessment of individual bioequivalence requests the comparison of the marginal distributions of bioavailability between the two drug

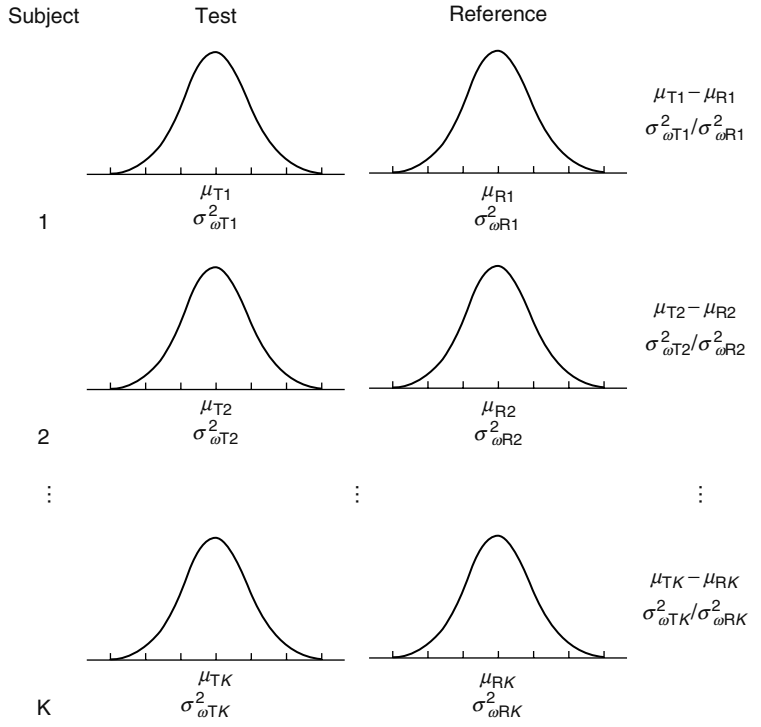


FIGURE 1.5.4: Concept of individual bioequivalence.

products within the same subject, then the replicated crossover designs are required to generate the multiple pharmacokinetic responses of the same formulations from an individual (Liu, 1995; Liu and Chow, 1995). The guidance on *Statistical Approaches to Establishing Bioequivalence* (FDA, 2001) provides the definitions and criteria for individual bioequivalence. The concept of individual bioequivalence is totally different from the usual average bioequivalence, and hence has a tremendous influence on statistical design, conduct, and analysis of bioequivalence trials. For more discussion, see Liu and Chow (1997a,b).

1.6 Decision Rules and Regulatory Aspects

1.6.1 Average Bioequivalence

The association between bioequivalence limits and clinical difference is difficult to assess in practice. The following decision rules were proposed by the FDA between 1977 and 2003 (Purich, 1980; FDA, 2003) for testing the bioequivalence in terms of average bioavailability of specific drugs, such as anticonvulsants, carbonic anhydrase inhibitors, and phenothiazines. Suppose AUC and $C_{\max}$ are the primary systematic exposure measures of the extent and rate of absorption. For each parameter, the following decision rules for assessment of average bioequivalence are applied.

1.6.1.1 75/75 Rule

Bioequivalence is claimed if at least 75% of individual subject ratios (relative individual bioavailability of the test formulation to the reference formulation) are within (75%, 125%) limits.

For the 75/75 rule, although it possesses some advantages, such as (1) it is easy to apply, (2) it compares the relative bioavailability within each subject, and (3) it removes the effect of heterogeneity of inter-subject variability from the comparison between the formulations, it is not viewed favorably by the FDA owing to some undesirable statistical properties. In a simulation study, Haynes (1981) showed that the 75/75 rule is very sensitive for drugs that have large inter- or intra-subject variabilities; even in the situation where the mean AUCs for the test and reference formulations are exactly the same. Metzler and Huang (1983), in another simulation study, also indicated that the 75/75 rule may reject as much as 56.3% of test products when the inter-subject variability is large. Thiyagarajan and Dobbins (1987) discussed the use of the 75/75 rule for assessment of bioequivalence. Chow (1989) and Chow and Shao (1991) provided an analytic evaluation of the 75/75 rule relative to the ± 20 rule. The results suggest that the 75/75 rule will never be met when the intra-subject variability is large (say 20%) for any given true ratio of means. For small variability (say 10%), only 61.3% of individual subject ratios will fall within (75%, 125%) limits when the true ratio of means is within 80% and 120% limits. Anderson and Hauck (1990) discussed the 75/75 rule and considered the use of individual subject ratios for assessment of individual bioequivalence.

1.6.1.2 80/20 Rule

If the average of the test product is not statistically significantly different from that of the reference product, and if there is at least 80% power for detection of a 20% difference of the reference average, then bioequivalence is concluded.

The 80/20 rule, which often requires that a study be large enough to provide at least 80% chance of correctly detecting a 20% difference in average bioavailability, is based on the concept of testing a hypothesis of equality for a single variable rather than equivalence. In the past three decades, however, hypothesis testing for the evaluation of bioequivalence has been questioned and was not encouraged. The 80/20 rule is considered only as a prestudy power calculation for sample size determination in the planning stage of study protocol.

1.6.1.3 ±20 Rule

Bioequivalence is concluded if the average bioavailability of the test formulation is within ±20% of that of the reference formulation with a certain assurance.

The ±20 rule, which allows a test formulation to exhibit up to a 20% variation in average bioavailability in comparison with a reference formulation, is commonly employed for most drug products. Levy (1986), however, indicated that the ±20 rule does not accommodate the effect that the 20% variation could have on the safety and efficacy of a specific drug. Another concern is the interchangeability of the formulations. As more generic products become available, the generic substitution for a brand name drug may involve the substitution of one generic product for another during a patient's therapy. Under the ±20 rule, interchanging the generic products can lead to a more than 20% difference from one to another. For example, substitution of a bioequivalent product providing 120% of the reference for a bioequivalent product providing 80% of the reference would result in an increase in the relative dose of 50%. In this case, the toxicity or efficacy is significantly magnified when bioequivalent products varying by as much as 50% are interchanged one for another. Thus, it is suggested that individualized drug-by-drug bioequivalence criteria (i.e., the acceptable degree of variation in bioavailability) be developed by the FDA at the time a generic product becomes eligible for approval.

On the basis of the report by the bioequivalence task force, the 75/75 rule is not required for the assessment of bioequivalence because it is not based on rigorous statistical tests. It appears that the ±20 rule was acceptable to the FDA for evaluation of average bioequivalence in early 1980. The 80/20 rule was recommended as the secondary analysis, which is often used as a supplement to the ±20 rule. However, frequently, the ±20 rule and the 80/20 rule may result in inconsistent conclusions. That is, the average bioequivalence is concluded based on the ±20 rule, but the power for detecting a 20% difference is far below 80% or vice versa. The possible causes of the inconsistency between the two decision rules were discussed by Chow and Shao (1991).

1.6.1.4 80/125 Rule

Bioequivalence is concluded if the average bioavailability of the test formulation is within (80%, 125%) that of the reference formulation, with a certain assurance.

From a multiplicative model for pharmacokinetic responses postulated by Westlake (1973, 1986), the logarithmic transformation is suggested for AUC $(0-\infty)$ or AUC$(0-t_{\text{last}})$ and $C_{\max}$ in the guidance (FDA, 2003). As a result, the Division of Bioequivalence, the FDA suggested use of an equivalence criterion of 80%–125% for assessment of bioequivalence based on the ratio of average bioavailability. This criterion is not symmetric about 1 on the original scale where the maximum probability of concluding average bioequivalence occurs. However, on the logarithmic scale, the criterion has a range of -0.2231 to 0.2231, which the symmetric about 0 where the probability of concluding average bioequivalence is at maximum. Had the criterion of the ± 20 rule been used for the logarithmic transformation, which is not linear, the maximum probability of concluding average bioequivalence would occur when the ratio of the average bioavailability is approximately about 0.98. The EMEA (2001) and WHO (2005) used the same equivalence criterion of 80%–125% for the log-transformed AUC$(0-\infty)$ or AUC$(0-t_{\text{last}})$ and $C_{\max}$. However, for $C_{\max}$, in certain cases, the EMEA and WHO allow a wider interval of 75%–133% for the ratio of average bioavailability to address any safety and efficacy concerns for patients switched between formulations. If a wider interval is used, it must be prespecified in the protocol.

It should be noted that bioequivalence determinations based on mean values do not account for the differences in inter- or intra-subject variabilities between formulations. Although the Pitman–Morgan test (Morgan, 1939; Pitman, 1939) is suggested for testing the equality of the variance between formulations, until recently, little or no attention in the literature has been given to address how much difference in variability, including inter- or intra-subject variabilities would be of clinical significance. In general, a much larger sample size is required for testing a difference in variances than that of testing for a difference in average. More details on the variability of bioavailability are given in Chapter 7.

1.6.2 Population and Individual Bioequivalence

Statistical evaluations for population/individual bioequivalence depend on different definitions of population/individual bioequivalence and their corresponding criteria. Basically, the criteria for evaluation of population/individual bioequivalence can be classified into the moment- and probability-based criteria, which are described below.

