

LINEAR LIBRARY
C01 0068 0712



UNIVERSITY OF CAPE TOWN
DEPARTMENT OF MATHEMATICAL STATISTICS

STATISTICAL ASPECTS OF BIOAVAILABILITY

by

John Lawrence Fresen

A thesis prepared under the supervision of
Associate Professor J.M. Juritz in fulfilment
of the requirements for the degree of
Master of Science in Mathematical Statistics

Copyright by the University of Cape Town
1985

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

*Mathematicians are like Frenchmen :
whatever you say to them they translate
into their own language and forthwith
it is something entirely different.*

Goethe

TO JILL AND DANIEL

TABLE OF CONTENTS

| | | Page |
|-----------|--|------|
| | ACKNOWLEDGEMENTS | i |
| | ABSTRACT | iii |
| CHAPTER 1 | OVERVIEW OF BIOAVAILABILITY | 1.1 |
| 1.1 | Concept of bioavailability | 1.1 |
| 1.2 | Definitions and assumptions | 1.3 |
| 1.3 | Measures of bioavailability | 1.4 |
| 1.4 | Pharmacokinetic modelling | 1.8 |
| 1.4.1 | Two compartment model | 1.9 |
| 1.4.2 | One compartment model | 1.12 |
| 1.5 | Statistical considerations | 1.14 |
| CHAPTER 2 | ESTIMATING THE BIOAVAILABILITY PARAMETERS | 2.1 |
| 2.1 | Introduction | 2.1 |
| 2.2 | Direct estimation of the AUC | 2.3 |
| 2.2.1 | The trapezoidal rule | 2.3 |
| 2.2.2 | The log-trapezoidal rule | 2.4 |
| 2.2.3 | Lagrange method | 2.5 |
| 2.2.4 | Cubic spline method | 2.6 |
| 2.2.5 | Extrapolating the AUC | 2.9 |
| 2.2.6 | Discussion on numerical integrating algorithms | 2.10 |
| 2.3 | Fitting a polyexponential model | 2.11 |
| 2.3.1 | The graphical method | 2.12 |
| 2.3.2 | Least squares | 2.14 |
| 2.3.3 | Weighted least squares | 2.15 |
| 2.3.4 | Maximum likelihood | 2.16 |
| 2.4 | Minimisation techniques | 2.17 |
| 2.4.1 | Steepest descent | 2.18 |
| 2.4.2 | Newton-Raphson | 2.19 |
| 2.4.3 | Gauss-Newton | 2.19 |
| 2.5 | Finding initial estimates | 2.21 |
| 2.5.1 | Foss's method for a sum of two exponentials | 2.22 |
| 2.5.2 | Comments on Foss's method | 2.25 |
| 2.5.3 | Modification | 2.26 |
| 2.5.4 | Extreme conditions | 2.26 |
| 2.6 | Asymptotic distribution of least squares/ maximal likelihood estimators | 2.29 |
| 2.7 | Other methods for estimating the pharmacokinetic parameters | 2.30 |
| 2.7.1 | Ordered simulation estimation procedure (OSEP) | 2.30 |
| 2.7.2 | Fourier analysis | 2.32 |
| 2.7.3 | L_p -norm estimation | 2.33 |
| 2.7.4 | Statistical moments | 2.33 |

| | | Page |
|-----------|--|------|
| CHAPTER 3 | ASSESSING BIOEQUIVALENCE | 3.1 |
| 3.1 | Introduction | 3.1 |
| 3.2 | The two period cross-over trial | 3.2 |
| 3.2.1 | Model I : With residual/carry over effects | 3.3 |
| 3.2.2 | Model II : Without residual/carry over effects | 3.6 |
| 3.2.3 | Problems of the cross-over design | 3.9 |
| 3.2.4 | Nonparametric analysis of the cross-over design | 3.10 |
| 3.2.5 | Formulating other hypotheses | 3.10 |
| 3.3 | Confidence intervals | 3.13 |
| 3.3.1 | Confidence intervals based on ANOVA | 3.14 |
| 3.3.2 | Westlake's modification of the ANOVA based confidence interval | 3.15 |
| 3.3.3 | Confidence intervals based on the paired t-test | 3.16 |
| 3.3.4 | Nonparametric confidence intervals based on Wilcoxon's signed rank tests | 3.17 |
| 3.3.5 | Nonparametric confidence intervals based on Pitman's permutation test | 3.19 |
| 3.4 | Bayesian approach to bioequivalence | 3.19 |
| 3.4.1 | Difference between formulation means without carry over effects | 3.20 |
| 3.4.2 | Difference between formulation means with carry over effects | 3.24 |
| 3.4.3 | Ratio of formulation means | 3.25 |
| 3.5 | Discussion | 3.26 |
| 3.6 | Recommendations | 3.28 |
| CHAPTER 4 | BOOTSTRAP METHOD | 4.1 |
| 4.1 | Introduction | 4.1 |
| 4.2 | The bootstrap approximation | 4.3 |
| 4.3 | Standard errors | 4.7 |
| 4.4 | Confidence intervals | 4.8 |
| 4.5 | The simple pivotal interval | 4.9 |
| 4.6 | Generalised pivotal intervals | 4.12 |
| 4.7 | Percentile intervals | 4.13 |
| 4.8 | Generalisation of the percentile interval | 4.13 |
| 4.9 | Bias-corrected percentile intervals | 4.14 |
| 4.10 | Discussion of bootstrap confidence intervals | 4.17 |
| 4.11 | Examples | 4.18 |
| 4.12 | Asymptotic accuracy of the bootstrap | 4.21 |
| 4.13 | Conditions for bootstrapping | 4.26 |

| | | Page |
|--------------|--|------|
| CHAPTER 5 | BOOTSTRAP APPROACH TO BIOEQUIVALENCE | 5.1 |
| 5.1 | Introduction | 5.1 |
| 5.2 | Assessing bioequivalence using intervals | 5.1 |
| 5.3 | Assumptions of bootstrap intervals | 5.3 |
| 5.4 | Choosing an interval | 5.4 |
| 5.5 | Algorithm for bias corrected interval | 5.6 |
| 5.6 | Example of bootstrap intervals | 5.10 |
| 5.7 | Wilcoxon and Pitman's procedures as resampling plans | 5.13 |
| 5.8 | Index of concordance | 5.16 |
| 5.9 | Example | 5.18 |
| 5.10 | Conclusions | 5.20 |
| CHAPTER 6 | BOOTSTRAPPING THE TOTAL RESPONSE TO A BIOAVAILABILITY STUDY | 6.1 |
| 6.1 | Introduction | 6.1 |
| 6.2 | Data, model and assumptions | 6.2 |
| 6.3 | The bootstrap algorithm | 6.4 |
| 6.4 | Results | 6.5 |
| 6.4.1 | Original data | 6.5 |
| 6.4.2 | Bootstrap distribution of (A, b_1, b_2) and $(AUC, TMAX, CMAX)$ | 6.7 |
| 6.5 | Conclusions | 6.17 |
| BIBLIOGRAPHY | | |
| APPENDIX | Computer program to compute bootstrap bias corrected interval for bioavailability parameters | |

ACKNOWLEDGEMENTS

The staff and postgraduate students of the Department of Mathematical Statistics at the University of Cape Town are a most approachable, friendly and enthusiastic group of statisticians. Whenever it was necessary to discuss a point, test an idea, or look for moral support their doors were open. Each discussion generated fresh enthusiasm and usually new ideas and perspectives. Working in this department as a postgraduate student has been a most stimulating experience.

The following colleagues deserve special mention:

- Associate Professor June Juritz for suggesting the topic and for supervising the research with such keen perception. Her contribution is strongly reflected in each chapter. Her enthusiasm and cheerful disposition are much appreciated.
- Dr Tertius de Wet, Dr Ross Sparks, Professor Theodor Stewart and Professor Walter Zucchini for their stimulating discussions, their interest, and their time.
- Professor Eon Smit, head of the Department of Statistics at the University of the Western Cape for organising a lecturing timetable which maximised time for research.
- Mrs Tib Cousins for her friendship, encouragement and serving as a general mentor. Her skilful and artistic typing has resulted in an excellent presentation.
- Mrs Minette Munitz for writing the computer program to compute the bootstrap confidence intervals discussed in Chapter 5.

- My wife Jill, who proofread the manuscript, checked the mathematics, wrote computer programs, sketched graphs and gave much encouragement.

When and if the degree is awarded it will be conferred in my name. In reality it should be awarded to the team of people listed above, all of whom have contributed to it to a greater or lesser extent.

A B S T R A C T

In 1984 it became legal for pharmacists to offer customers a cheaper generic alternative for a given prescription. The motivation for this was the excessively high cost of brand name drugs. The substitution of a generic alternative for a brand name drug is based on the assumption that drugs with a comparable chemical composition will have a similar therapeutic effect. The fact that this supposition is not always true has been demonstrated by a number of particular drugs, digoxon being perhaps the most vivid example.

The objective of this thesis is to review the statistical aspects associated with

- (i) measuring the bioavailability of a drug (Chapter 2)
- (ii) establishing the equivalence of a new and standard formulation of a drug (Chapter 3).

In the process of reviewing the literature two problems were identified. Firstly, it is commonly assumed that bioavailability parameters follow either the normal or lognormal distribution. This assumption is difficult to defend, hence procedures based on such assumptions became suspect. Secondly, bioavailability is inherently multivariate whereas in practice univariate procedures are employed.

Efron's bootstrap method, which does not rest on assumptions about the underlying distribution, is proposed as a tool for assessing bio-equivalence. A new measure of bioequivalence, the Index of Concordance,

is proposed. This index can be computed with equal ease for univariate or multivariate data using the bootstrap (Chapter 5).

The bootstrap idea of resampling the data can also be applied to compartmental modelling of bioavailability data. One result of this is a nonparametric estimate of the underlying distribution of the bioavailability parameters (Chapter 6).

The bootstrap is, on its own, a fascinating concept. A review of the bootstrap is given in Chapter 4.

CHAPTER 1

OVERVIEW OF BIOAVAILABILITY

1.1 Concept of bioavailability

When a drug is administered it undergoes numerous processes before it enters the systemic system from which it is eventually delivered to the site of action. For example, tablets must disintegrate and dissolve in the gastric juices. After dissolution the drug is absorbed through the gastro-intestinal wall into the gastro-intestinal portal blood. During this process the drug may be altered metabolically by the gastric juices and the blood. The drug and any metabolites that have formed are then transported by the blood to the liver which may alter the drug and its metabolites even further. From the liver the blood goes to the lungs where further biotransformation may take place. Only after all this has occurred does the drug and its metabolites reach the systemic circulation. Of course, the drug is being metabolised continually by the blood. Thus, the amount of unchanged drug that eventually reaches the systemic system is a fractional part of the original dose.