1.6.2.1 Moment-Based Criteria

The current moment-based criteria are based on the expected squared error loss in the form of the intra-subject difference of bioavailabilities in a subject who receives the test and reference formulations on two different occasions, and intra-subject variability which can be expressed as the expected squared error loss in the form of the intra-subject difference of bioavailabilities in a subject who receives the reference formulation of two different occasions. These moment-based criteria are then functions of difference in average bioavailability, the variability of the

subject-by-formulation interaction, and the ratio of the test intra-subject variability to the reference intra-subject variability. These three components, in fact, represent the three characteristics for quality assurance of a generic drug product, as compared with the approved reference product. As a result, individual bioequivalence can be assessed either by the aggregate or disaggregate moment-based criteria. The aggregate moment-based criteria are linear combinations of the three components, such that the decision-making process for conclusion of individual bioequivalence can be made by comparing the sample observed value of the combined criterion with some prespecified upper bioequivalence limit without consideration of the contributions made by individual components. For more details about the aggregate criteria, see Sheiner (1992), Schall and Luus (1993), Holder and Hsuan (1993), Chen (1996, 1997), Chen et al. (2000), Hyslop et al. (2000), Endrenyi et al. (2000). McNally et al. (2003), Chow et al. (2002a,b), Hsuan and Reeve (2003), and FDA (2001).

Liu and Chow (1996) and Chow (1999) suggested the use of the disaggregate criterion for which individual components must meet their respective prespecified limits to conclude individual bioequivalence. The disaggregate criteria are more intuitive appealing and appreciate contributions made by individual components. Vuorinen and Turunen (1996) and Vuorinen (1997) applied disaggregate criteria in a stepwise manner for assessment of average, population, and individual bioequivalence sequentially. Carrasco and Jover (2003) suggested using the structural equation model for assessment of individual bioequivalence in a disaggregate manner. Discussion on advantages and drawbacks of aggregate and disaggregate criteria can be found in Liu and Chow (1997a).

1.6.2.2 Probability-Based Criteria

The probability-based criteria are based on the probability that the intra-subject difference of bioavailabilities in a subject receiving the test and reference formulations on two different occasions is within some prespecified limits. Anderson and Hauck (1990) first introduced the individual equivalence ratio (IER) as a probability-based criterion for assessment of individual bioequivalence and proposed a nonparametric binomial test (TIER). Under the normality assumption, Liu and Chow (1997b) showed that the usual *t*-statistic for evaluation of average bioequivalence can also be used for TIER, but with different critical values from noncentral *t*-distribution. Chinchilli and Esinhart (1994) also suggested that, under the normality assumption, the concept of tolerance interval be applied for assessment of individual bioequivalence. Schall and Luus (1993) and Schall (1995) suggest that the probability, based on the intra-subject difference between test and reference formulations, should be compared with that based on the intra-subject difference in bioavailabilities in the same subject who receives the reference formulations on two different occasions. Other probability-based criteria and their procedures can also be found in Wellek (1993) and Liu and Chow (1997a). Schall and Luus (1993) and Schall (1995) provided the discussion of the relationship between the moment- and probability-based criteria.

Note that decision rules, regulatory aspects, and statistical evaluations regarding population bioequivalence and individual bioequivalence are discussed further in Chapters 11 and 12, respectively.

1.7 Statistical Considerations

In this section, some statistical considerations that may occur in the assessment of bioequivalence are summarized.

1.7.1 AUC Calculation

As indicated in Section 1.3, among the pharmacokinetic parameters, AUC is the primary systematic exposure measure of the extent of absorption; or the amount of drug absorbed in the body, which is often used to assess bioequivalence between drug products. AUC is usually calculated using the trapezoidal rule based on the blood or plasma concentrations obtained at various sampling time points. In practice, a few missing values or unexpected observations may occur at some sampling time points owing to laboratory error, data transcription error, or other causes unrelated to bioequivalence. Generally, missing values or unexpected observations between two end sampling time points have little effect on the comparison of bioavailability (Rodda, 1986). However, if many missing values or unexpected observations occur in the plasma concentration–time curve, especially at two end sampling time points, the bias of the estimated AUC could be substantial and, consequently, may affect the comparison of bioavailability. Thus, how to justify the bias in the calculation of AUC is an important statistical issue. Furthermore, because the concentration at time zero (i.e., immediately before drug administration) may be different from subject to subject, whether or not the AUC should be adjusted from the baseline concentration is an interesting problem for both the clinician and biostatistician.

1.7.2 Model Selection and Normality Assumptions

Let μ_T and μ_R be the true averages for test and reference products, respectively. According to the 80/125 rule for assessment of average bioequivalence, the ratio of true averages (μ_T/μ_R) must be within (80%, 125%), with 90% assurance to claim bioequivalence. A typical approach is to construct a 90% confidence interval for μ_T/μ_R and compare it with (80%, 125%). If the constructed confidence interval is within (80%, 125%), then average bioequivalence is concluded. To construct a 90% confidence interval for μ_T/μ_R, two statistical models, namely, the raw data model (or additive model) and the log-transformed model (or multiplicative model), are often considered.

For the raw data model, an exact 90% confidence interval for $\mu_T - \mu_R$ is constructed based on the original data (raw data) and is converted to the confidence

interval for μ_T/μ_R by dividing by the observed reference mean ($\overline{Y}_R$) (assuming that $\overline{Y}_R$ is the true μ_R). The constructed confidence interval, however, is not at the exact 90% confidence level because the method ignores the variability of $\overline{Y}_R$. Another method is to use Fieller's theorem (Locke, 1984; Schuirmann, 1989) to construct an exact 90% confidence interval for μ_T/μ_R. This method is derived based on the ratio of sample means for test and reference products. The disadvantage of this method is that the distribution of the ratio of sample means is rather complicated and its moments may not exist (Hinkley, 1969). In practice, it is important to provide a further statistical evaluation of the above confidence intervals because the decision of bioequivalence is made based on whether or not the confidence interval is within 80% and 125% (Schuirmann, 1989).

The primary assumptions of the raw data model are normality assumptions. Since the AUCs, t_{max}, and C_{max} are positive quantities, the underlying distributions are, in fact, normal distribution truncated at 0. This is a valid argument against the raw data model. In addition, the distribution of AUC is often skewed. Thus, a log transformation on AUC is usually performed to remove the skewness. The log-transformed data is then analyzed using the raw data model, which is equivalent to analyzing the raw data using the log-transformed model. Under the normality assumptions, the log-transformed model can provide an exact confidence interval for μ_T/μ_R (Mandallaz and Mau, 1981). Thus, compared with the raw data model, the FDA recommends that the log-transformed model should be used for the analysis of AUC and C_{max} in the bioequivalence studies (FDA, 2003b; see also, Attachment 5, report by the Bioequivalence Task Force, 1988).

The above methods, based on either the raw data model or the log-transformed model, are derived under the assumptions of normality or lognormality for between subject (inter-subject) and within subject (intra-subject) variabilities. One of the difficulties commonly encountered is whether or not the assumption of normality or lognormality is valid. Thus, it is suggested that the normality or lognormality assumptions be checked before an appropriate statistical model is used. The tests for normality or lognormality assumptions are critical for choosing an appropriate model. Unfortunately, thus far, there exist no convincing statistical tests for normality or lognormaltiy assumptions for inter- and intra-subject variabilities in bioequivalence studies. Jones and Kenward (2003) recommended a method using studentized residuals, which are obtained under the model (they are approximately independent) for testing normality of an intra-subject variability based on the Shapiro–Walk statistic (Shapiro and Wilk, 1965). A similar approach is also suggested for testing the normality of an inter-subject variability. Owing to the difficulty of testing normality assumptions, in the past four decades, some research efforts were directed to the search for nonparametric alternatives (see e.g., Koch, 1972; Cornell, 1980; Hauschke et al., 1990).

1.7.3 Inter- and Intra-Subject Variabilities

Because individual subjects may differ widely in their responses to the drug, the knowledge of inter- and intra-subject variabilities may provide valuable information in the assessment of bioequivalence (Wagner, 1971). To improve the intra-subject

variability from the comparison of bioavailability between drug products, a crossover design, which is the design of choice by many investigators and is acceptable to the FDA (21 CFR, 320.26 and 320.27), is often considered. The advantages of using a crossover design are

1. Each subject can serve as his or her own control.
2. The assessment of bioequivalence is based on the intra-subject variability.
3. Fewer subjects are required to provide the desired degree of accuracy and power compared with other designs, such as parallel design.

However, in a crossover design, the intra-subject variability may be confoudned with some expected and unexpected variabilities, such as lot-to-lot, product-to-product, and subject-by-product variabilities. These sources of variabilities are difficult to assess based on a nonreplicated crossover design or other currently available designs (Ekbohm and Melander, 1989). Thus, appropriate replicated crossover designs or methods are necessary for assessing these variabilities (Liu and Chow, 1995; Chow, 1996b; FDA, 2001, 2003b).

1.7.4 Interval Hypothesis and Two One-Sided Tests

As early as the 1970s, statisticians became aware that the usual hypothesis testing for equality was not appropriate for bioavailability studies (Metzler, 1974). The purpose of bioequivalence is to verify that two formulations are indeed bioequivalent. Thus, from a statistical viewpoint, it may be more appropriate to reverse the null hypothesis of bioequivalence and the alternative hypothesis of bioinequivalence. Let θ_1 and θ_2 be two known bioequivalence limits and θ be the parameter of interest. The hypotheses for assessment of bioequivalence are given as follows:

$$\begin{aligned} & H_0\colon\ \theta \le \theta_1 \text{ or } \theta \ge \theta_2 \\ \text{versus} \quad & H_a\colon\ \theta_1 < \theta < \theta_2, \end{aligned}$$

which can be further decomposed into two one-sided hypotheses as

$$\begin{aligned} & H_{01}\colon\ \theta \le \theta_1 \\ \text{versus} \quad & H_{a1}\colon\ \theta_1 < \theta, \end{aligned}$$

and

$$\begin{aligned} & H_{02}\colon\ \theta \ge \theta_2 \\ \text{versus} \quad & H_{a2}\colon\ \theta < \theta_2. \end{aligned}$$

Since the hypothesis of bioequivalence in H_a is expressed as an interval, it is referred to as the interval hypothesis. The test procedures for the average bioavailability based on the interval hypothesis were proposed by Schuirmann (1981, 1987)

and Anderson and Hauck (1983). The distribution of the observed test statistic proposed by Anderson and Hauck can be approximated by a central *t*-distribution. Schuirmann's procedure uses two one-sided tests for assessment of equivalence in average bioavailability. In this approach, two *p*-values are obtained to evaluate whether the bioavailability of the test product is not too low for one side (H_{01} vs. H_{a1}) and whether the bioavailability is not too high for the other side (H_{02} vs. H_{a2}). However, it is unclear what the exact *p*-value is for H_0 versus H_a because for any given θ_1 and θ_2 and the observed statistic for H_{01} versus H_{a1}, the *p*-value for H_{02} versus H_{a2} is not a random variable, but a fixed known quantity. In addition, the above two approaches suffer from the fact that under the normality assumption and unknown variances, in finite samples, there is no unconditional uniformly most powerful unbiased (UMPU) (nor invariant) test (Kendall and Stuart, 1979; Hsu et al., 1994; Lehmann and Romano, 2005). In other words, there always exist procedures with greater power for the same hypotheses under certain conditions. Alternatively, several nonparametric procedures have been proposed (Hauschke et al., 1990; Liu, 1991). However, there is little or no information available on the relative efficiency of the nonparametric procedures to the parametric methods (Liu and Weng, 1994).

1.7.5 Outlier Detection

As indicated in the report by the bioequivalence task force, the detection and treatment of outlying data in bioequivalence studies are important issues because the results and decisions of bioequivalence could be totally different by including or excluding the outlying data in the analysis. Several tests have been proposed for the detection of outlaying data (Chow and Tse, 1990a; Lin and Tsong, 1990; Liu and Weng, 1991; Wang et al. 1995, Wang and Chow, 2003). However, additional research and the development of some robust procedures are needed in this area.

1.7.6 Subject-by-Formulation Interaction

The concept of individual bioequivalence is first to investigate the closeness of the marginal distributions in bioavailabilities between the test and reference formulations in a subject, and then to assemble this information over a group of K subjects for assessment of individual bioequivalence. As a result, the individual difference in average and the individual ratio of intra-subject variabilities of the marginal distributions may be different from subject to subject. This phenomenon is referred to as the subject-by-formulation interaction. Thus, the subject-by-formulation interaction, in a general sense, should consider both average and intra-subject variability. However, the current state-of-art moment-based criteria for individual bioequivalence take into account only the difference in averages in the form of the variance of the deviations of the individual differences from the population difference. They ignore the differences (ratios) of the individual ratios of intra-subject variabilities from the population ratio. Most recently, Endrenyi and Tothfalusi (1999) and Endrenyi et al. (2000) studied

statistical properties of the estimated variance component for subject by formulation in studies of individual bioequivalence. However, further research is required to understand whether this information is important in assessment of bioequivalence.

1.7.7 Meta-Analysis of Bioequivalence

The current regulations only request that bioequivalence of the generic copies to the innovator drug product be established. As a result, all development of concepts, such as prescribability and switchability, and definitions of average, population, and individual bioequivalence are concentrated only on the comparison of the test formulation with the approved innovator reference product. However, as more generic copies become available, switch between different generic copies of the same innovator product is inevitable. This situation is particularly true for developing countries when only cheaper generic copies are available. Even in the well-developed countries, such as the United States, owing to a desire to contain spiral increasing health cost, switch between generic copies is still possible under certain circumstance; for example, change of health care providers because of cheaper premium paid by employers or job changes by employees. As a result, the safety of generic copies has become a public issue not only because the number of generic copies for the same approved reference product can be as many as 160, but also they are not identical in terms of inactive ingredients that are binded and bulked, coated and colored, and may vary from one version to another. Chow and Liu (1997) and Chow and Shao (1999) showed how to apply meta-analysis to a systemic overview of independent bioequivalence trials for assessment of prescribability between generic copies. This is that one area in bioequivalence that is often ignored, but truly required for immediate attention.

1.7.8 Other Issues

Several issues concerning the assessment of bioequivalence have been discussed. See, e.g., Chow and Ju (1994), Chow and Liu (1995a), Chow (1996a), Chow (1997a), and Liu (2004). These include the determination of the bioequivalence limit for individual bioequivalence (Chen, 1996). The equivalence limit of 80%–125% has been accepted by the regulatory agencies, academia, and pharmaceutical industry of most countries for average bioequivalence. However, debates for selection of the equivalence limits for subject-by-formulation interaction and ratio of intra-subject variabilities is still going on and will last for the foreseeable future. More research is required for a procedure of determination of equivalence limits for aggregate and disaggregate moment-based criteria and probability-based criteria and their justifications. On the other hand, the sponsor needs to use more resources for assessment of individual bioequivalence. As a result, search for the optimal or nearly optimal replicated crossover designs in terms of relative efficiency is also urgently needed for individual bioequivalence. In most of developing countries, due to the cost, only the generic copies of the innovator from the original country are available.

Then how to assess bioequivalence between the generic drugs developed by the local generic sponsors with that from the original country is not only an important regulatory issue but also a critical public health issue for the developing countries.

1.8 Aims and Structure of the Book

This is intended to be the first book entirely devoted to the design and analysis of bioavailability/bioequivalence studies. It covers all of the statistical issues that may occur in the various stages of design and data analysis in bioavailability/bioequivalence studies. It is our goal to provide a useful desk reference and state-of-the-art examination of this area to scientists engaged in pharmaceutical research, those in government regulatory agencies who have to make decisions on the bioequivalence between drug products, and to biostatisticians who provide the statistical support for bioavailability/bioequivalence studies and related clinical projects. More importantly we would like to provide graduate students in pharmacokinetics, clinical pharmacology, biopharmaceutics, and biostatistics an advanced textbook in bioavailability studies. We hope that this book can serve as a bridge among the pharmaceutical industry, government regulatory agencies, and academia.

This book is configured into the following five components: preliminaries, average bioequivalence, population/individual bioequivalence, *in vitro* and alternative evaluation of bioequivalence, and other bioequivalence studies. The preliminary part covers from Chapters 1 through 3. In this chapter, the history, definition, decision rules, and some statistical considerations for bioavailability studies have been discussed. In Chapter 2, some basic considerations on the concerns of the investigator, monitor, and biostatistician for the designs of bioavailability studies are discussed. We then introduce some designs that are currently available for bioavailability studies. The relative advantages of a crossover design that is acceptable to the FDA are extensively discussed in this chapter. In Chapter 3, statistical inference for a variety of effects from a standard 2×2 crossover design is discussed. Because currently the regulatory agencies of most countries in the world require the evidence of average bioequivalence, Chapters 4 through 10 of this book are entirely devoted to the design and analysis of average bioequivalence. Statistical methods currently available for the assessment of average bioequivalence are provided in Chapter 4. The nonparametric methods including bootstrap resampling procedure will also be extensively explored in this chapter. These methods are compared in terms of power and relative efficiency in Chapter 5. Sampling size determination for average bioequivalence is also included in this chapter. The log-transformed model and the approach using individual subject ratios are given in Chapter 6. In addition to the examination of intra-subject variability and inter-subject variability, the assessment of bioequivalence using the variability of bioavailability is explored in Chapter 7. In Chapter 8, some tests for normality assumptions and procedures for detection of outliers are derived. Chapter 9 provides statistical methods for assessing average

bioequivalence under a higher-order crossover design for two formulations. Assessment of average bioequivalence for more than two formulations is outlined in Chapter 10. Population and individual bioequivalence are covered in Chapters 11 and 12. In Chapter 11, the merits, desirable features of population bioequivalence and individual bioequivalence, and different criteria of individual bioequivalence, and different criteria of individual bioequivalence and their rationales and relations are discussed. In addition, different replicated crossover designs are introduced and compared for evaluation of individual bioequivalence in this chapter. Different statistical procedures for aggregated and disaggregate moment-based criteria and probability-based criteria are provided in Chapter 12. Chapters 13 through 15 discuss the topics on *in vitro* and alternative evaluation of bioequivalence. Chapter 13 gives an introduction for assessment of bioequivalence based on clinical endpoints, such as response data and time to onset of a therapeutic response when plasma concentrations are negligible. In Chapter 14, statistical methods for assessment of bioequivalence based on *in vitro* bioequivalence testing for local delivery drug products such as nasal aerosols and nasal sprays are described. Criteria and statistical procedures for assessment of similarity between dissolution profiles are given in Chapter 15. Chapters 16 through 20 review some other bioequivalence studies. Chapter 16 proposes meta-analysis approaches to bioequivalence review based on average bioequivalence. The approach can be applied to the concept of population and individual bioequivalence. Comparison between these methods and power and sample size determination are also discussed in this chapter. In Chapter 17, objectives as well as the design and procedures for population pharmacokinetics are provided. Also included in this chapter is the assessment of inter- and intra-subject variabilities in multicompartmental PK model. In Chapter 18, other pharmacokinetic studies such as drug interaction studies, dose proportionality studies, steady-state analyses for multiple doses, and food effects studies are given. Chapter 19 provides a thorough review of the FDA guidances on bioequivalence including the 2001 FDA guidance on statistical approaches (FDA, 2001), the 2003 FDA guidance on general considerations for orally administrated drug products (FDA, 2003b), and other FDA guidances such as the guidance on fed bioequivalence, Clozapine tablets, and the SUPAC guidances. Chapter 20 addresses some frequently asked questions and future challenges on bioequivalence, which include assessment of bioequivalence with genomic data, bridging bioequivalence studies, and bioequivalence for biological products (follow-on biologics or biosimilar drug products).