The concept of bioavailability (biological availability) describes the net result of this process and is used to define the rate and extent of drug appearance in the systemic circulation (Melander (1984)). However, the term bioavailability is often used as a shortened form of "comparative bioavailability" or "bioequivalence". Two formulations

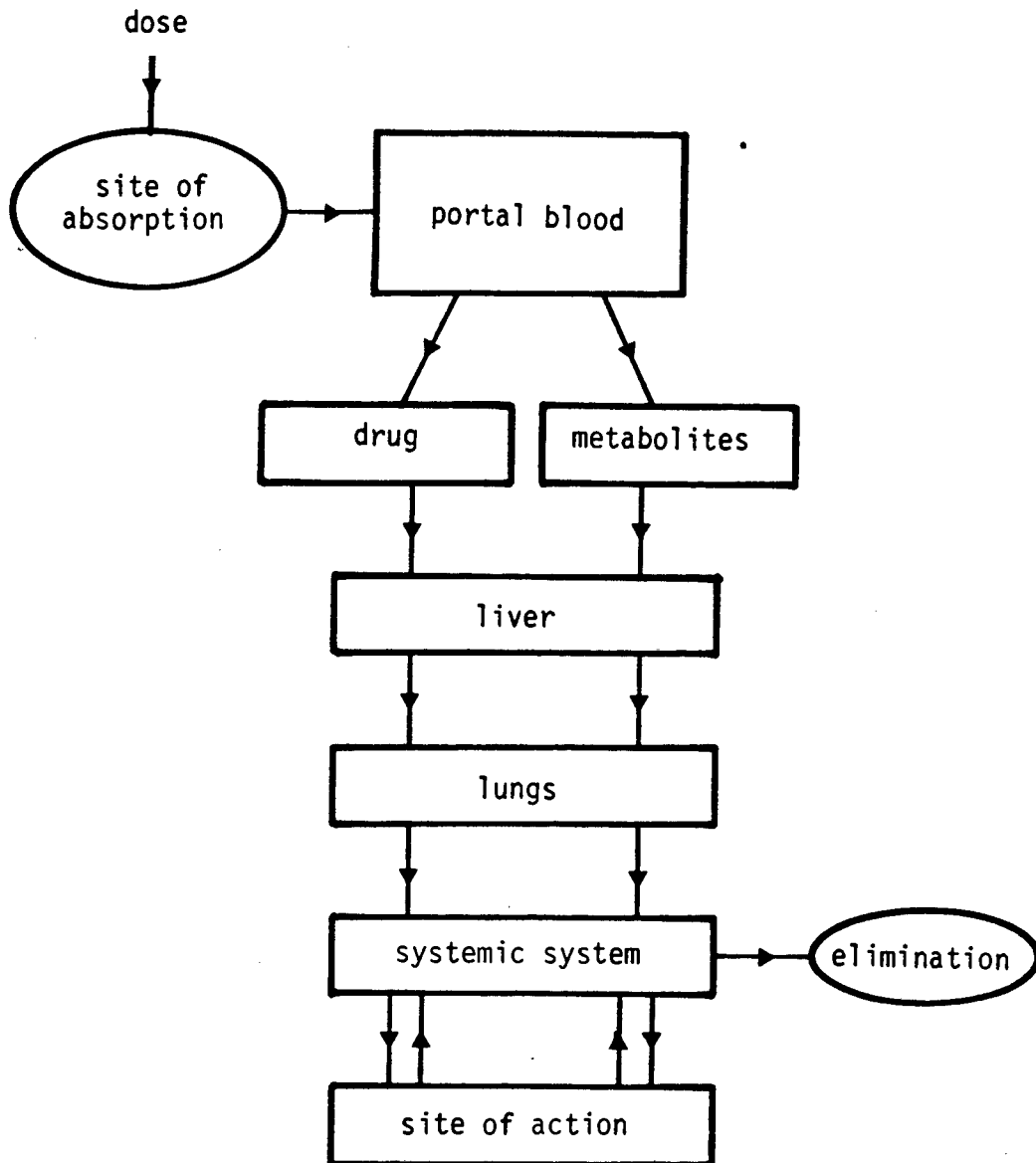


Figure 1.1 Schematic diagram of processes through which a drug, administered orally, must pass.

of the same drug are considered bioequivalent if they contain the same quantity of active drug and deliver this active drug to the circulating blood at the same rate and extent (Metzler (1974), Wagner (1975), Westlake (1979)). Of course, as Metzler (1974) points out the primary question in bioequivalence is *"whether two or more formulations containing the same active ingredient are therapeutically equivalent."*

1.2 Definitions and assumptions

Metzler (1974) suggests that three types of equivalents are involved:

chemical equivalents: drug products of the same dosage form which contain equal amounts of the same active ingredient as indicated by official standards;

biological equivalents: those chemical equivalents which deliver the same amount of active ingredient to the circulating blood. (Although he does not say so, one feels that the wording "at the same rate" should be added to this.)

therapeutic equivalents: those chemical equivalents which produce the same therapeutic effect as measured by the control of a symptom or disease.

He goes on to point out that chemical equivalents are not necessarily therapeutic equivalents, as was once assumed. Therapeutic equivalence of two chemically equivalent formulations can only be assessed via a clinical efficacy trial. Efficacy trials are both expensive and difficult to carry out, and bioavailability trials are an attempt to infer therapeutic equivalence without doing efficacy trials. The basic assumption of bioequivalence is one of continuity:

two formulations that have similar bioavailability characteristics will have similar therapeutic efficacy.

This assumption says that once the active ingredient is in the circulating blood, distribution, metabolism and excretion will not be influenced by formulation.

1.3 Measures of bioavailability

In order to be of any practical value the definition of bioequivalence must be interpreted in terms of some measurable quantity. The usual measurement is the level of the drug in the blood or serum from the time a single dose of the drug is taken until it has been completely metabolised or excreted (see Figure 1.2). Another measure is the amount of drug excreted in the urine.

According to Metzler (1974) the least controversial comparison of bioavailability would result if two average continuous concentration-time curves from a number of subjects were superimposable and had equal variability. However, in general continuous sampling is not possible and hence comparison is made on the basis of discrete sampling times denoted by t_1, \dots, t_k . If these are well chosen a good comparison can result.

And so we have

First interpretation of bioequivalence: Chemical equivalents that have essentially similar concentration-time profiles.

The inherent difficulty of sensibly comparing two sets of blood/serum profiles has focused attention on certain aspects of these profiles. The area under the concentration-time curve (AUC) is by far the most popular measure of bioavailability. The AUC is believed to be proportional to the total amount of active drug delivered to the systemic system. The maximum concentration (C_{MAX}) and time to maximum concentration (T_{MAX}) contain information about the rate and extent of absorption. More recently the half-life of the drug ($T_{\frac{1}{2}}$) has also been

shown to be important in the case of repeated dosages (Bruce (1984)), Greenblatt et al (1984)).

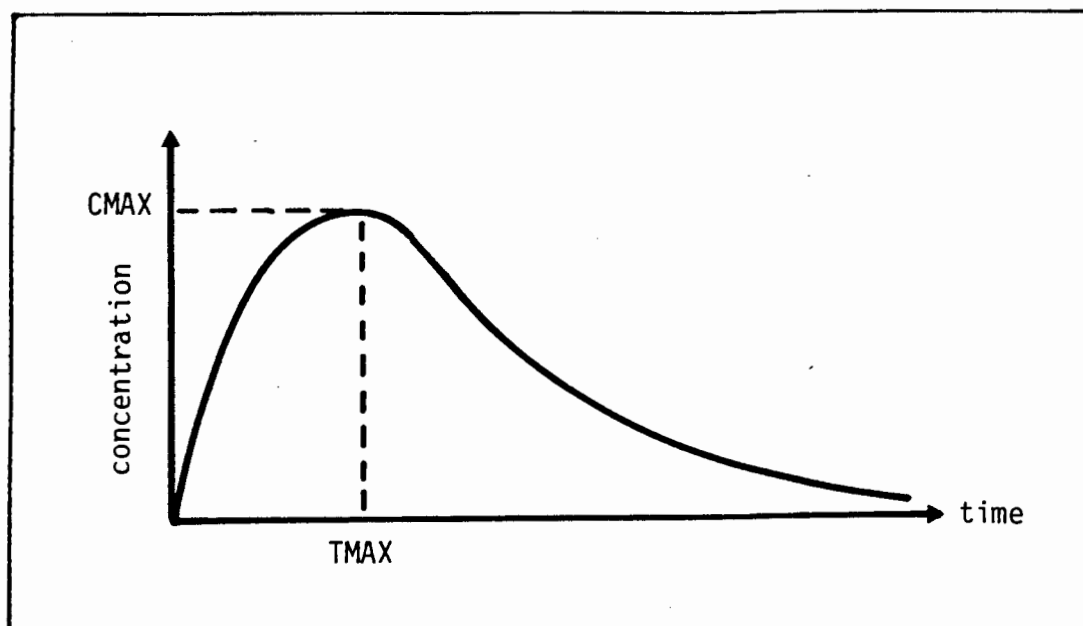


Figure 1.2 A typical concentration vs time curve showing the usual bioavailability parameters.

From this we obtain:

Second interpretation of bioequivalence: Chemical equivalents that have the same AUC, C_{MAX}, T_{MAX} and T_{1/2}. (This is based on the basic bioavailability assumption and the assumption that the four parameters AUC etc, adequately describe the concentration-time profiles.)

The great advantage of this interpretation is the reduction in dimensionality of the problem. In fact many authors consider that AUC is by far the most important bioavailability parameter (Melander (1984)) and, for many drugs, the only parameter of importance. Hence we have:

Third interpretation of bioequivalence: Chemical equivalents that have the same AUC.

The majority of statistical papers on bioequivalence are based on this interpretation of bioequivalence. They assume that bioequivalence will be assessed on the basis of one clinical parameter, usually AUC.

However, Westlake (1975) warns against the use of overall rules of thumb for deciding on bioequivalence. He proposes the general principle: *"Statistical tests should be carried out only on those characteristics of the blood-level sequence that have some meaningful relationship to the therapeutic use of the drug."*

He cites the following examples: The drug imipramine has an extremely delayed onset of action. As a result of this, TMAX has no clinical significance. Bioequivalence can only be judged on the basis of AUC. Another example is the drug Chlorpheniramine which has a very long elimination half-life. However therapeutic effect is dependent on repeated dosing at close intervals. Blood levels have little to do with therapeutic effect. Yet a third example are antibiotics that have a minimum inhibitory concentration for effective therapy. The two parameters of interest are

- (i) the time above the minimum inhibitory concentration and
- (ii) the time required to first reach this level (Westlake (1975)).

Wagner (1975 p.339) discusses fallacies that have arisen around the concept of bioavailability. He discusses factors that should be considered when assessing bioavailability and when extrapolating the results of a bioavailability study to other similar drugs or formulations.

Briefly these are

- (a) even if a drug passes official compendial standards this does not guarantee bioavailability in man,
- (b) differences in bioavailability will not necessarily be recognised in the clinical use of the drug, nor will differences necessarily be reported in the literature,
- (c) if two formulations of the same drug are shown to be bioequivalent it cannot be assumed that all formulations of the drug are bioequivalent,
- (d) bioavailability cannot be assessed from "in vitro" dissolution tests alone,
- (e) differences in bioavailability from one manufacturer's products to the next are at least as important as differences between the label dose,
- (f) bioavailability is not necessarily related to pharmacological effects or clinical response.

For this thesis we shall assume that the basic assumption of bioavailability (see earlier) holds good. We shall also adopt the definition given in Wagner (1975, p.340) of bioavailability : the extent and rate of absorption for a dosage form as reflected by the time-concentration curve of the administered dose in the systemic circulation. The term bioequivalents will mean chemical equivalents that have comparable bioavailabilities.

1.4 Pharmacokinetic Modelling

The absorption of a drug as described in section 1.1 is a complex process involving many independent variables. Pharmacokinetics deals essentially with the description of concentration changes of drugs in the blood as a function of time. Pharmacokinetic models provide a highly simplified, but useful mathematical description of the process.

These models view the absorption, distribution and elimination process as though occurring in a series of compartments with linear drug transfer rates between compartments. These compartmentalised models give rise to a system of linear differential equations whose solution are functions that are polyexponential in form

$$\text{i.e. } Y_j(t) = \sum_i c_{ij} \exp(-\lambda_{ij}t) \quad \text{for } j = 1, \dots, k$$

where $Y_j(t)$ = concentration in j th compartment at time t

k = number of compartments

c_{ij} , λ_{ij} are coefficients determined by the transfer rates and initial conditions.

Wagner (1975) gives an extensive account of one, two and three compartment models. He discusses models for intravenous injection, intravenous infusion and oral administration. He also considers the case where the drug is converted into a single metabolite as well as the case in which the drug is converted into a primary metabolite, which is turn is converted into a secondary metabolite and the drug and both metabolites are excreted in the urine. Over and above this, he also considers single and multiple dosing.

Steinijans (1975), Steyn and van Wyk (1977) and Greenblatt et al (1984) give a simple but useful account of one and two compartment models. These authors achieve simplification by combining two mechanisms of drug elimination namely metabolism and excretion under the label "elimination."

Following Steyn and van Wyk (1977) and Greenblatt et al (1984) we shall restrict attention to one and two compartment models and to single dosing. The three compartment model leads to a four term polyexponential function. Parameter estimates for this function become very unstable and therefore, useless. (Lanczos (1957), Hibbert and Steyn (1982).)

1.4.1 Two compartment model

A basic interpretation of the two compartment model supplies a useful abstraction of the absorption process. Since we will be concerned only with the blood/serum concentration of the drug and not its metabolites, we use the word elimination to include metabolism and excretion. This simplifies Wagner's classification somewhat but leads to a similar expression for blood levels. Our development follows that of Steinijans (1975), Steyn and van Wyk (1977) and Greenblatt et al (1984).

A drug dose (D) is introduced into the absorption site (usually the gastro-intestinal tract) at time $t = 0$ (Figure 1.3). The drug then passes into the central compartment (the blood) with absorption rate proportional to the concentration of the drug at the site of absorption. The constant of proportionality is denoted by k_a . Reversible drug distribution occurs between the central and peripheral compartments.

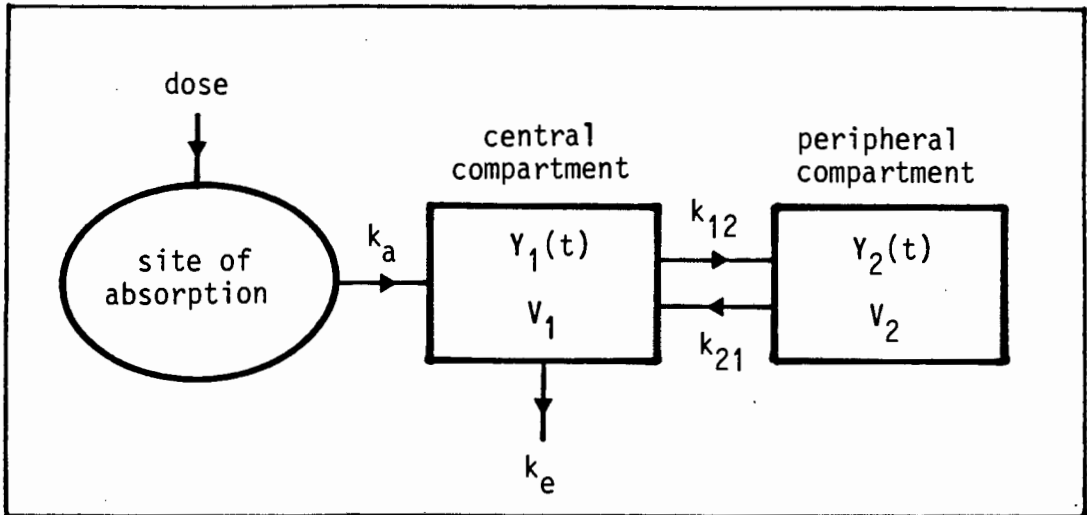


Figure 1.3 Schematic representation of the two compartment open model

The rate of distribution from the central to the peripheral compartment is denoted by k_{12} and that from the peripheral to the central compartment by k_{21} . Irreversible drug elimination takes place only from the central compartment (the blood) at rate k_e .

This model predicts that the concentration in the central compartment (i.e. blood/serum concentration) Y_1 will be a tri-exponential function of time (t) after dosing:

$$Y_1(t) = -(P_1 + P_2) \exp(-k_a t) + P_1 \exp(-\alpha t) + P_2 \exp(-\beta t)$$

where

$$\alpha = \frac{1}{2} [k_{12} + k_{21} + k_e] + \sqrt{(k_{12} + k_{21} + k_e)^2 - 4k_{21}k_e}$$

$$\beta = \frac{1}{2} [k_{12} + k_{21} + k_e] - \sqrt{(k_{12} + k_{21} + k_e)^2 - 4k_{21}k_e}$$

$$P_1 = \frac{k_a FD}{V_1} \cdot \frac{k_{21}^{-\alpha}}{(k_a - \alpha)(\beta - \alpha)}$$

$$P_2 = \frac{k_a FD}{V_1} \cdot \frac{k_{21}^{-\beta}}{(k_a - \beta)(\alpha - \beta)}$$

F = fraction of dose absorbed¹

D = administered dose

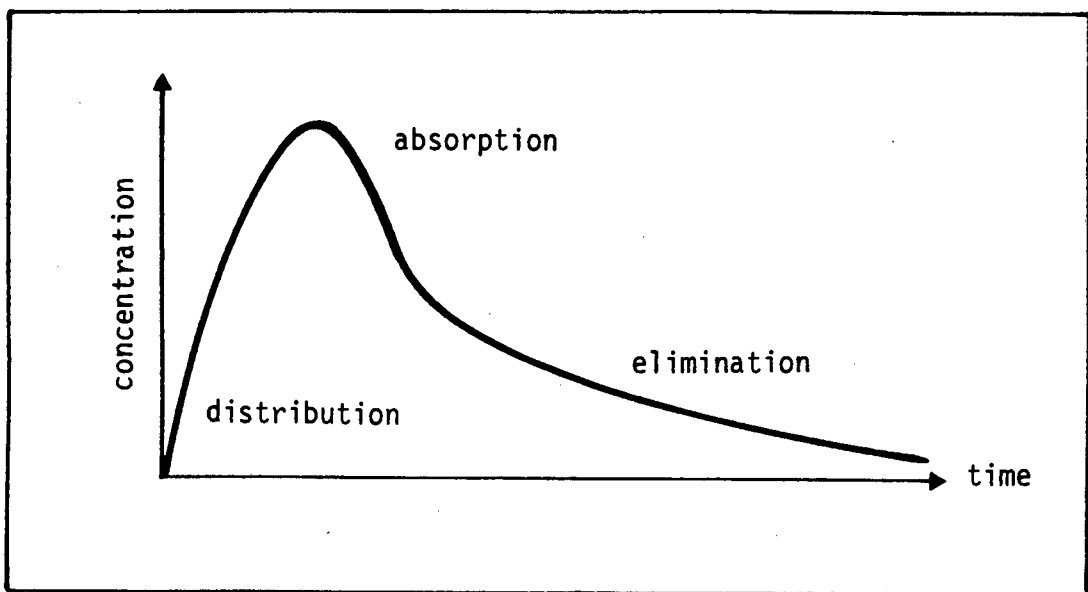
V_1 = volume of compartment no 1.

A semi-logarithmic plot of $Y_1(t)$ versus t (Figure 1.4) will reveal three distinct phases that correspond to

phase 1 - absorption of drug into central compartment or blood,

phase 2 - distribution of drug from central to peripheral compartment,

phase 3 - elimination of drug.



$$Y_1(t) = \frac{F D}{V_1} \left[\frac{k_{12}}{k_{21} - k_a} \exp(-k_a t) + \frac{k_a}{k_{21} - k_a} \exp(-k_{21} t) + \frac{k_{21}}{k_{21} - k_a} \exp(-k_a t) - \frac{k_{12}}{k_{21} - k_a} \exp(-k_{21} t) \right]$$

Figure 1.4 Blood/serum concentration curve plotted on semi-logarithmic scale for the two compartment open model.

Except for AUC the bioavailability parameters C_{MAX} , T_{MAX} and $T_{1/2}$ cannot be expressed in closed form and will have to be found numerically.

1. This is the classical measure of bioavailability.

$$\text{AUC} = \int_0^{\infty} Y_1(t) dt = \frac{P_1 + P_2}{k_a} + \frac{P_1}{\alpha} + \frac{P_2}{\beta}$$

1.4.2 One compartment model

If the distribution process occurs much more rapidly than the absorption and elimination, then the body may be considered as a single homogeneous compartment. In this instance there is no need for a peripheral compartment (Figure 1.5).

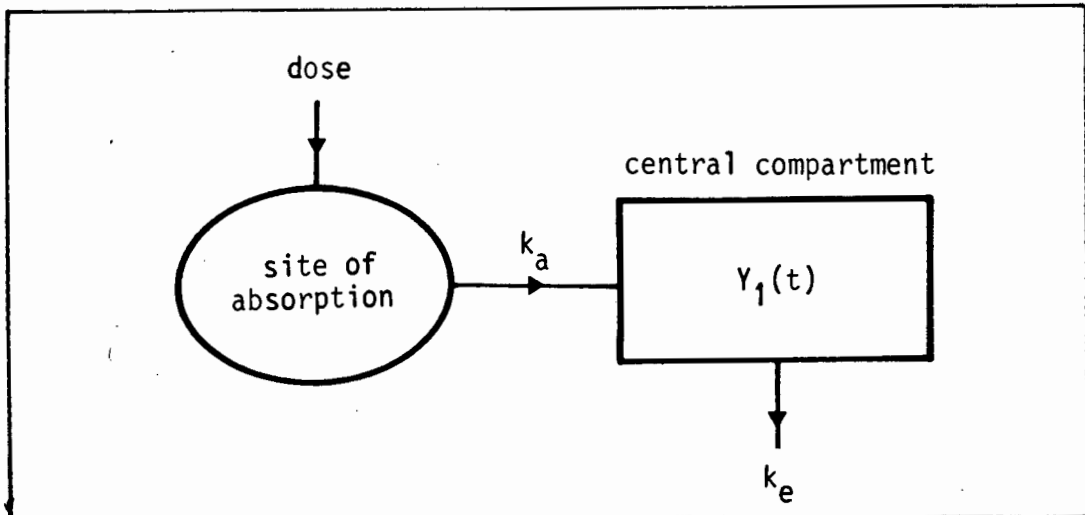


Figure 1.5 Schematic representation of one compartment open models

Again it is assumed that the dose is introduced to the site of absorption as a bolus at time $t = 0$. The drug is absorbed into the central compartment with absorption rate constant k_a which is assumed to be first order. Drug elimination is also first order with rate constant k_e .

According to this model the predicted concentration $Y_1(t)$ at time t after dosing is

$$Y_1(t) = FD \left(\frac{k_a}{k_a - k_e} \right) \{ \exp(-k_e t) - \exp(-k_a t) \}$$

where F = fraction of dose absorbed

D = administered dose

A semi-logarithmic plot of the blood/serum concentration has two phases:

phase 1 - increasing concentration that corresponds to absorption

phase 2 - decreasing concentration that corresponds to elimination.

This is shown in Figure 1.6

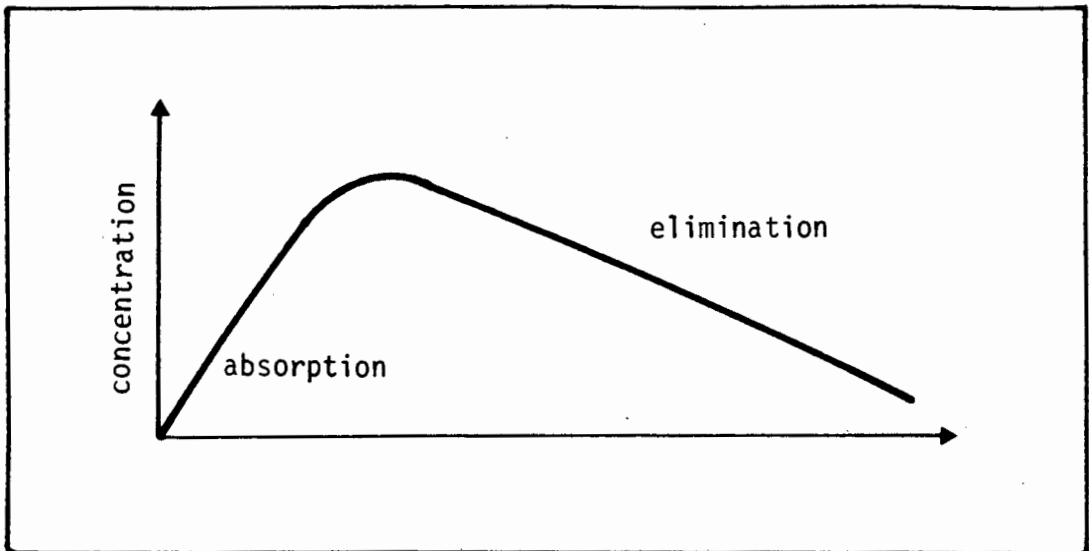


Figure 1.6 Blood/serum concentration curves corresponding to a one compartment open model

The bioavailability parameters are easily expressible in closed form:

$$AUC = \int_0^{\infty} Y_1(t) dt = FD \left(\frac{k_a}{k_a - k_e} \right) \left(\frac{1}{k_e} + \frac{1}{k_a} \right)$$

$$TMAX = (\ln k_a - \ln k_e) / (k_a - k_e)$$

$$C_{MAX} = Y(TMAX)$$

1.5 Statistical Considerations

Bioavailability trials consist of two types, depending on the objective: One may be interested in assessing the absolute bioavailability of a new but unknown drug or in the comparative bioavailability of a new formulation/preparation of a generic alternative. In either case, there are a number of interrelated problems of a statistical nature.