Chapter 2

Design of Bioavailability Studies

2.1 Introduction

Before a clinical trial is conducted, a protocol that details the conduct of the trial is usually developed. A thoughtful and well-organized protocol includes study objectives, design, patient selection criteria, dosing schedules, and statistical methods. Unlike clinical trials, bioavailability studies are often conducted with healthy volunteers. Thus, the choice of the design and the statistical methods for the analysis of data becomes two important aspects in planning a bioavailability study. These two aspects are closely related to each other because the method of analysis depends on the design employed. Generally meaningful conclusions can only be drawn based on data collected from a valid scientific design using appropriate statistical methods. General considerations that one should consider when planning a bioavailability study include

1. What is to be studied, or what are the study objectives?
2. How are the data to be collected, or what design is to be employed?
3. How are the data to be analyzed, or what statistical methods are to be used?

In this chapter, our efforts will be directed to the determination of study objectives and the selection of an appropriate design for a bioavailability study. We intend to explore and compare some basic designs that are currently available for such studies. Some specific designs that are used for different purposes under various circumstances are discussed further in Chapter 9. Unless otherwise specified, throughout this book, for the sake of convenience, we restrict our attention to the comparison of different formulations of the same drug product. The comparison of different drug products of the same active ingredient and different ways of administration can be treated similarly.

The choice of the design depends primarily on the variability in the observations. For example, as indicated in Section 1.7.3, the individual subjects may differ very widely in their responses to the drug products. Thus, one major source of variability arises from differences between subjects. As a result, a criterion for choosing an appropriate design is whether or not the selected design can identify, estimate, and

isolate the inter-subject variability in data analysis. Any design that can remove this variation from the comparison in average bioavailability between formulations would be appropriate. Such a design is generally more efficient than a design that cannot account for the inter-subject variability. In this chapter, we induce several designs that are often considered for bioavailability and bioequivalence studies. These designs include the complete randomized designs (or the parallel designs), the randomized block designs, the crossover designs, the Latin square designs, and the (balanced) incomplete block designs. These designs, which may remove the expected variability from the comparison of bioavailability between formulations, may be useful, depending on the parameters to be evaluated, the characteristics of the drug, or the medical restrictions.

The remainder of this chapter is organized as follows. In Section 2.2, objectives for some studies related to bioavailability, such as bioequivalence studies, proportionality studies, and steady-state analyses are discussed. In Section 2.3, we provide some design considerations when planning a bioavailability study. In Section 2.4, a brief description of a parallel design is given. An extensive discussion on crossover designs is presented in Section 2.5. Balanced incomplete block designs are introduced in Section 2.6. Some factors for choosing an appropriate design for bioavailability studies are discussed in Section 2.7.

2.2 Study Objective

In clinical trials, a description of the general aims of the study is a useful preliminary that helps explain why the study is considered worthwhile (Pocock, 1983). The statement of study objectives is a concise and precise definition of prespecified hypotheses or parameters concerning the drug products that are to be examined or estimated. In clinical trials, a clear statement of study objectives not only ensures that the investigator adhere to the hypotheses at the time of analysis and interpretation of results, but also enables statisticians to select an appropriate design and statistical methods for data analysis.

In the following, some examples of study objectives and corresponding hypotheses or parameters of interest in bioavailability and related studies are given.

2.2.1 Bioequivalence Studies

One of the objectives of a bioequivalence study is to compare bioavailability between two formulations (a test and a reference formulation) of a drug product and to determine bioequivalence in terms of the rate and extent of absorption. The primary hypothesis may be whether the difference in average bioavailability between a test and reference product is within 80%–125% of the reference mean with certain assurance. On the other hand, individual bioequivalence might be the objective of other bioequivalence studies. To achieve this objective, a crossover design is often considered. Several statistical methods are available for the evaluation of different hypotheses.

2.2.2 Dose Proportionality Studies

For a dose proportionality (or dose linearity study), the objective is to evaluate whether the relationship between dose level and a pharmacokinetic parameter (such as AUC) is linear over a given dose range. The results may provide useful information in determining dose levels at which the minimum concentration for therapeutic effect and toxic concentration will be achieved. The hypothesis of interest is that there is a linear relation between dose level and AUC. Several statistical tests for the hypothesis of dose proportionality are available for both serial blood collection and single time-point blood collection. More details on dose proportionality studies are discussed further in Chapter 18.

2.2.3 Steady-State Studies

For a steady-state study, comparison of the blood (or plasma) concentration is made after steady state is achieved (generally, after multiple dosing). The objective of such a study is to determine whether a steady state has been reached and when it was reached. This may be evaluated by testing the hypothesis that there is no difference in concentrations at the end of each dosing interval. In Chapter 18, more details on a steady-state analysis are given.

2.2.4 Variability and Interchangeability

Because the determination of bioequivalence may not adequately characterize different types of variation that can occur both within a given individual as well as among different individuals, an appropriate design may be considered to provide information on the inter-subject and the intra-subject variabilities and the interchangeability of one formulation for another. The objective of such design is to estimate the inter-subject and intra-subject variabilities and provide statistical inference on both the variability and the interchangeability. This issue is examined in Chapters 7 and 11.

Pocock (1983) indicated that, in clinical trials, the study objectives are built on more expansive descriptions of patient selection criteria, treatment schedules, and the methods of patient evaluation. Although a precise and detailed explanation of these issues can help ensure that an unbiased assessment of the study objectives is achieved, a valid scientific design with appropriate statistical methods for the analysis of data is the key to carry out the study objectives.

2.3 Basis Design Considerations

In the *Federal Register* [Vol. 42, No. 5, Sect. 320.25(b), 1977], the U.S. Food and Drug Administration indicated that a basic design for an *in vivo* bioavailability study is determined by the following:

1. Scientific questions to be answered
2. Nature of the reference material and the dosage form to be tested
3. Availability of analytical methods
4. Benefit–risk considerations relative to human testing

Consideration of the reference dosage form is critical. For example, a suspension may not be an appropriate reference material because of high variability in bioavailability of the suspension dosage form. In many instances, a suspension of poorly soluble active drug ingredient may be more poorly absorbed than a well-formulated tablet.

The availability of the analytical method that is used to measure the immediate pharmacological effect or concentration of the active drug ingredient, therapeutic moiety, or metabolites is important. The FDA requires that the analytical method used in bioavailability studies be of sufficient accuracy, sensitivity, and reproducibility to discriminate between inequivalent products. The requirement implies that a product of known poor bioavailability must be compared against the reference product to determine whether the method can detect differences between the two products.

Finally, in practice, most bioavailability studies are conducted with healthy normal subjects. Bioavailability studies conducted on critically ill patients may not be appropriate and be contrary to the best medical practice unless there is a definitive benefit to the patients. For example, a bioavailability study with kanamycin in patients with stable renal disease would permit dosage adjustments based on renal creatinine clearance and serum kanamycin levels.

In addition to these basis design considerations, some specific considerations when planning a design for a bioavailability study are given in the following.

2.3.1 Experimental Design

The *Federal Register* [Vol. 42, No. 5, Secs. 320.26(b) and 320.27(b), 1977] indicated that a bioavailability study (single-dose or multidose) should be crossover in design, unless a parallel or other design is more appropriate for valid scientific reasons. For a parallel design, each subject receives one and only one formulation in random fashion, whereas for a crossover design each subject receives more than one formulation at different periods. In practice, subjects account for a large source of variability in plasma or blood drug concentrations. Thus, an appropriate design should allow estimation and removal of the inter-subject variability from drug comparisons. More details on the parallel, crossover, and other designs are discussed in the following sections.

2.3.2 Randomization

Valid statistical inferences are usually drawn based on the assumption that the errors in observations are independently distributed, random variables. Randomization

usually ensures the validity of this assumption. The randomization schedules depend on the design selected. For example, for a parallel design comparing two formulations of a drug product, the subjects are assigned to receive each formulation at random. For a crossover design, each subject is a block that represents a restriction on complete randomization because the formulations are randomized within the subject. An example of randomization for a standard two-sequence, two-period (2×2) crossover design is given in Section 2.5.4.

2.3.3 Sampling Time Intervals

For the estimation of the rate and extent of absorption, although the sampling time intervals for both the test and reference formulations need not be the same, it is preferred that sampling time intervals are identical to assure true equivalence. It, however, should be noted that the actual sampling times may deviate from the scheduled sampling times in practice. On the other hand, blood or plasma samples should be collected at the time before dosing and over an interval of sufficient time (e.g., three to five half-lives of the drug active ingredient or therapeutic moiety) to accurately determine the individual terminal disposition curve.

2.3.4 Drug Elimination Period

For a single-dose study, the terminal drug elimination period should allow at least three or more terminal half-lives of the active drug ingredient or therapeutic moiety, or its metabolite, either measured in the blood or as the decay of the immediate pharmacological effect. For a multiple dose study, the elimination period should allow at least five half-lives.

2.3.5 Number of Subjects

For a bioavailability study, usually 18–24 healthy normal subjects are used. To detect a clinically important average difference (e.g., 20%), a prestudy power calculation is often performed to determine the number of subjects needed for detection of such difference with a desired probability (e.g., 80%). The issue of power and sample size determination is discussed further in Chapter 5.

2.4 Parallel Design

A parallel design is complete randomized design in which each subject receives one and only one formulation of a drug in a random fashion. The simplest parallel design is the two-group parallel design, which compares two formulations of a drug.