In absolute bioavailability studies one is attempting to assess the rate and extent of absorption of a drug and/or its active metabolites. Such a study would typically involve a panel of subjects (from six to twenty four). These subjects are screened medically and, under controlled conditions, given the drug dosage form. Blood samples are taken from each individual at predetermined sampling times t_1, \dots, t_k after application. The extent and rate of absorption are then estimated from these blood concentration profiles.

In comparative bioavailability studies one is attempting to assess the relative magnitudes of the rate and extent of absorption of a standard product and one or more test products. A guideline protocol for comparative bioavailability trials has been devised by the Food and Drug Administration (FDA). Wagner (1975, p.353) gives a draft of the proposed guidelines but warns that this draft should not be construed as a summary of the guidelines. The draft is comprehensive covering ten sections:

Title; Names of investigators; Synopsis; Background information; Objectives of the study; Clinical facilities available; Institutional

review; Experimental plans; References and Appendices.

Section VIII on Experimental Plans is relevant to the statistical aspects. The Experimental Plans cover the following topics:

(A) Subject selection. This covers who the subjects will be and how they will be selected; Age range, body weight range etc; Clinical screening tests that are to be given; Exclusion of concurrent medication.

(B) The drug. Various details about the drug are to be recorded. For example, tests used to characterise the drug, etc.

(C) Control or Reference: Various details, similar to that of the test drug, are to be recorded.

(D) Treatment schedule and Doses: A table should be given which clearly indicates the treatment schedule; the assignment of subjects to treatments; the number of subjects; fasting conditions and the time of administration of the doses in relation to food intake; the level of activity of the patients; the amount of water to be ingested during each test day.

Wagner (1975, p.354) quotes the FDA draft as saying: *"The number of subjects employed in the test should be adequate to demonstrate the lack of a statistically significant practical difference among the bio-availability parameters selected with an alpha of 0.05 and a beta of 0.20."* The draft also stated: *"Generally 12 to 20 subjects are sufficient for bioavailability studies."*

(E) Observations: Specify when the blood and/or urine is to be sampled i.e. specify t_1, \dots, t_k ; how these samples are to be treated, stored

and transported. Statements as to how side effects and intolerance are to be evaluated and reported.

(F) Assay Method(s): Details of assay methods to be used to analyse blood and/or urine should be stated.

(G) Data Analysis: Method of data analysis should be specified. Statistical considerations should be clearly stated. Computer programs to be used should be specified and referenced.

It is from these data that the bioavailability parameters AUC, etc, are calculated and bioequivalence assessment is made. The statistical considerations involved and methods used will be the subject matter of chapters 2 and 3. Chapter 4 is an account of the statistical method called the bootstrap and methods for deriving bootstrap confidence intervals. In chapter 5 we propose a new method for assessing bioequivalence using the bootstrap.

CHAPTER 2

ESTIMATING THE BIOAVAILABILITY
PARAMETERS2.1 Introduction

The result of a bioavailability trial, either absolute or comparative, is a set of data for each of n individuals consisting of sampling times t_1, \dots, t_k and corresponding blood/serum concentrations $y_i(t_1), \dots, y_i(t_k)$ $i = 1, \dots, n$. A typical set of data is given in Table 2.1 from Button (1979).

The data analysis proceeds in two stages. Firstly the data from each individual are used to estimate the pharmacokinetic parameters AUC, TMAX, CMAX and $T_{1/2}$, or some subset of these. The individual estimates obtained in the first step are then used to make inference about the population parameters.

In this chapter the focus will be on the first stage of the data analysis. We consider data $(t_0, y_0), (t_1, y_1), \dots, (t_k, y_k)$ for an individual and our task is to estimate the pharmacokinetic parameters for that individual.

Table 2.1 Six horses received 15 mg/kg theophylline as aminophylline by intragastric administration. Theophylline assay by high performance liquid chromatography. Concentrations in $\mu\text{g}/\text{m}\ell$

| Time | h o r s e | | | | | | mean | ± | SD |
|-------|-----------|------|------|------|------|------|------|---|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | | | |
| 0,166 | | 4,2 | 10,1 | 3,7 | 12,6 | 11,9 | 7,1 | ± | 5,1 |
| 0,333 | 1,5 | 15,1 | 14,8 | 11,5 | 18,1 | 16,3 | 12,9 | ± | 6,0 |
| 0,5 | 6,8 | 17,5 | 19,9 | 19,7 | 20,9 | 18,3 | 17,2 | ± | 5,2 |
| 0,666 | 14,0 | 19,5 | 22,1 | 21,4 | 22,0 | 19,3 | 19,7 | ± | 3,1 |
| 1 | 16,5 | 20,3 | 20,8 | 23,9 | 20,3 | 20,9 | 20,5 | ± | 2,4 |
| 1,5 | 21,3 | 22,9 | 20,3 | 21,8 | 19,7 | 19,0 | 20,8 | ± | 1,4 |
| 2 | 19,8 | 20,4 | 19,7 | 18,9 | 20,1 | 18,2 | 19,5 | ± | 0,8 |
| 2,5 | 18,2 | 19,1 | 18,9 | 18,7 | 18,4 | 17,0 | 18,4 | ± | 0,8 |
| 3 | 17,3 | 17,7 | 17,3 | 16,2 | 17,7 | 16,2 | 17,1 | ± | 0,7 |
| 4 | 15,6 | 16,8 | 16,1 | 15,5 | 16,1 | 14,3 | 15,7 | ± | 0,8 |
| 5 | 14,6 | 15,5 | 15,0 | 14,4 | 15,4 | 13,9 | 14,8 | ± | 0,6 |
| 6 | 13,3 | 14,6 | 14,2 | 13,4 | 14,8 | 12,9 | 13,9 | ± | 0,8 |
| 8 | 11,6 | 13,6 | 13,2 | 13,0 | 13,2 | 11,9 | 12,8 | ± | 0,8 |
| 10 | 11,2 | 12,1 | 12,3 | 12,0 | 12,4 | 10,6 | 11,8 | ± | 0,7 |
| 12 | 10,0 | 10,6 | 10,8 | 11,6 | 11,1 | 10,2 | 10,7 | ± | 0,6 |
| 24 | 5,5 | 6,9 | 6,5 | 7,0 | 6,3 | 5,7 | 6,3 | ± | 0,6 |
| 30 | 4,4 | 4,6 | 4,6 | 5,1 | 4,8 | 4,8 | 4,7 | ± | 0,2 |
| 48 | 2,5 | 2,1 | 1,7 | 2,2 | 1,9 | 1,9 | 2,1 | ± | 0,3 |

2.2 Direct estimation of the AUC

The AUC may be estimated

- (i) directly by numerical integration, or
- (ii) derived from the estimated parameters of a compartmental model or some curve fitted to the concentration-time data. In this section we will consider estimating the AUC directly by numerical integration. Deriving the AUC from estimated parameters will be considered in section 2.3.

In performing a numerical integration the true curve between two successive concentrations y_j and y_{j+1} is approximated by some simple function such as a straight line, polynomial or exponential. Four algorithms are commonly used and differ mainly in the form of the approximating function. The algorithms, which will be discussed in more detail below are

- (i) the Trapezoidal rule
- (ii) the Log-trapezoidal rule
- (iii) the Lagrange method
- (iv) Cubic splines

Numerical integration methods only provide the $AUC(0, t_k)$ - the area under the curve from time zero to the final observation time, t_k . The total $AUC(0, \infty)$ must be found by some extrapolation method. This will be discussed in section 2.2.5.

2.2.1 The Trapezoidal Rule

The simplest method of estimating $AUC(0, t_k)$ numerically is to join the

points $(t_0, y_0), (t_1, y_1), \dots, (t_k, y_k)$ by a series of straight lines.

The resulting figure will be a polygon consisting of k trapezia.

The total area under this polygon from time $t = t_0 = 0$ to $t = t_k$ is then the sum of the areas of the k trapezia.

$$\hat{AUC}(0, t_k) = \sum_{j=0}^{k-1} (y_j + y_{j+1})(t_{j+1} - t_j) / 2 \quad (2.1)$$

2.2.2 The Log-trapezoidal rule

The Log-trapezoidal method approximates y between any two observations y_{i-1} and y_i using a single exponential. This is equivalent to approximating $\ln y$ by a straight line between successive points.

Interpolating on the interval (t_{i-1}, t_i) we have

$$\ln y = \ln y_{i-1} + (t - t_{i-1}) \ln (y_i / y_{i-1}) / (t_i - t_{i-1})$$

or alternately

$$y = y_{i-1} \exp\{-(t - t_{i-1}) \ln (y_i / y_{i-1}) / (t_i - t_{i-1})\} \quad (2.2)$$

Integrating (2.2) we obtain

$$\begin{aligned} \hat{AUC}(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} y \, dt \\ &= (y_i - y_{i-1})(t_i - t_{i-1}) / \ln (y_i / y_{i-1}) \end{aligned}$$

This log-trapezoidal method is best suited for data that is monotonically decreasing. It cannot be used if any observed y value is zero or if two consecutive values are equal. Furthermore the method may produce large errors when used on ascending curves, near a peak or on a steeply descending polyexponential curve (Yeh and Kwan (1978)).

2.2.3 Lagrange method

In the Lagrange method interpolation is achieved using cubic polynomials. To interpolate in the interval (t_{i-1}, t_i) the equation

$$y = a_i + b_i t + c_i t^2 + d_i t^3 \quad (2.3)$$

is fitted to the four adjacent data points (t_{i-2}, y_{i-2}) , (t_{i-1}, y_{i-1}) , (t_i, y_i) and (t_{i+1}, y_{i+1}) . The four coefficients a_i , b_i , c_i and d_i may be obtained by solving the following system of linear equations:

$$\begin{pmatrix} 1 & t_{i-2} & t_{i-2}^2 & t_{i-2}^3 \\ 1 & t_{i-1} & t_{i-1}^2 & t_{i-1}^3 \\ 1 & t_i & t_i^2 & t_i^3 \\ 1 & t_{i+1} & t_{i+1}^2 & t_{i+1}^3 \end{pmatrix} \begin{pmatrix} a_i \\ b_i \\ c_i \\ d_i \end{pmatrix} = \begin{pmatrix} y_{i-2} \\ y_{i-1} \\ y_i \\ y_{i+1} \end{pmatrix} \quad (2.4)$$

The area under the concentration-time curve between t_{i-1} and t_i is estimated by integrating (2.3) over this interval, to give

$$\begin{aligned} \hat{AUC}(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} (a_i + b_i t + c_i t^2 + d_i t^3) dt \\ &= a_i(t_i - t_{i-1}) + b_i(t_i^2 - t_{i-1}^2)/2 + c_i(t_i^3 - t_{i-1}^3)/3 \\ &\quad + d_i(t_i^4 - t_{i-1}^4)/4 \end{aligned} \quad (2.5)$$

Equation (2.4) can be applied serially for each $i = 2, 3, \dots, n-1$ but not for the two end intervals (t_0, t_1) and (t_{k-1}, t_k) . For these two intervals, the nearest three points are used to fit a parabola

$$y = a_i + b_i t + c_i t^2 \quad (2.6)$$

The three coefficients a_i , b_i and c_i are calculated by solving a system of three simultaneous linear equations, analogous to equation

(2.4). The corresponding areas are obtained by integrating equation (2.6).