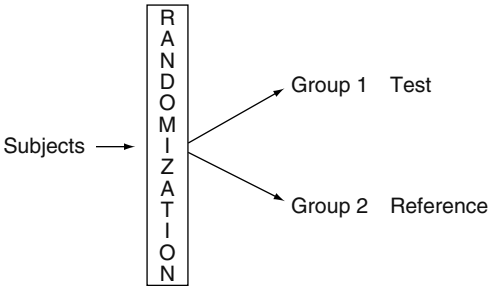


FIGURE 2.4.1: Two-group parallel design.

Each group usually contains the same number of subjects. An example of a two-group parallel design is illustrated in Figure 2.4.1.

For phase II and III clinical trials, the parallel design probably is the one most frequently used. However, it may not be an appropriate design for bioavailability and bioequivalence studies. This is because the variability in observations (e.g., AUC) consists of the inter-subject and intra-subject variabilities and the assessment of bioequivalence between formulations is usually made based on the intra-subject variability. A parallel design, however, is not able to identify and separate these two sources of variations because each subject in the parallel design usually receives the same drug during the entire course of study. Although the equivalence in average bioavailability between formulations can still be established through this design, the comparison is made based on the inter-subject and intra-subject variabilities. As a result, for a fixed number of subjects, the parallel design would, in general, provide a less precise statistical inference for the difference in average bioavailability between formulations than that of a crossover design.

Although the parallel design is not widely used for bioavailability studies owing to the incapability of identifying and removing the inter-subject variability from the comparison between formulations, there are some rare occasions in which a parallel design may be more appropriate than a crossover design. For example, for generic topical antifungals bioequivalence study, the FDA requires a three-arm parallel design (i.e., test, reference, and vehicle control). If the drug is known to have a very long half-life, it is not desirable to adapt a crossover design. In a crossover design, a sufficient length of washout is necessary to eliminate the possible carryover effects and, consequently, the study may take considerable time. This, in turn, may increase the number of dropouts and make the completion of a study difficult. In addition, if the study is to be conducted with very ill patients, a parallel design is usually recommended so that the study can be completed quickly. As a result a parallel design may be considered as an alternative to a crossover design if (1) the inter-subject variability is relatively small compared with the intra-subject variability; (2) the drug is potentially toxic or has a very long elimination half-life; (3) the population of interest consists of very ill patients; and (4) the cost for increasing the number of subjects is much less than that of adding an additional treatment period.

2.5 Crossover Design

2.5.1 Introduction

A crossover design is a modified, randomized block design in which each block receives more than one formulation of a drug at different periods. A block may be a subject or a group of subjects. Subjects in each block receive a different sequence of formulations. A crossover design is called a complete crossover design if each sequence contains each of the formulations. For a crossover design, it is not necessary that the number of formulations in each sequence be greater than or equal to the number of formulations to be compared. We shall refer to a crossover design as a $g \times p$ crossover design if there are g sequences of formulations administered at p different periods. For bioavailability and bioequivalence studies, the crossover design is viewed favorably by the FDA and other regulatory agencies such as EMEA in the world because of the following advantages:

1. Each subject serves as his or her own control. It allows a within-subject comparison between formulations.
2. It removes the inter-subject variability from the comparison between formulations.
3. With a proper randomization of subjects to the sequence of formulation administrations, it provides the best unbiased estimates for the differences (or ratios) between formulations.

The use of crossover designs for clinical trails has been extensively discussed in the literature. See, for example, Brown (1980), Huitson et al. (1982), Jones and Kenward (2003), and Senn (1993).

In the following, we introduce several different types of crossover designs that are often used in bioavailability studies. The relative advantages and drawbacks of these designs are also discussed.

2.5.2 Washout and Carryover Effects

It is helpful to introduce the concepts of washout and carryover effects (or residual effects) in a crossover design because the presence of carryover effects usually has an influence on statistical inference of bioavailability between formulations.

The washout period is defined as the rest period between two treatment periods for which the effect of one formulation administered at one treatment period does not carry over to the next. In a crossover design, the washout period should be long enough for the formulation effects to wear off so that there is no carryover effect from one treatment period to the next. The washout period depends on the nature of the drug. A suitable washout period should be long enough to return any relevant changes that influence bioavailability to baseline (usually, at least five times the blood–plasma elimination half-life of the active ingredient, therapeutic moiety or its metabolite, or the decay of the immediate pharmacological effect since the last sampling time point of the previous period).

If a drug has a long half-life or if the washout period between treatment periods is too short, the effect of the drug might persist after the end of dosing period. In this case, it is necessary to distinguish the difference between the direct drug effect and the carryover effects. The direct drug effect is the effect that a drug product has during the period in which the drug is administered, whereas the carryover effect is the drug effect that persists after the end of the dosing period. Carryover effects that last only one treatment period are called first-order carryover effects. A drug is said to have c-order carryover effects if the carryover effects last up to c treatment periods. In bioavailability and bioequivalence studies, however, it is unlikely that a drug effect will carry over more than one treatment period because a sufficient length of washout is usually considered. In this book, therefore, we consider only the first-order carryover effects if they are present.

2.5.3 Statistical Model and Linear Contrast

In a crossover design, because the direct drug effect may be confounded with any carryover effects, it is important to remove the carryover effects from the comparison if possible. To account for these effects, the following statistical model is usually considered. Let Y_{ijk} be the response (e.g., AUC) of the ith subject in the kth sequence at the jth period.

$$Y_{ijk} = \mu + S_{ik} + P_j + F_{(j,k)} + C_{(j-1,k)} + e_{ijk}, \tag{2.5.1}$$

where

μ is the overall mean
S_{ik} is the random effect of the ith subject in the kth sequence, where $i = 1, 2, \ldots, g$
P_j is the fixed effect of the jth period, where $j = 1, \ldots, p$ and $\Sigma_j P_j = 0$
$F_{(j,k)}$ is the direct fixed effect of the formulation in the kth sequence which is administered at the jth period, and $\Sigma F_{(j,k)} = 0$
$C_{(j-1,k)}$ is the fixed first-order carryover effect of the formulation in the kth sequence which is administered at the $(j-1)$th period, where $C_{(0,k)} = 0$; and $\Sigma C_{(j-1,k)} = 0$
e_{ijk} is the (within-subject) random error in observing Y_{ijk}.

It is assumed that $\{S_{ik}\}$ are independently and identically distributed (i.i.d.) with mean 0 and variance σ_S^2, and $\{e_{ijk}\}$ are independently distributed with mean 0 and variance σ_t^2, where $t = 1, 2, \ldots, L$ (the number of formulations to be compared). $\{S_{ik}\}$ and $\{e_{ijk}\}$ are assumed mutually independent. The estimate of σ_S^2 is usually used to explain the inter-subject variability, and the estimates of σ_t^2 are used to assess the intra-subject variabilities for the tth formulation.

Let $\bar{Y}_{\cdot 1k}, \bar{Y}_{\cdot 2k}, \ldots, \bar{Y}_{\cdot pk}$ be the observed means for periods in the kth sequence. That is,

$$\overline{Y}_{\cdot jk} = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{ijk}, \quad j = 1, \ldots, p \quad \text{and} \quad k = 1, \ldots, g. \tag{2.5.2}$$

Under the normality assumptions, the carryover effects and other fixed effects, such as the direct drug effect and the period effect, can be estimated based on these gp means because there are $(gp-1)$ degrees of freedom (df) among these gp means, which can be decomposed as follows:

$$(gp-1) = (p-1) + (g-1) + (p-1)(g-1),$$

where $(p-1)$ df are attributed to the period effect, $(g-1)$ df are assigned to the sequence effect, and $(p-1)(g-1)$ are associated with the sequence-by-period interaction. The $(p-1)(g-1)$ df are of particular interest because they preserve the information related to the direct drug effect and the carryover effects. For example, for a standard 2×2 crossover design, there are 3 df associated with four sequence-by period means: 1 for the sequence effect, 1 for the period effect, and 1 for the sequence-by-period interaction which is, in fact, the direct drug effect when there are no carryover effects.

A within-subject linear contrast for the kth sequence is defined as a linear combination of $\overline{Y}_{\cdot 1k}$, $\overline{Y}_{\cdot 2k}, \ldots,$ and $\overline{Y}_{\cdot pk}$. That is,

$$l = c_1\overline{Y}_{\cdot 1k} + c_2\overline{Y}_{\cdot 2k} + \cdots + c_p\overline{Y}_{\cdot pk},$$

where $\Sigma_j c_j = 0$.

Two linear combinations of $\overline{Y}_{\cdot jk}, j = 1, 2, \ldots, p$ are said to be orthogonal if the sum of the cross-products of the coefficients of the two contrasts is 0. In other words, let

$$l_1 = \sum_{j=1}^{p} c_{1j}\overline{Y}_{\cdot jk} \quad \text{and} \quad l_2 = \sum_{j=1}^{p} c_{2j}\overline{Y}_{\cdot jk}$$

be two linear contrasts, then l_1 and l_2 are orthogonal if

$$\sum_{j=1}^{p} c_{1j}c_{2j} = 0.$$

It can be seen that the variance of l involves only the intra-subject variabilities $\sigma_t^2, t = 1, 2, \ldots, L$. Thus, statistical inferences for the fixed effects, such as the period effects, the direct drug effects, and the carryover effects, can be made based on within-subject variabilities using appropriate linear contrasts of these gp means.