The cumulative area $\hat{AUC}(0, t_k)$ is then computed by summation

$$\hat{AUC}(0, t_k) = \sum_{i=1}^k \hat{AUC}(t_{i-1}, t_i) \quad (2.7)$$

2.2.4 Cubic Spline Method

The cubic spline method is similar to the method of Lagrange in that interpolation is achieved using cubic polynomials. However there is an additional constraint of differentiability at each data point.

General spline functions are defined as piecewise polynomials of degree k , connected at several knots, such that the fitted curve and its first $k-1$ derivatives are continuously differentiable. For cubic splines k is defined to be 3 and the knots are taken to be the data points themselves.

The derivation presented below follows that of Dunfield and Read (1972) and Yeh and Kwan (1978).

The cubic polynomial in equation (2.3) is differentiated three times to give

$$y' = b_i + 2c_i t + 3d_i t^2 \quad (2.8)$$

$$y'' = 2c_i + 6d_i t \quad (2.9)$$

$$y''' = 6d_i \quad (2.10)$$

From equation (2.9) it is evident that y'' is linear over each interval $[t_{i-1}, t_i]$. Because of the linearity it may be rewritten in the following form

$$y'' = y''_{i-1}(t_i-t)/h_i + y''_i(t-t_{i-1})/h_i \quad (2.11)$$

where $h_i = t_i - t_{i-1}$. This equation is integrated twice to give

$$y' = -y''_{i-1}(t_i-t)^2/2h_i + y''_i(t-t_{i-1})^2/2h_i + s_1 \quad (2.12)$$

$$y = y''_{i-1}(t_i-t)^3/6h_i + y''_i(t-t_{i-1})^3/6h_i + s_1t + s_2 \quad (2.13)$$

where s_1 and s_2 are constants of integration. These constants are determined by evaluating equation (2.13) at t_{i-1} and t_i , giving

$$y_{i-1} = y''_{i-1}h_i^2/6 + s_1t_{i-1} + s_2 \quad (2.14)$$

$$y_i = y''_i h_i^2/6 + s_1t_{i-1} + s_2 \quad (2.15)$$

Solving these two equations for s_1 and s_2 gives

$$s_1 = (y_i - y_{i-1})/h_i - h_i(y''_i - y''_{i-1})/6 \quad (2.16)$$

$$s_2 = (t_i y_{i-1} - y_i t_{i-1})/h_i - h_i(t_i y''_{i-1} - y''_i t_{i-1})/6 \quad (2.17)$$

All quantities in equation (2.13) are known except for y''_{i-1} and y''_i .

These values are determined as follows:

Equation (2.12) is evaluated at t_{i-1} from the interval $[t_{i-1}, t_i]$ and from the interval $[t_{i-2}, t_{i-1}]$ to give the following two equations, respectively:

$$y'_{i-1} = -y''_{i-1}h_i/2 + (y_i - y_{i-1})/h_i - h_i(y''_i - y''_{i-1})/6 \quad (2.18)$$

$$y'_{i-1} = -y''_{i-1}h_{i-1}/2 + (y_{i-1} - y_{i-2})/h_{i-1} - h_{i-1}(y''_{i-1} - y''_{i-2})/6 \quad (2.19)$$

Combining equations (2.18) and (2.19) and rearranging gives

$$\begin{aligned} h_{i-1}y''_{i-2}/6 + (h_i + h_{i-1})y''_{i-1}/3 + h_i y''_i/6 &= (y_i - y_{i-1})/h_i \\ &- (y_{i-1} - y_{i-2})/h_{i-1} \end{aligned} \quad (2.20)$$

Equation (2.20) can be applied to all intervals except the first and last. This will yield $n-1$ equations. Since there are $n+1$ unknowns two additional equations are required. These are obtained by specifying two extra conditions. In the present case these are $y_1''' = y_2'''$ and $y_{n-1}''' = y_n'''$. The third derivatives are given by equation (2.10).

$$y_1''' = (y_1'' - y_0'')/h_1 \quad (2.21)$$

$$y_2''' = (y_2'' - y_1'')/h_2 \quad (2.22)$$

$$y_{n-1}''' = (y_{n-1}'' - y_{n-2}'')/h_{n-1} \quad (2.23)$$

$$y_n''' = (y_n'' - y_{n-1}'')/h_n \quad (2.24)$$

Combining equations (2.21) and (2.22) and equations (2.23) and (2.24), respectively, gives

$$y_0''/h_1 - (h_0^{-1} + h_1^{-1}) y_1'' + y_2''/h_2 = 0 \quad (2.25)$$

$$y_{n-2}''/h_{n-1} - (h_{n-1}^{-1} + h_n^{-1}) y_{n-1}'' + y_n''/h_n = 0 \quad (2.26)$$

The $n+1$ unknowns y_0'', \dots, y_n'' may be obtained by solving (2.20), (2.25) and (2.26) simultaneously. Once these are known equation (2.13) may be integrated over each interval $[t_{i-1}, t_i]$ to give

$$\begin{aligned} \hat{AUC}(t_{i-1}, t_i) &= \int_{t_{i-1}}^{t_i} y \, dt \\ &= h_i^3 (y_i'' + y_{i-1}'')/24 + h_i [s_1(t_i + t_{i-1})/2 + s_2] \end{aligned} \quad (2.27)$$

The cumulative area from time $t = 0$ to $t = t_k$ is

$$\hat{AUC}(0, t_k) = \sum_{i=1}^k \hat{AUC}(t_{i-1}, t_i) \quad (2.28)$$

2.2.5 Extrapolating the AUC

Four methods have been given estimating $AUC(0, t_k)$ directly. However, in bioavailability studies it is the total area $AUC(0, \infty)$ from zero time to time infinity that is of interest (Wagner 1975, p.344).

Using any method $AUC(0, t_k)$ underestimates $AUC(0, \infty)$. But without making assumptions about the form of the concentration-time curve it is not possible to correct or extrapolate $AUC(0, t_k)$ to give an estimate of the desired quantity $AUC(0, \infty)$.

If one is willing to assume a compartmentalised model with linear transfer rates, then the concentration-time curve will be polyexponential in form. Suppose for example this is

$$\begin{aligned} y(t) &= A \left(e^{-k_e t} - e^{-k_a t} \right) & k_a > k_e \\ &= A e^{-k_e t} - A e^{-k_a t} \end{aligned} \quad (2.29)$$

Since $k_a > k_e$, $e^{-k_a t}$ approaches zero more rapidly than does $e^{-k_e t}$. If the ratio k_a/k_e is sufficiently large then for large t we effectively have

$$y(t) \approx A e^{-k_e t} \quad (2.30)$$

The unobserved area from t_k to ∞ can now be approximated by the area under the curve defined by (2.30) between t_k and ∞ . Denoting this unobserved area by $AUC(t_k, \infty)$ we have

$$\begin{aligned} AUC(t_k, \infty) &\approx \int_{t_k}^{\infty} A e^{-k_e t} dt \\ &= A e^{-k_e t_k / k_e} \\ &\approx y_k / k_e \end{aligned} \quad (2.31)$$

A log-linear plot of the data will achieve two objectives : it will display whether or not (2.30) provides a reasonable description of the data for large t . If (2.30) is reasonable then the last few data points will lie approximately on a straight line. Secondly, the slope of the line through these last few points is an estimate of $-k_e$. This may be obtained either by eye or by linear regression of $\log y_j$ against t_j .

Hence we have an estimate of the unobserved area

$$\hat{AUC}(t_k, \infty) = y_k / \hat{k}_e \quad (2.32)$$

Using (2.32) as a correction term we have an estimate of the total $AUC(0, \infty)$

$$\hat{AUC}(0, \infty) = \hat{AUC}(0, t_k) + \hat{AUC}(t_k, \infty)$$

where $\hat{AUC}(0, t_k)$ is obtained by any one of the four methods that have been described earlier.

Although the concentration-time curve was assumed to be a sum of two exponentials to obtain the correction term (2.32) a similar argument applies to any polyexponential function, provided that a log-linear plot of the data lie eventually on a straight line.

2.2.6 Discussion on numerical integrating algorithms

Yeh and Kwan (1978) have, through a series of five simulation experiments, tested the relative merits of the four algorithms presented above on various types of data. They considered simulated sets of data

- (i) that vary linearly between points without error,
- (ii) exponentially decreasing data without error,
- (iii) from a two compartment open model without error,
- (iv) from a two compartment open model with error introduced in various ways.

In conclusion Yeh and Kwan suggest that if the data are functionally smooth and error free the Lagrange and cubic spline method will give the best approximation to the system. However since errors are experimentally inevitable, the superiority afforded by these methods becomes less certain, and this uncertainty may increase with increasing noise in the data.

Both the Lagrange and cubic spline methods may produce spurious and unrealistic oscillations and need to be monitored.

Although the two trapezoidal methods are less accurate they may be the logical choice because of their simplicity. They are particularly suitable when estimates of AUC are the data that will be used for testing bioequivalence.

2.3 Fitting a polyexponential model

For numerical integration we do not assume any particular model for the concentration-time curve, except to derive the correction term.

From the theory of compartmental models many concentration-time curves can be represented as a polyexponential function

$$y = \sum_{i=1}^p a_i e^{-b_i t}$$

If this approach is taken some estimation procedure is used to find or estimate the coefficients a_1, \dots, a_p and b_1, \dots, b_p . The pharmacokinetic parameters AUC etc., are then calculated from the fitted equation.

The methods considered in this thesis are

- (i) The graphical method
- (ii) Least squares
- (iii) Weighted least squares
- (iv) Maximum likelihood

2.3.1 The Graphical method

The graphical method of finding estimates has been well discussed and documented by Gurrupide et al. (1964), Atkins (1969 pp.101-106), Foss (1969), Wagner (1975) and Steyn and van Wyk (1977). The method has also been given various other names; the "stripping" procedure (Wagner); the "back-projection" technique (Wagner); the "peeling-off" technique (Foss).

The method is based on the observation that a polyexponential function

$$\begin{aligned} \text{e.g. } y &= A_1 e^{-b_1 t} + A_2 e^{-b_2 t} + A_3 e^{-b_3 t} & (2.33) \\ &= y_1 + y_2 + y_3 \\ &\text{with } 0 < b_3 < b_2 < b_1 \end{aligned}$$

will, for large values of t , behave approximately as

$$y_3 = A_3 e^{-b_3 t} \quad (2.34)$$

if b_1 , b_2 and b_3 are sufficiently separated. This can be detected by plotting $\ln y$ vs t . If this plot is approximately linear for large

values of t then by taking logarithms we obtain

$$\ln y \simeq \ln y_3 = \ln A_3 - b_3 t \quad (2.35)$$

Using the last few "linear points" as data the slope ($-b_3$) and intercept ($\ln A_3$) can be estimated either by eye or by linear regression. From these estimates we obtain

$$\hat{y}_3 = \hat{A}_3 e^{-\hat{b}_3 t} \quad (2.36)$$

Subtracting (2.36) from (2.33) we obtain

$$y - \hat{y}_3 \simeq A_1 e^{-b_1 t} + A_2 e^{-b_2 t} \quad (2.37)$$

Since $0 < b_2 < b_1$ (2.37) will, for "large" values of t behave as

$$y - \hat{y}_3 \simeq A_2 e^{-b_2 t} \quad (2.38)$$

This can be detected by plotting $\ln(y - \hat{y}_3)$ vs t . If this plot is approximately linear for "large" t then approximately

$$\ln(y - \hat{y}_3) \simeq \ln A_2 - b_2 t \quad (2.39)$$

Again, using the last few points as data, one can obtain estimates of A_2 and b_2 . We have

$$\hat{y}_2 = \hat{A}_2 e^{-\hat{b}_2 t} \quad (2.40)$$

Subtracting (2.40) from (2.37) we obtain

$$y - \hat{y}_3 - \hat{y}_2 \simeq A_1 e^{-b_1 t} \quad (2.41)$$

A plot of $\ln(y - \hat{y}_3 - \hat{y}_2)$ vs t should be approximately linear and yield estimates of A_1 and b_1 .

Examples are given in Foss (1969), Wagner (1975) and Steyn and van Wyk (1977).

2.3.2 Least squares

Consider the statistical model

$$y_i = f(t_i, \underline{\theta}) + \varepsilon_i \quad (2.42)$$

where y_i represents the blood/serum concentration at time t_i

$\underline{\theta}$ is a vector of m unknown model parameters which are to be estimated

$f(t_i, \underline{\theta})$ is the model or response function that represents the assumed functional relationship between t , $\underline{\theta}$ and y .

For our purposes $f(t_i, \underline{\theta})$ is a polyexponential function.

The least squares estimate of $\underline{\theta}$ is defined as that vector $\hat{\underline{\theta}}$ which minimises the objective function

$$S(\underline{\theta}) = \sum_{i=1}^k \{y_i - f(t_i, \underline{\theta})\}^2 \quad (2.43)$$

Provided that $S(\underline{\theta})$ is differentiable the least squares estimate $\hat{\underline{\theta}}$ is the solution of the so called normal equations

$$\nabla S(\underline{\theta}) = \begin{pmatrix} \frac{\partial S}{\partial \theta_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial S}{\partial \theta_m} \end{pmatrix} = \underline{0} \quad (2.43a)$$

where

$$\frac{\partial S}{\partial \theta_j} = -2 \sum_{i=1}^k \{y_i - f(t_i, \underline{\theta})\} \frac{\partial f}{\partial \theta_j} \quad j = 1, \dots, m \quad (2.43b)$$

Note that if $f(t, \underline{\theta})$ is nonlinear in the parameters $\underline{\theta}$, then the system of equations (2.43a) can only be solved iteratively.

Algorithms for computing $\hat{\underline{\theta}}$ are described in section 2.4.

2.3.3 Weighted least squares

Let w_1, \dots, w_k be a set of non-negative numbers, usually called weights. The number w_i reflects the relative contribution of the i th observation (t_i, y_i) to the estimate of $\underline{\theta}$. The weighted least squares estimate corresponding to $\underline{w}' = (w_1, \dots, w_k)$ is defined as that vector $\hat{\underline{\theta}}(\underline{w})$ which minimises the objective function

$$S(\hat{\underline{\theta}}(\underline{w})) = \sum_{i=1}^k w_i \{y_i - f(t_i, \hat{\underline{\theta}})\}^2 \quad (2.44)$$

Setting $w_i = 1$, $i = 1, \dots, k$ produces the ordinary least squares estimate.

A common choice of w_i is

$$w_i = 1/\sigma_i^2$$

where σ_i^2 is the variance of y_i . Brownlee (1960), Wagner (1975, pp.288-289) suggests the following weighting factors

$$w_i = 1/y_i$$

or alternately

$$w_i = 1/y_i^2$$

The problem of choosing the weights w_i is by no means settled. As Wagner mentions the choice $w_i = 1$ will result in the terminal concentrations having almost no effect on $\underline{\theta}$. This is because the residuals $y_i - f(t_i, \underline{\theta})$ are extremely small for terminal values of t . However, any of the choices $1/\sigma_i^2$, $1/y_i$ or $1/y_i^2$ result in the terminal con-

centrations carrying more and more weight. The smaller the concentration the more weight it will have. Concentrations of zero have infinite weight. So these choices also seem unsatisfactory.

2.3.4 Maximum Likelihood

If one is prepared to make assumptions about the distributional structure of the error terms in the model then one can estimate the unknown parameter $\underline{\theta}$ using the likelihood principle i.e. select that vector $\underline{\theta}$ which maximises the likelihood. The problem is, of course, to find assumptions which seem reasonable. This is another unsettled problem; to find distributional assumptions that are in accordance with the observations.

If one assumes that

$$EY_i = f(t_i, \underline{\theta})$$

and that

$$\varepsilon_i \text{ are i.i.d. as } N(0, \sigma^2) \quad (2.45)$$

then the log-likelihood is

$$\ell(\underline{\theta}; \sigma^2) = (k/2) \ln (2\pi\sigma^2) - \sum_{i=1}^k \{y_i - f(t_i, \underline{\theta})\}^2 / 2\sigma^2 \quad (2.46)$$

From equation (2.46) it is clear that the maximum likelihood estimator is identical to the least squares estimator, under the assumption of normality.

However, for bioavailability studies the model is a repeated measures model and the assumption of independent errors does not seem reasonable. Further, it seems evident from the column of standard deviations in

Table 2.1 that the assumption of constant variance is rather dubious. The data in Table 2.1 suggests that the standard deviation may be approximately proportional to y_i the observed concentration.

2.4 Minimisation techniques

For nonlinear models coefficients cannot be estimated directly. An iterative estimation procedure is required. Consider the problem of selecting a vector $\underline{\theta}$ that minimises some objective function $S(\underline{\theta})$. This problem occurs in least squares, weighted least squares and maximum likelihood estimation. Many iterative algorithms have been developed for solving this problem. The usual paradigm is

Step 1 : Select an initial estimate $\underline{\theta}_0$

Step 2 : iteratively obtain a new estimate $\underline{\theta}_{n+1}$ from the estimate $\underline{\theta}_n$ as follows

$$\underline{\theta}_{n+1} = \underline{\theta}_n - \underline{e}_n$$

rules are given for computing the correction term \underline{e}_n . These rules are usually dependent on S .

Step 3 : stop when \underline{e}_n lies within some specified neighbourhood of zero i.e. if $|\underline{e}_n| < \delta$ where $\delta > 0$ is specified.

The major difference between algorithms is the manner in which the correction term \underline{e}_n is computed. We will consider three of the most well known algorithms; steepest descent, Newton-Raphson and the Gauss-Newton procedure. These methods are discussed in detail in Royce Sadler (1975)

2.4.1 Steepest descent

Let the gradient of S , evaluated at $\underline{\theta}_n$, be denoted by \underline{g}_n

$$\underline{g}_n = \underline{g}(\underline{\theta}_n) = \nabla S(\underline{\theta}_n) = \begin{pmatrix} \frac{\partial S}{\partial \theta_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial S}{\partial \theta_m} \end{pmatrix} \quad \underline{\theta} = \underline{\theta}_n$$

The steepest descent iteration is defined by

$$\underline{\theta}_{n+1} = \underline{\theta}_n - h_n I \underline{g}_n$$

where: h_n is a scalar which determines the length of the step taken in the direction $-\underline{g}_n$; and

I is the $m \times m$ identity matrix, introduced here to unify the treatment.

The step length h_n is calculated by line search i.e. solve the one dimensional minimisation problem

$$\min_{h>0} S(\underline{\theta}_n - h I \underline{g}_n)$$

Provided that $\underline{g}_n \neq \underline{0}$; $S(\theta_{n+1}) < S(\theta_n)$

Although steepest descent will invariably steer clear of troublesome saddle points it is not in general a finite process. It may also converge too slowly. Royce Sadler (p.15) suggests that as a practical minimisation method its use is not recommended.

2.4.2 Newton-Raphson

Let H_n denote the Hessian of S evaluated at $\underline{\theta}_n$

$$H_n = H_n(\underline{\theta}_n) = \left(\frac{\partial^2 S}{\partial \theta_i \partial \theta_j} \right)_{ij} \Big|_{\underline{\theta} = \underline{\theta}_n}$$

The Newton-Raphson iterative formula is then

$$\underline{\theta}_{n+1} = \underline{\theta}_n - H_n^{-1} \underline{g}_n$$

Newton's method specifies direction and step length simultaneously.

The positive definiteness of H plays a role analogous to the sign of the second derivative in the case of one variable.

Geometrically this is equivalent to approximating the surface S with the osculating paraboloid resulting from a three term Taylor approximation to S . At each step one moves to the minimum of the osculating paraboloid.

A problem arises if H is ill-conditioned and difficult to invert. A possible solution would be to replace H with a positive definite matrix. This is done in the Gauss-Newton method that follows.

2.4.3 Gauss-Newton

Let $f(t, \underline{\theta})$ be the model for the data and let ∇f_{i_n} denote its gradient evaluated at $\underline{\theta}_n$ for the sampling time t_i

$$\nabla f_{in} = \begin{pmatrix} \frac{\partial f(t_i, \underline{\theta})}{\partial \theta_i} \\ \vdots \\ \frac{\partial f(t_i, \underline{\theta})}{\partial \theta_m} \end{pmatrix} \quad i = 1, \dots, k$$

$\underline{\theta} = \underline{\theta}_n$

The Gaussian approximation to the Hessian is

$$H_n \approx G_n = 2 \sum_{i=1}^k \nabla f_{in} \nabla f_{in}^T$$

Royce Sadler (1975, p.21) derives this result.

The iterative formula is

$$\underline{\theta}_{n+1} = \underline{\theta}_n - G_n^{-1} g_n$$

The Gaussian approximation to the Hessian is attractive for two reasons:

- 1) G is always positive definite
- 2) it requires only first derivatives of the regression function.

Since these are also required for computing g anyway, obtaining G involves very little extra work.

Although G is always positive definite it may be near singular. This occurs especially in the fitting of sums of exponential functions Juritz et al. (1983). In order to overcome the near singularity the Marquardt-Levenberg method may be used. The ill-conditioning is reduced by adding a term $\lambda_n I$ to G_n where $\lambda_n > 0$ is chosen so that

$$G_n + \lambda_n I$$

is positive definite.

2.5 Finding initial estimates

To start an iterative estimation procedure such as the three methods outlined in the previous section initial estimates θ_0 must be supplied. Although general rules cannot be given for finding initial estimates a number of procedures have been developed for polyexponential functions.

Steyn and van Wyk (1977) and Hibbert and Steyn (1982) discuss the following initial estimation methods

- (i) Graphical method. This has been described in section 2.3.1. Although this method is often used to give final parameter estimates, these estimates may also be used as initial values in an iterative algorithm,
- (ii) methods for the one exponential case,
- (iii) Fourier transform method for exponentially spaced data, Steyn (1980),
- (iv) regression-difference equation method for equally spaced data, Shah (1973),
- (v) method of partial sums for equally spaced data, Cornell (1962, 1965), Agha (1971), Della Corta *et al.* (1974),
- (vi) numerical integration method, Foss (1970), Fresen and Juritz (1985).

Since in practice bioavailability data are seldom equally or exponentially spaced methods (i) and (vi) are most useful and practical. Since the graphical method has been described in section 2.3.1 we will only discuss the numerical integration method proposed by Foss (1970) and the modifi-

cation proposed by Fresen and Juritz (1985).

We shall consider methods for obtaining initial parameter estimates for the polyexponential function

$$y = y(t) = \sum_{i=1}^m a_i e^{-b_i t} \quad (2.47)$$

and propose a simple modification to Foss's method (Foss 1970) which will accommodate fitted values that must pass through the origin. Furthermore, Foss's method becomes unstable under certain extreme conditions. Because a practitioner needs to know about the conditions under which the procedure will fail we examine these conditions.

For simplicity we consider only the sum of two exponentials. However the method can be extended to polyexponential functions consisting of three or more terms.