2.5.4 Crossover Designs for Two Formulations

In this section, we focus on the assessment of bioequivalence between test formulation (T) and a reference (or standard) formulation (R) of a drug product. The most commonly used statistical design for comparing average bioavailability between two formulations of a drug probably is a two-sequence, two-period, crossover design, we shall refer to this design as the standard 2×2 crossover design. For the standard 2×2 crossover design, each subject is randomly assigned to either

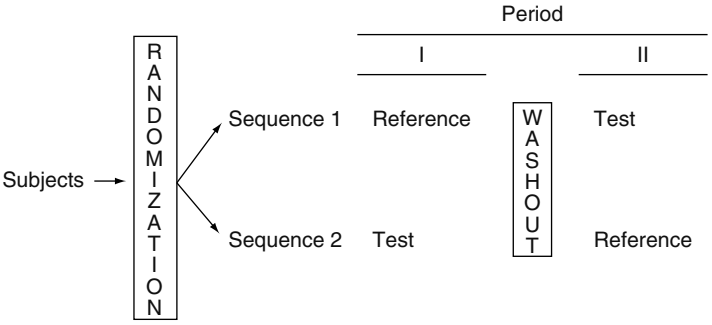


FIGURE 2.5.1: 2×2 crossover designs.

sequence RT or sequence TR at two dosing periods. In other words, subjects within RT (RT) receive for mulation R(T) at the first dosing period and formulation T(R) at the second dosing period. The dosing periods are separated by a washout period of sufficient length for the drug received in the first period to be completely metabolized or excreted from the body. An example of a 2×2 crossover design is illustrated in Figure 2.5.1. Although the crossover design is a variant of the Latin square design, the number of the formulations in a crossover design does not necessarily have to be equal to the number of periods. One example is a 2×3 crossover design for comparing two formulations as illustrated in Figure 2.5.2. In this design, there are two formulations, but three periods. Subjects in each sequence receive one of the formulations twice at two different dosing periods. The design of this kind is known as a higher-order crossover design which is discussed in detail in Chapter 9.

Randomization for the standard 2×2 crossover design can be carried out by using either a table of random numbers or an SAS procedure, PROC PLAN (SAS®*, 2005). For example, suppose the standard 2×2 crossover design is to be conducted with 24 healthy volunteers to assess bioequivalence between a test formulation and a

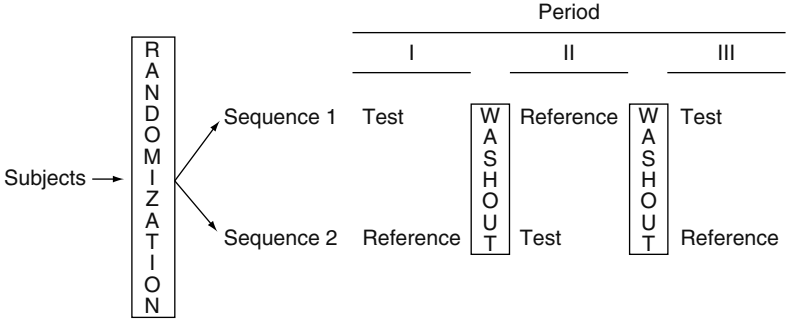


FIGURE 2.5.2: 2×3 crossover designs.

* Registered trademark SAS Institute, Cary, North Carolina.

reference formulation of a drug product. Because there are two sequences of formulations (RT and TR), 12 subjects are to be assigned to each of the two sequences. In other words, one group will receive the first sequence of formulations (RT) and the other group will receive the second sequence of formulations (TR). Thus, we first generate a set of random permutations from 1 to 24 using PROC PLAN, which follows

16, 19, 20, 11, 4, 24, 1, 12, 5, 23, 15, 6,
17, 2, 10, 14, 18, 13, 21, 3, 7, 8, 22, 9

Then, subjects are sequentially assigned a number from 1 through 24. Subjects with numbers in the first half of the above random order are assigned to the first sequence (RT) and the rest are assigned to the second sequence (TR) (see Table 2.5.1). In practice, a set of randomization code for more than the total number of subjects planed is usually prepared to account for the possible replacement of dropouts.

TABLE 2.5.1: Randomization codes for the standard 2×2 crossover design with 24 subjects.

Subject	Sequence	Formulations
1	1	RT
2	2	TR
3	2	TR
4	1	RT
5	1	RT
6	1	RT
7	2	TR
8	2	TR
9	2	TR
10	2	TR
11	1	RT
12	1	RT
13	2	TR
14	2	TR
15	1	RT
16	1	RT
17	2	TR
18	2	TR
19	1	RT
20	1	RT
21	2	TR
22	2	TR
23	1	RT
24	1	RT

For the standard 2×2 crossover design, from Equation 2.5.1, the two responses for the ith subject in each sequence are given as

$$\begin{aligned} \text{Sequence 1} \quad & Y_{i11} = \mu + S_{i1} + P_1 + F_1 + e_{i11} \\ & Y_{i21} = \mu + S_{i1} + P_2 + F_2 + C_1 + e_{i21} \\ \text{Sequence 2} \quad & Y_{i12} = \mu + S_{i2} + P_1 + F_2 + e_{i12} \\ & Y_{i22} = \mu + S_{i2} + P_2 + F_1 + C_2 + e_{i22} \end{aligned} \tag{2.5.3}$$

where

$P_1 + P_2 = 0$
$F_1 + F_2 = 0$
$C_1 + C_2 = 0.$

For each subject, a pair of observations is observed at periods 1 and 2. Thus, we may consider a bivariate random vector (i.e., [period 1, period 2]) as follows:

$$\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})', \quad i = 1, \ 2, \ \ldots, n_k \quad \text{and} \quad k = 1, \ 2. \tag{2.5.4}$$

Then, $\mathbf{Y}_{ik}$ values are independently distributed with the following mean vector and covariance matrix

$$\begin{aligned} \text{Sequence 1} \quad \boldsymbol{\alpha}_1 &= \begin{bmatrix} \mu + P_1 + F_1 \\ \mu + P_2 + F_2 + C_1 \end{bmatrix} \\ \boldsymbol{\Sigma}_1 &= \begin{bmatrix} \sigma_1^2 + \sigma_S^2 & \sigma_S^2 \\ \sigma_S^2 & \sigma_2^2 + \sigma_S^2 \end{bmatrix} \\ \text{Sequence 2} \quad \boldsymbol{\alpha}_2 &= \begin{bmatrix} \mu + P_1 + F_2 \\ \mu + P_2 + F_1 + C_2 \end{bmatrix} \\ \boldsymbol{\Sigma}_2 &= \begin{bmatrix} \sigma_2^2 + \sigma_S^2 & \sigma_S^2 \\ \sigma_S^2 & \sigma_1^2 + \sigma_S^2 \end{bmatrix} \end{aligned} \tag{2.5.5}$$

It can be seen that the intra-subject variabilities are different between formulations. If, however, $\sigma_1^2 = \sigma_2^2 = \sigma_e^2$, then $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_e^2 + \sigma_S^2 & \sigma_S^2 \\ \sigma_S^2 & \sigma_e^2 + \sigma_S^2 \end{bmatrix} \tag{2.5.6}$$

When the carryover effects are present (i.e., $C_1 \neq 0$ and $C_2 \neq 0$), the standard 2×2 crossover design may not be desirable, for it may not provide estimates for some fixed effects. For example, as indicated in the Subsection 2.5.3, there is only 1 degree of freedom, which is attributed to the sequence effect. The sequence effect, which cannot be estimated separately, is confounded (or aliased) with any carryover effects. If the

carryover effects are unequal (i.e., $C_1 \neq C_2 \neq 0$), then there exists no unbiased estimated for the direct drug effect from both periods. In addition, the carryover effects cannot be precisely estimated because it can be evaluated based on only the between subject comparison. Furthermore, the intra-subject variabilities σ_1^2 and σ_2^2 cannot be estimated independently and directly from the observed data because each subject receives either the test formulation or the reference formulation only once during the study. In other words, there are no replicates for each formulation within each subject.

To overcome these undesirable properties, a higher-order crossover design may be useful. A higher-order crossover design is defined as a crossover design in which either the number of periods is greater than the number of formulations to be compared, or the number of sequences is greater than the number of formulations to be compared. There are several higher-order crossover designs available in the literature (Kershner and Federer, 1981; Laska et al., 1983; Laska and Meinser, 1985; Jones and Kenward, 2003). These designs, however, have their own advantages and disadvantages. An in-depth discussion can be found in Jones and Kenward (2003).

In the following, we discuss three commonly used higher-order crossover designs, which possess some optimal statistical properties, for comparing average bioavailability between two formulations. We shall refer to these three designs as design A, B, and C, respectively. Designs A, B, and C are given in Table 2.5.2. In each of the

TABLE 2.5.2: Optimal crossover designs for two formulations.

Design A

	Period	
Sequence	**I**	**II**
1	T	T
2	R	R
3	R	T
4	T	R

Design B

	Period		
Sequence	**I**	**II**	**II**
1	T	R	R
2	R	T	T

Design C

	Period			
Sequence	**I**	**II**	**III**	**IV**
1	T	T	R	R
2	R	R	T	T
3	T	R	R	T
4	R	T	T	R

TABLE 2.5.3: Variances for designs A, B, and C in multiples of $\hat{\sigma}_e^2/n$.