2.5.1 Foss's method for a sum of two exponentials

Consider the two exponential function

$$y = a_1 e^{-b_1 t} + a_2 e^{-b_2 t} \quad (2.48)$$

By differentiating this function twice with respect to t it can be shown to satisfy the differential equation

$$y'' + (b_1 + b_2) y' + b_1 b_2 y = 0 \quad (2.49)$$

The initial conditions are

$$y_0 = a_1 + a_2, \quad y'_0 = -a_1 b_1 - a_2 b_2 \quad (2.50)$$

Integrating (2.49) over t , we have

$$\int_0^t y''(\xi) d\xi = -(b_1+b_2) \int_0^t y'(\xi) d\xi - b_1b_2 \int_0^t y(\xi) d\xi \quad (2.51)$$

on integrating this becomes

$$y'(t) - y'_0 = (b_1+b_2)(y_0-y(t)) - b_1b_2 \int_0^t y(\xi) d\xi \quad (2.52)$$

Integrating again over t , we obtain

$$\int_0^t \{y'(\lambda)-y'_0\} d\lambda = (b_1+b_2) \int_0^t \{y_0-y(\lambda)\} d\lambda - b_1b_2 \int_0^t \int_0^\lambda y(\xi) d\xi d\lambda \quad (2.53)$$

$$\therefore y-y_0-y'_0t = (b_1+b_2)\{y_0t - \int_0^t y(\lambda) d\lambda\} - b_1b_2 \int_0^t \int_0^\lambda y(\xi) d\xi d\lambda \quad (2.54)$$

Now let $F(t) = \int_0^t y(\lambda) d\lambda$ and $G(t) = \int_0^t \int_0^\lambda y(\xi) d\xi d\lambda$

Substituting F and G and the initial conditions into (2.54) we obtain

$$\begin{aligned} y &= a_1 + a_2 + (a_1b_2+a_2b_1)t - (b_1+b_2)F(t) - b_1b_2G(t) \\ &= a + bt + cF(t) + dG(t) \end{aligned} \quad (2.55)$$

where $a = a_1 + a_2$; $b = a_1b_2 + a_2b_1$; $c = -(b_1+b_2)$; $d = -b_1b_2$ (2.56)

However, if $y_0 = 0$, then $a_1 + a_2 = 0$ and (2.55) becomes

$$y = bt + cF(t) + dG(t) \quad (2.57)$$

The functions $F(t)$ and $G(t)$ are unknown but we can approximate them using numerical quadrature: Let t_r be a typical sampling point then

$$F(t_r) = \int_0^{t_r} y(\lambda) d\lambda \quad (2.58)$$

The numerical quadrature procedure that Foss (1970) invokes in order to estimate F assumes that y may be interpolated between observations using a single exponential function, i.e.

$$y = \alpha_i e^{-\beta_i t} \quad t_i < t \leq t_{i+1} \quad i = 0, \dots, n-1 \quad (2.59)$$

Suppose that the observations are $(t_0, y_0), (t_1, y_1), \dots, (t_n, y_n)$ then the incremental area between t_i and t_{i+1} is approximately

$$\int_{t_i}^{t_{i+1}} y \, dt = (y_i + y_{i+1})(t_{i+1} - t_i) / \ln(y_i / y_{i+1}) \quad (2.60)$$

for $y_i \neq y_{i+1}$. If $y_i = y_{i+1}$ then

$$\int_{t_i}^{t_{i+1}} y \, dt = y_i(t_{i+1} - t_i) \quad (2.61)$$

The estimate of $F(t_r)$ is then given by

$$\hat{F}(t_r) = \sum_{i=0}^{r-1} \int_{t_i}^{t_{i+1}} y \, dt \quad (2.62)$$

This numerical quadrature procedure is a slight modification of what Yeh and Kwan (1978) describe as the log-trapezoidal method.

$$\begin{aligned} \text{For } G(t_r) &= \int_0^{t_r} \int_0^\lambda y(\xi) \, d\xi \, d\lambda \\ &= \int_0^{t_r} F(\lambda) \, d\lambda \end{aligned} \quad (2.63)$$

Foss (1970) uses the trapezoidal rule applied to the estimated values of F to give

$$\hat{G}(t_r) = \sum_{j=0}^{r-1} \{F(t_j) + F(t_{j+1})\}(t_{j+1} - t_j) / 2 \quad (2.64)$$

After making these substitutions equations (2.55) and (2.57) may be regarded as linear models in the parameters a, b, c and d . These parameters can be estimated by a multiple linear regression using least squares.

Having obtained estimates for a, b, c and d estimates of the initial

parameters are obtained by solving the system of equations (2.56).

The solutions are:

$$\begin{aligned} b_1 &= \frac{1}{2}\{-c + \sqrt{c^2+4d}\} & b_2 &= -d/b_1 \\ a_1 &= (ab_1-b)/(b_1-b_2) & a_2 &= a-a_1 \end{aligned} \quad (2.65)$$

2.5.2 Comments on Foss's method

Foss's method works for data that decreases monotonically. He gives an example to demonstrate this. However, his method does not work if applied to data that builds up from zero to a peak and then decreases back to zero. Such data is encountered in bioavailability trials of orally administered drugs (Wagner (1975)).

There are two reasons why his procedure does not work for such data; if the first observation y_0 is zero then formula (2.60) breaks down attempting to take the logarithm of zero.

Secondly if two consecutive observations y_j and y_{j+1} have approximately the same values, as may happen near the peak, then the denominator $\ln(y_j/y_{j+1})$ will be very close to zero. In this instance formula (2.60) may be inaccurate. Yeh and Kwan (1978) report that the log-trapezoidal method may produce large errors when used in an ascending curve, near a peak or in a steeply descending polyexponential curve. These large errors result in a poor estimate of F and hence also of G . The end effect is a breakdown of the linear relationship as given in equations (2.55) and (2.57).

2.5.3 Modification

In order to remedy both problems outlined in the previous section one need only alter the numerical quadrature method for estimating F . The simplest method to achieve this is to apply the trapezoidal method to give (Lanczos 1957, Yeh and Kwan 1978)

$$\hat{F}(t_r) = \sum_{j=0}^{r-1} (y_j + y_{j+1})(t_{j+1} - t_j)/2$$

for $r = 1, 2, \dots, n$.

Alternately one may use a combination of methods as outlined in Yeh and Kwan (1978). For example, use the trapezoidal method until just after the peak and thereafter the log-trapezoidal method. The author has found the trapezoidal method gives satisfactory results.

2.5.4 Extreme Conditions

As the ratio b_2/b_1 increases the functions t , $F(t)$ and $G(t)$ become almost linearly dependent. This causes a co-linearity problem when attempting to fit the linear model defined by (2.55) and (2.57).

Because of this the coefficients become unstable and the procedure breaks down.

The condition can be detected by plotting concentration or log-concentration against time; an early peak indicates that b_2/b_1 is large. The earlier the peak, the larger the ratio b_2/b_1 .

In order to see how this collinearity develops we may, without loss of generality, suppose that t and y have been scaled so that

$$y = e^{-t} - e^{-kt} \quad k > 1 \quad (2.66)$$

Here k represents the ratio b_2/b_1 . Figure 1 gives a plot of this function for $k = 1,5; 2; 5; 20$ and ∞ . It is evident from the plot that, as k increases (2.66) approximates e^{-t} very closely.

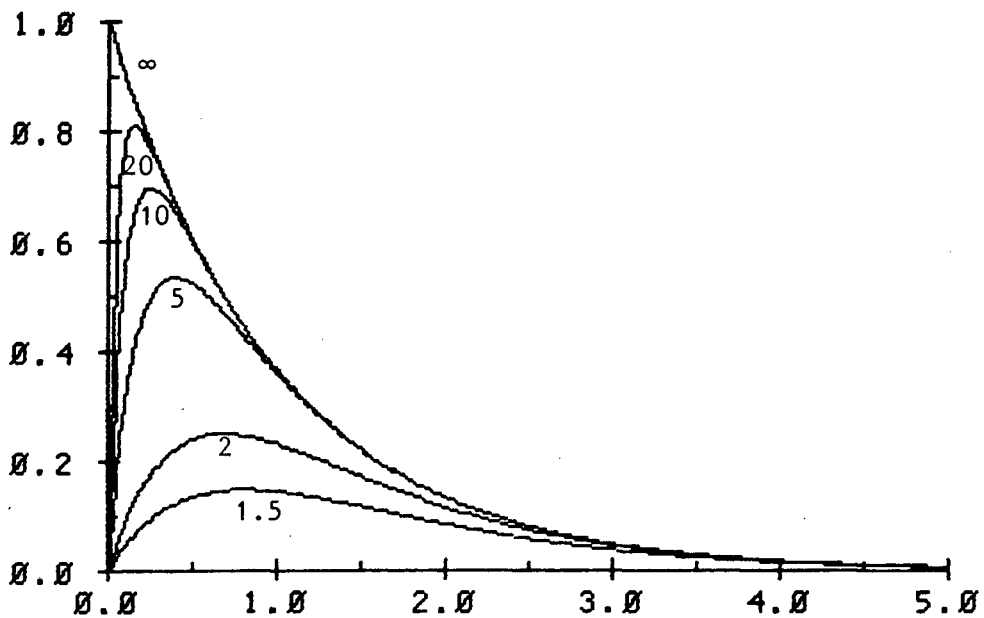


Figure 2.1 : Plot of the function $y = e^{-t} - e^{-kt}$ for various values of k . $k = 1,5; 2; 5; 10; 20$ and ∞ .

Consider first the case where $k = \infty$ so that

$$y = e^{-t} \quad (2.67)$$

This function satisfies the differential equation

$$y' = -y \quad (2.68)$$

with the initial conditions

$$y_0 = 1; \quad y'_0 = -1 \quad (2.69)$$

Integrating (2.68) over time gives

$$\int_0^t y'(\xi) d\xi = -\int_0^t y(\xi) d\xi$$

$$y(t) - 1 = -\int_0^t y(\xi) d\xi \quad (2.70)$$

Integrating (2.70) over time gives

$$\int_0^t y(\lambda) d\lambda - \int_0^t d\lambda = -\int_0^t \int_0^\lambda y(\xi) d\xi d\lambda \quad (2.71)$$

Corresponding to $k = \infty$ we define $F_\infty(t) = \int_0^t y(\lambda) d\lambda$ and

$$G_\infty(t) = \int_0^t \int_0^\lambda y(\xi) d\xi d\lambda$$

Substituting F_∞ and G_∞ into (2.71) we have

$$F_\infty(t) - t = -G_\infty(t)$$

$$F_\infty(t) + G_\infty(t) - t = 0 \quad t > 0 \quad (2.72)$$

Hence for $k = \infty$ the functions t , F_∞ and G_∞ are linearly dependent.

We now consider the case for which $k < \infty$. In this instance define

$$F_k(t) = \int_0^t (e^{-\lambda} - e^{-k\lambda}) d\lambda \quad (2.73)$$

$$G_k(t) = \int_0^t F_k(\lambda) d\lambda = \int_0^t \int_0^\lambda (e^{-\xi} - e^{-k\xi}) d\xi d\lambda \quad (2.74)$$

Now consider

$$F_\infty(t) - F_k(t) = \int_0^t e^{-k\lambda} d\lambda = \frac{1}{k} (1 - e^{-kt}) \quad (2.75)$$

and

$$\begin{aligned} G_\infty(t) - G_k(t) &= \int_0^t (F_\infty(\lambda) - F_k(\lambda)) d\lambda \\ &= \frac{1}{k} \int_0^t (1 - e^{-k\lambda}) d\lambda \\ &= \frac{1}{k} t - \frac{1}{k^2} (1 - e^{-kt}) \end{aligned} \quad (2.76)$$

From (2.75) and (2.76) it is clear that $F_k(t) \rightarrow F_\infty(t)$ and $G_k(t) \rightarrow G_\infty(t)$ as $k \rightarrow \infty$.

Therefore $F_k(t) + G_k(t) - t \rightarrow 0$ as $k \rightarrow \infty$ (2.77)

Hence for large k the functions t , F_k and G_k or simply t , F and G are almost linearly dependent.

Fresen and Juritz (1985) and Fresen (1985) give examples where the method works and where the method breaks down.

2.6 Asymptotic distribution of Least Squares/Maximal Likelihood estimators

Consider the statistical model given by (2.42) or alternately the model given by (2.45). The partial derivative of the response function with respect to the r th parameter θ_r for the u th sampling point t_u , evaluated at the unknown true parameter $\underline{\theta}^*$, is denoted by

$$f_{ru} = \left[\frac{\partial f(t_u, \underline{\theta})}{\partial \theta_r} \right]_{\underline{\theta} = \underline{\theta}^*} \quad (2.78)$$

There are k sampling points and m parameters; therefore the $k \times m$ matrix of these derivatives is

$$F = (f_{ru})$$

It is well known that the least squares estimator has variance-covariance matrix which is approximated by (Box and Lucas, 1959)

$$V = (F'F)^{-1} \sigma^2 \quad (2.79)$$

If the errors are assumed independent and normal then the Fisher information is $F'F/\sigma^2$. Hence (2.79) is also the asymptotic variance-covariance matrix for the maximum likelihood estimates.