Design	$V(\hat{C}\|F)$[a]	$V(\hat{F}\|C)$	$V(\hat{F})$
S[b]	—	—[c]	1.0000
A	4.0000	2.0000	1.0000
B	1.0000	0.7500	0.7500
C	0.3636	0.2500	0.2500

[a] $V(\hat{C}|F) = 4\,(2\hat{\sigma}_s^2 + \hat{\sigma}_e^2)/\text{n}$.
[b] S is the standard 2×2 crossover design.
[c] The direct drug effect is not estimable in the presence of the carryover effects.

three designs, the estimates of the direct drug effect and carryover effects are obtained based on the within-subject linear contrasts. As a result, statistical inferences for direct drug effect and carryover effects are mainly based on the intra-subject variability. For the comparisons of these three designs with the standard 2×2 crossover design, it is helpful to use the following notations. The direct drug effect after adjustment for the carryover effects is denoted by $F|C$. Then, F simply refers to the unadjusted direct drug effect. Also the variance of the estimator of $F|C$ (i.e., $\hat{F}|C$) is denoted by $V(\hat{F}|C)$. Table 2.5.3 gives the variances (in the multiples of σ_e^2/n) of the direct drug effect and carryover effects for the three designs and the standard 2×2 crossover designs (Senn, 1993; Jones and Kenward, 2003). The variances of designs A, B, and C are derived under the assumptions that (1) $n_k = n$ for all k; (2) $\sigma_1^2 = \sigma_2^2 = \sigma_e^2$; and (3) there is no direct drug-by-carryover interaction. For designs B and C, the direct drug effect adjusted for the carryover effects is the same as the unadjusted direct drug effect (i.e., no carryover effects). This is because the direct drug effect and carryover effects for designs B and C are estimated by the linear contrasts which are orthogonal to each other. Note that an orthogonality of linear contrasts for the direct drug effect and carryover effects implies that their covariance is zero. In other words, the estimators of the direct drug effect and carryover effects in designs B and C are not correlated (or independent).

Design A is also known as Balaam's design (Balaam, 1968). It is an optimal design in the class of the crossover designs with two periods and two formulations. This design is formed by adding two more sequences (sequences 1 and 2) to the standard 2×2 crossover design (sequences 3 and 4). These two augmented sequences are TT and RR. With additional information provided by the two augmented sequences, not only can the carryover effects be estimated using the within-subject contrasts, but the intra-subject variability for both test and reference formulations can also be obtained because there are replicates for each formulation each subject.

Design B is an optimal design in the class of the crossover designs with two sequences, three periods, and two formulations. It can be obtained by adding an additional period to the standard 2×2 crossover designs. The treatments administered in the third period are the same as those in the second period. This type of designs is also known as the extended-period or extraperiod designs. Note that this design is made of a pair of dual sequences TRR and RTT. Two sequences, the treatments

of which are mirror images of each other, are said to be a pair of dual sequences. As pointed out by Jones and Kenward (2003), the only crossover designs worth considering are those that are made up of dual sequences. Compared with the standard 2×2 crossover design, the variance for the direct drug effect is reduced by 25%.

For the carryover effects, the variance is reduced by about 75% as compared with the Balaam's design. In addition, the intra-subject variability can be estimated based on the data collected from periods 2 and 3.

Design C is an optimal design in the class of the crossover designs with four sequences, four periods, and two formulations. It is also made up of two pairs of dual sequences (TTRR, RRTT) and (TRRT, RTTR). Note that the first two periods of design C are the same as those in Balaam's design and the last two periods are the mirror image of the first two periods. The design is much more complicated than designs A and B, although it produces the maximum in variance reduction for both the direct drug effect and the carryover effects among the designs considered.

2.5.5 Crossover Designs for Three or More Formulations

The crossover designs for comparing three or more formulations are much more complicated than those for comparing two formulations. For simplicity, in this section, we restrict our attention to those designs in which the number of periods equals the number of formulations to be compared. In Section 2.6, the designs for comparing a large number of formulations with a small number of treatment periods are discussed.

For comparing three formulations of a drug, there are a total of three possible pairwise comparisons between formulations: formulation 1 versus formulation 2, formulation 1 versus formulation 3, and formulation 2 versus formulation 3. It is desirable to estimate these pairwise differences in average bioavailability between formulations with the same degree of precision. In other words, it is desirable to have equal variances for each pairwise differences in average bioavailability between formulations (i.e., $V(\hat{F}_i - \hat{F}_j) = \nu\sigma_e^2$) where ν is a constant and σ_e^2 is the intra-subject variability. Designs with this property are known as variance-balanced designs. It should be noted that, in practice, ν may vary from design to design. Thus, and ideal design is one with the smallest ν, such that all pairwise differences between formulations can be estimated with the same and possibly best precision. However, to achieve this goal, the design must be balanced. A design is said to be balanced if it satisfies the following conditions (Jones and Kenward, 1989, 2003):

1. Each formulation occurs only once with each subject.
2. Each formulation occurs the same number of times in each period.
3. The number of subjects who receive formulation i in some period followed by formulation j in the next period is the same for all $i \neq j$.

Under the constraint of the number of periods (p) being equal to the number of formulations (t), balance can be achieved by using a complete set of ''orthogonal Latin squares'' (John, 1971; Jones and Kenward, 2003). However, if $p = t$, a complete set of orthogonal Latin squares consists of $t(t-1)$ sequences except for $t = 6$. Some

TABLE 2.5.4: Orthogonal latin squares for $t = 3$ and 4.

Three Formulations ($t = 3$)

	Period		
Sequence	**I**	**II**	**III**
1	R[a]	T_1	T_2
2	T_1	T_2	R
3	T_2	R	T_1
4	R	T_2	T_1
5	T_1	R	T_2
6	T_2	T_1	R

Four Formulations ($t = 4$)

	Period			
Sequence	**I**	**II**	**III**	**IV**
1	R[a]	T_1	T_2	T_3
2	T_1	R	T_3	T_2
3	T_2	T_3	R	T_1
4	T_3	T_2	T_1	R
5	R	T_3	T_1	T_2
6	T_1	T_2	R	T_3
7	T_2	T_1	T_3	R
8	T_3	R	T_2	T_1
9	R	T_2	T_3	T_1
10	T_1	T_3	T_2	R
11	T_2	R	T_1	T_3
12	T_3	T_1	R	T_2

[a] R is the reference formulation and T_1, T_2, and T_3 are the test formulations 1, 2, 3, respectively.

examples of orthogonal Latin squares with $t = 3$ and $t = 4$ are presented in Table 2.5.4. As a result, when the number of formulations to be compared is large, more sequences and consequently more subjects are required. This, however, may not be of practical use.

A more practical design has been proposed by Williams (1949). We shall refer to this as a Williams design. A Williams design possesses balance property and requires fewer sequences and periods. The algorithm for constructing a Williams design with t periods and t formulations is summarized in the following numerical steps (Jones and Kenward, 2003):

1. Number of formulations from $1, 2, \ldots, t$.
2. Start with the $t \times t$ standard Latin square. In this square, the formulations in the ith row are given by $i, i+1, \ldots, t, 1, 2, \ldots, i-1$.
3. Obtain a mirror image of the standard Latin square.

4. Interlace each row of the standard Latin square with the corresponding mirror image to obtain a $t \times 2t$ arrangement.

5. Slice the $t \times 2t$ arrangement down to the middle to yield two $t \times t$ squares. The columns of each $t \times t$ squares correspond to the periods and the rows are the sequences. The numbers within the square are the formulations.

6. If t is even, choose any one of the two $t \times t$ squares. If t is odd, use both squares.

In the following, to illustrate the use of this algorithm as an example, we will construct a Williams design with $t = 4$ (one reference and three test formulations) by following the above steps.

1. Denote the reference formulations by 1, and test formulations 1, 2, and 3 by 2, 3, and 4.

2. The 4×4 standard Latin square is given as

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{array}$$

3. The minor image of the 4×4 standard Latin square is then given by

$$\begin{array}{cccc} 4 & 3 & 2 & 1 \\ 1 & 4 & 3 & 2 \\ 2 & 1 & 4 & 3 \\ 3 & 2 & 1 & 4 \end{array}$$

4. The 4×8 arrangement after interlacing the 4×4 standard Latin square with its mirror image is

$$\begin{array}{cccc|cccc} 1 & 4 & 2 & 3 & 3 & 2 & 4 & 1 \\ 2 & 1 & 3 & 4 & 4 & 3 & 1 & 2 \\ 3 & 2 & 4 & 1 & 1 & 4 & 2 & 3 \\ 4 & 3 & 1 & 2 & 2 & 1 & 3 & 4 \end{array}$$

5. The 4×4 squares obtained by slicing the above 4×8 arrangement are

		Period			
Square	**Sequence**	**I**	**II**	**III**	**IV**
1	1	1	4	2	3
	2	2	1	3	4
	3	3	2	4	1
	4	4	3	1	2
2	1	3	2	4	1
	2	4	3	1	2
	3	1	4	2	3
	4	2	1	3	4

6. Because $t = 4$, we can choose either square 1 or square 2. The resultant Williams design from square 1 is given in Table 2.5.5 by replacing 1, 2, 3, 4, with R, T_1, T_2, and T_3.

From the above example, it can be seen that a Williams design requires only 4 sequences to achieve the property of "variance-balanced," whereas a complete set of 4×4 orthogonal Latin squares requires 12 sequences. The Williams designs with $t = 3$ and 5 using the above algorithm are also given in Table 2.5.5.

TABLE 2.5.5: Williams designs for $t = 3$, 4, and 5.

Three Formulations ($t = 3$)

	Period		
Sequence	**I**	**II**	**III**
1	R[a]	T_2	T_1
2	T_1	R	T_2
3	T_2	T_1	R
4	T_1	T_2	R
5	T_2	R	T_1
6	R	T_1	T_2

Four Formulations ($t = 4$)

	Period			
Sequence	**I**	**II**	**III**	**IV**
1	R	T_3	T_1	T_2
2	T_1	R	T_2	T_3
3	T_2	T_1	T_3	R
4	T_3	T_2	R	T_1

Five Formulations ($t = 5$)

	Period				
Sequence	**I**	**II**	**III**	**IV**	**V**
1	R	T_4	T_1	T_3	T_2
2	T_1	R	T_2	T_4	T_3
3	T_2	T_1	T_3	R	T_4
4	T_3	T_2	T_4	T_1	R
5	T_4	T_3	R	T_2	T_1
6	T_2	T_3	T_1	T_4	R
7	T_3	T_4	T_2	R	T_1
8	T_4	R	T_3	T_1	T_2
9	R	T_1	T_4	T_2	T_3
10	T_1	T_2	R	T_3	T_4

[a] R is the reference formulation, and T_1, T_2, T_3, and T_4 are test formulations 1, 2, 3, and 4, respectively.

2.6 Balanced Incomplete Block Design

When comparing three or more formulations of a drug product, a complete crossover design may not be of practical interest for the following reasons (Westlake, 1973):

1. If the number of formulations to be compared is large, the study may be too time consuming, since t formulations require $t-1$ washout periods.
2. It may not be desirable to draw many blood samples for each subject owing to medical concerns.
3. Moreover, a subject is more likely to drop out when he or she is required to return frequently for tests.

These considerations suggest that one should keep the number of formulations that a subject receives as small as possible when planning a bioavailability study. For this, a randomized incomplete block design may be useful. An incomplete block design is a randomized block design in which not all formulations are present in every block. A block is called incomplete if the number of formulations in the block is less than the number of formulations to be compared. For an incomplete block design the blocks and formulations are not orthogonal to each other; that is, the block effects and formulation effects may not be estimated separately.

When an incomplete block design is used, it is recommended that the formulations in each block be randomly assigned in a balanced way so that the design will possess some optimal statistical properties. We shall refer to such a design as a balanced incomplete block design. A balanced incomplete block design is an incomplete block design in which any two formulations appear together an equal number of times. The advantages of using a balanced incomplete block design, rather than an incomplete design, are given as follows:

1. Difference in average bioavailability between the effects of any two formulations can always be estimated with the same degree of precision.
2. Analysis is simple in spite of the nonorthogonality provided that the balance is preserved.
3. Unbiased estimates of formulation effects are available.

Suppose that there are t formulations to be compared and each subject can only receive exactly p formulations ($t > p$). A balanced incomplete block design may be constructed by taking $C(t,p)$, the combinations of p out of t formulations, and assigning a different combination of formulations to each subject. However, to minimize the period effect, it is preferable to assign the formulations in such a way that the design is balanced over period (i.e., each formulation appears the same number of times in each period). In general, if the number of formulations is even (i.e., $t = 2n$) and $p = 2$, the number of blocks (sequences) required is $g = 2n(2n-1)$. On the other hand, if the number of formulations is odd (i.e., $t = 2n+1$) and $p = 2$, then $g = (2n+1)n$. Some examples for balanced incomplete block design are

TABLE 2.6.1: Balanced incomplete block designs for $t = 4$ with $p = 2$ and 3.

I. Each Sequence Receives Two Formulations ($p = 2$)

	Period	
Sequence[a]	**I**	**II**
1	R[b]	T_1
2	T_1	T_2
3	T_2	T_3
4	T_3	R
5	R	T_2
6	T_1	T_3
7	T_3	T_1
8	T_2	R
9	R	T_3
10	T_3	T_2
11	T_2	T_1
12	T_1	R

II. Each Sequence Receives Three Formulations ($p = 3$)

	Period		
Sequence	**I**	**II**	**III**
1	T_1	T_2	T_3
2	T_2	T_3	R
3	T_3	R	T_1
4	R	T_1	T_2

[a] A sequence (or block) may represent a subject or a group of homogeneous subjects.

[b] R is the reference formulation, and T_1, T_2, and T_3 are test formulations 1, 2, and 3, respectively.

given in Tables 2.6.1 and 2.6.2. Table 2.6.1 gives examples for $p = 2$ and 3 when four formulations ($t = 4$) are to be compared. For $p = 2$, the first six blocks are required for a balanced incomplete block design. However, to ensure the balance over period, an additional six blocks (7 through 12) are needed. For $t = 5$, Table 2.6.2 lists examples for a balanced incomplete block design with $p = 2$, 3, and 4. A balanced incomplete block design for $p = 3$ is the complementary part of that balanced incomplete block design for $p = 2$. The design for $p = 4$ can be constructed by deleting each formulation in turn to obtain five blocks successively.

For $t > 5$, several methods for constructing balanced incomplete block designs are available. Among these, the easiest way is probably the method of cyclic substitution. For this method to work, we first choose an appropriate initial block. The other blocks can be obtained successively by changing formulations A to B, B to C, . . . , and so on in each block. For example, for $t = 6$ and $p = 3$, if we start with (A, B, D), then the second block is (B, C, E), and the third block is (C, D, F), and so on.

TABLE 2.6.2: Balanced incomplete block designs for $t = 5$ with $p = 2$, 3, and 4.

I. Each Sequence Receives Two and Three Formulations

$p = 2$			$p = 3$			
	Period			Period		
Sequence[a]	I	II	Sequence	I	II	III
1	R[b]	T_1	1	T_2	T_3	T_4
2	T_1	T_2	2	T_3	T_4	R
3	T_2	T_3	3	T_4	R	T_1
4	T_3	T_4	4	R	T_1	T_2
5	T_4	R	5	T_1	T_2	T_3
6	R	T_2	6	T_1	T_3	T_4
7	T_2	T_4	7	T_3	R	T_1
8	T_4	T_1	8	R	T_2	T_3
9	T_1	T_3	9	T_2	T_4	R
10	T_3	R	10	T_4	T_1	T_2

II. Each Sequence Receives Four Formulations ($p = 4$)

	Period			
Sequence	I	II	III	IV
1	T_1	T_2	T_3	T_4
2	T_2	T_3	T_4	R
3	T_3	T_4	R	T_1
4	T_4	R	T_1	T_2
5	R	T_1	T_2	T_3

[a] A sequence (or block) may represent a subject or a group of homogeneous subjects.

[b] R is the reference formulation, and T_1, T_2, T_3, and T_4 are test formulations 1, 2, 3, and 4, respectively.

Note that a balanced incomplete block design is, in fact, a special case of variance-balanced design, which are discussed in Chapter 10. For an incomplete block design, balance may be achieved with fewer than $C(t,p)$ blocks. Such designs are known as partially balanced incomplete block designs. The analysis of these designs, however, is complicated, and hence, of little practical interest. More details on balanced incomplete black designs and partially balanced incomplete block designs can be found in Fisher ad Yates (1953), Bose et al. (1954), Cochran and Cox (1957), John (1971), and Cox and Reid (2000).

2.7 Selection of Design

In Sections 2.4 through 2.6, we briefly discussed three basic statistical designs, the parallel design, the crossover design, and the balanced incomplete block design for bioavailability and bioequivalence studies. Each of these has its own advantages

and drawbacks under different circumstances. How to select an appropriate design when planning a bioavailability study is an important question. The answer to this question depends on many factors that are summarized as follows:

1. Number of formulations to be compared
2. Characteristics of the drug and its disposition
3. Study objectives
4. Availability of subjects
5. Inter- and intra-subject variabilities
6. Duration of the study or the number of periods allowed
7. Cost of adding a subject relative to that of adding one period
8. Dropout rates

For example, if the intra-subject variability is the same as or larger than the inter-subject variability, the inference on the difference in average bioavailability would be the same regardless of which design is used. Actually, a crossover design in this situation would be a poor choice, because blocking results in the loss of some degrees of freedom and will actually lead to a wider confidence interval on the difference between formulations.

If a bioavailability and bioequivalence study compares more than three formulations, a crossover design may not be appropriate. The reasons, as indicated in Section 2.6, are (1) it may be too time consuming to complete the study because a washout is required between treatment periods; (2) it may not be desirable to draw many blood samples for each subject owing to medical concerns; and (3) too many periods may increase the number of dropouts. Here, a balanced incomplete block design is preferred. However, if we compare several test formulations with a reference formulation, the within-subject comparison is not reliable, as subjects in some sequences may not receive the reference formulation.

If the drug has a very long half-life, or it possesses a potential toxicity, or bioequivalence must be established by clinical endpoint because some drugs do not work through systemic absorption, then a parallel design may be a possible choice. With this design, the study avoids a possible cumulative toxicity from the carryover effects from one treatment period to the next. In addition, the study can be completed quickly. However, the drawback is that the comparison of average bioavailability is made based on the inter-subject variability. If the inter-subject variability is large relative to the inter-subject variability, the statistical inference on the difference in average bioavailability between formulations is unreliable. Even if the inter-subject variability is relatively small, a parallel design may still require more subjects to reach the same degree of precision achieved by a crossover design.

In practice, a crossover design, which can remove the inter-subject variability from the comparison of average bioavailability between formulations, is often considered to be the design of choice if the number of formulations to be compared

is small, say no more than three. If the drug has a very short half-life (i.e., there may not be carryover effects if the length of washout is long enough to eliminate the residual effects), a crossover design may be useful for the assessment of the intra-subject variability, provided that the cost for adding one period is comparable with that of adding a subject.

In summary, to choose an appropriate design for a bioavailability/bioequivalence study is an important issue in the development of a study protocol. The selected design may affect the data analysis, the interpretation of the results, and the determination of bioequivalence between formulations. Thus, all factors listed in the above should be carefully evaluated before an appropriate design is chosen.