2.7 Other methods for estimating the pharmacokinetic parameters

Four methods for estimating the model coefficients have been described in section 2.3; graphical method, least squares, weighted least squares and maximum likelihood. These four methods are perhaps the best known. However least squares is known to be sensitive to outliers. Rhodda et al. (1975) claim that a single outlier can have a devastating effect on the least squares estimates. For weighted least squares one is faced with the unsolved problem of choosing weights and maximum likelihood is dependent on distributional assumptions which seem hard to justify.

To avoid these weaknesses other approaches have been proposed. These include

- (i) Ordered Simulation Estimation Procedure (OSEP)
- (ii) Fourier analysis
- (iii) L_p estimation
- (iv) Statistical moments

We shall give a brief description of each method.

2.7.1 Ordered Simulation Estimation Procedure (OSEP)

Rhodda et al. (1975) propose a robust non-parametric procedure for the simplest pharmacokinetic model, the one compartment open model. They call their method the Ordered Simulation Estimation Procedure.

The one compartment open model leads to a sum of two exponentials

$$y = A \left(e^{-b_1 t} - e^{-b_2 t} \right) \quad b_2 > b_1$$

The domain is partitioned into three phases; absorption, peak and elimination. See Figure 2.2.

Selecting one observation (t_i, y_i) from each of these three regions generates a system of three equations in three unknowns, viz

$$y_i = A(e^{-b_1 t_i} - e^{-b_2 t_i}) \quad \text{for } i = 1, 2, 3$$

where $t_1 \in$ (absorption phase), $t_2 \in$ (peak phase) and $t_3 \in$ (elimination phase).

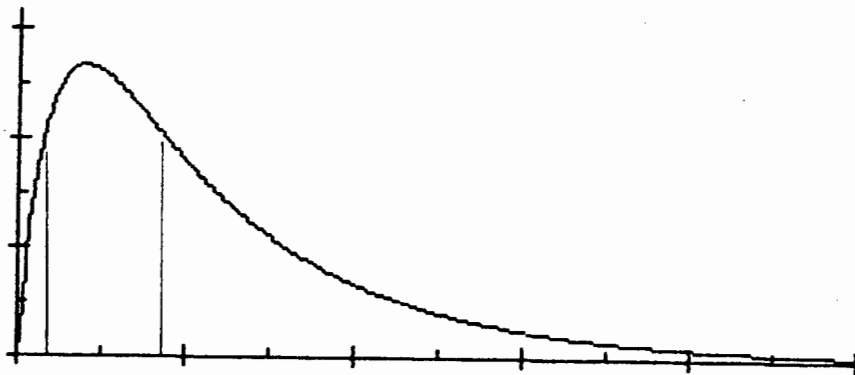


Figure 2.2 : Typical plot of $y = A(e^{-b_1 t} - e^{-b_2 t})$ showing partitioning of domain into absorption, peak and elimination phases

These three equations are solved for A , b_1 and b_2 using an iterative procedure such as Newton-Raphson or Steepest descent approach.

This procedure is repeated for each of the $N = n_1 n_2 n_3$ possible combinations and produces N estimates of A , b_1 and b_2 . These N estimates are ordered and the median of each estimated parameter is used as an estimate of the unknown parameter.

Rhodda et al. (1975) compare this estimation procedure with least squares using a Monte Carlo study. They recommend this estimation procedure

whenever outliers are known to exist, otherwise they recommend least squares.

As motivation for this procedure they cite an analogous procedure for the linear case as described by P.K. Sen (1971).

The drawback of this procedure is the arbitrary manner of partitioning the domain into the three regions absorption, peak and elimination. The effect of different partitions has not been studied.

2.7.2 Fourier Analysis

Gardner et al. (1959) proposed a method for fitting polyexponential functions based on Fourier analysis. Smith et al. (1976) proposed a discrete analogue to Gardner's method via spectral analysis. According to Steyn (1980) these methods did not gain wide acceptance because the numerical evaluation of Fourier integrals was then extremely difficult and tedious. He showed how the Fast Fourier Transform together with a filtering and smoothing technique can be used with success to estimate the coefficients or parameters in the model. However, the method is derived for exponentially spaced observations

$$\text{i.e. } t_x = \exp\{2a(x-m)/N\} \quad x = 0, 1, \dots, N-1.$$

If the observations are not precisely exponentially spaced Steyn suggests interpolating the y-values at the precise t_x points and then applying his method.

He compares this method with the graphical method and concludes that this method yields parameter values that describe the model more accurately.

It is clear that this method may be viewed as a method of obtaining initial estimates.

The details of the method are given in Steyn (1980).

2.7.3 L_p-norm estimation

The L_p-norm estimator is defined as that vector $\underline{\theta}$ that minimises the objective function

$$SL_p(\theta) = \sum_{i=1}^k |y_i - f(t_i, \underline{\theta})|^p \quad \text{where } 1 \leq p < \infty$$

For $p = 2$ this is equivalent to the least squares estimator.

Gonin and Money (1985a,b) discuss the choice of p as well as its asymptotic distribution.

2.7.4 Statistical Moments

Two pharmacokinetic parameters not yet mentioned are the Mean Residence Time (MRT) and the Mean Absorption Time (MAT). The MRT is defined in terms of the AUC and the first moment of the AUC (AUMC). Let

$$AUC(0, \infty) = \int_0^{\infty} y \, dt$$

and

$$AUMC(0, \infty) = \int_0^{\infty} ty \, dt$$

Then the MRT is defined as

$$MRT = AUC(0, \infty) / AUMC(0, \infty)$$

Both $AUC(0, \infty)$ and $AUMC(0, \infty)$ can be evaluated using numerical integration or by fitting a polyexponential model.

3.2 The two period cross-over trial

The two period cross-over trial compares two formulations of a drug, a new (N say) with a standard (S), each received by every subject over two periods. The n subjects are randomly divided into two groups of sizes n_1 and n_2 . It is preferable to have equal group sizes i.e. $n_1 = n_2$. One group receives the treatments in order N-S the other group in the reverse order. Often, a washout period between treatments is allowed to eliminate carry over effects. A typical form of the design is given in Figure 3.1.

| | PERIOD 1 | | PERIOD 2 |
|----------|----------|---------|----------|
| GROUP I | N | Washout | S |
| GROUP II | S | Period | N |

Figure 3.1 Two period cross-over trial with washout period

For bioavailability trials it is usual to have between 12 and 20 subjects. However, as was mentioned in section 1.5, the FDA draft suggests that *"The number of subjects employed in the test should be adequate to demonstrate the lack of statistically significant practical difference among the bioavailability parameters selected with an alpha of 0.05 and a beta of 0.20."*

There are two models to consider for this design; one with residual or carry over effect and the other without.

3.2.1 Model I : With residual/carry over effects

Consider the j th patient in the i th sequence or group in the k th period and denote the response by y_{ijk} . The model is

$$\begin{aligned}
 y_{ijk} = \mu + \xi_{ij} + \pi_k + \phi_\ell + \lambda_\ell + \varepsilon_{ijk} \quad & i = 1,2 \\
 & j = 1, \dots, n_i \\
 & k = 1,2 \\
 & \ell = 1,2
 \end{aligned} \tag{3.1}$$

where μ = general mean

ξ_{ij} = effect of j th patient within i th sequence; a random variable with mean zero and variance σ_S^2

π_k = effect of the k th period

ϕ_ℓ = direct effect of the ℓ th drug

λ_ℓ = residual effect of the ℓ th drug

ε_{ijk} = random fluctuation distributed with mean 0 and variance σ_e^2 , and is independent of ξ_{ij} .

With these assumptions the variance of an observation is

$$\text{Var}(y_{ijk}) = \sigma_e^2 + \sigma_S^2$$

and any two observations on an individual have covariance

$$\text{Cov}(y_{ij1}, y_{ij2}) = \sigma_S^2$$

Observations made on different subjects are independent, i.e.

$$\text{Cov}(y_{irk}, y_{itk}) = 0 \quad r \neq t$$

Grizzle (1965, corrected 1969) discusses the estimation and testing procedures for this model in detail.

The contrasts of interest and their variances are given in Table 3.1.

Table 3.1 Contrasts of interest (Grizzle 1965)

| Contrast | Estimate | Variance |
|-------------------------|---|------------------------|
| $\pi_1 - \pi_2$ | Not estimable | - |
| $\phi_1 - \phi_2$ | $\bar{y}_{1.1} - \bar{y}_{2.1}$ | $4 \sigma^2/n$ |
| $\lambda_1 - \lambda_2$ | $\bar{y}_{1.1} - \bar{y}_{1.2} - \bar{y}_{2.1} + \bar{y}_{2.2}$ | $2 \sigma^2(1+\rho)/n$ |

In the table: $\sigma^2 = \sigma_e^2 + \sigma_s^2$
 $\rho = \sigma_s^2/(\sigma_e^2 + \sigma_s^2)$

In most clinical trials, the hypotheses of interest are, either individually or jointly, $\lambda_1 = \lambda_2$ and $\phi_1 = \phi_2$. The period contrast $\pi_1 - \pi_2$ is not estimable under the model assumed. If there is interest in this contrast, a different design should therefore be used.

The analysis of variance for individual tests of hypothesis is given in Table 3.2.

The MAT is the difference between the MRT for a non instantaneous input ($MRT_{n.i.v.}$) and the MRT as they occur in the in vivo release and absorption process ($MRT_{i.v.}$)

$$MAT = MRT_{n.i.v.} - MRT_{i.v.}$$

Riegelman and Collier (1980) discuss the estimation and relevance of these concepts.

CHAPTER 3

ASSESSING BIOEQUIVALENCE

3.1 Introduction

Methods of assessing bioequivalence may be broadly classified in the following manner:

- (i) methodology employed; hypothesis testing, the use of confidence intervals or Bayesian procedures,
- (ii) parametric or non-parametric, and
- (iii) difference or ratio of the parameters involved.

The simple cross-over design is used so frequently in bioequivalence trials that it might be thought of as the standard design (Steinijans and Diletti (1983), Huitson et al (1982)). We shall therefore restrict attention to the simple two period cross-over trial and assume that the data is a result of such a trial.

Most of the procedures discussed in this chapter are univariate. However, in practice bioequivalence is assessed on a number of parameters. This is achieved by applying a univariate procedure to each parameter individually. The effect is a reduction of the significance levels and coverage probabilities.

The purpose of this chapter is to give a brief review of the two way cross-over trial and the methods for assessing bioequivalence.

Table 3.2 Analysis of Variance for Cross-over Design with Residual Effects (Model I)

| Source of Variation | df | Sum of squares | Expected mean squares |
|---------------------|-----|---|---|
| Residual Effect | 1 | $\frac{1}{2n_1n_2n}(n_2Y_{1..} - n_1Y_{2..})^2$ | $\sigma_e^2 + 2\sigma_s^2 + \frac{n_1n_2}{2n}(\lambda_1 - \lambda_2)^2$ |
| Subject (seq) | n-2 | $\frac{1}{2}\left(\sum_{j=1}^{n_1} Y_{1j}^2 + \sum_{j=1}^{n_2} Y_{2j}^2 - \frac{Y_{1..}^2}{n_1} + \frac{Y_{2..}^2}{n_2}\right)$ | $\sigma_e^2 + 2\sigma_s^2$ |
| Treatment | 1 | $\frac{1}{n_1n_2n}(n_2Y_{1.1} - n_1Y_{2.1})^2$ | $\sigma_e^2 + \sigma_s^2 + \frac{n_1n_2(\phi_1 - \phi_2)^2}{n}$ |
| Error | n-2 | $\sum_{i=1}^2\left(\sum_{j=1}^{n_i} Y_{ij1}^2 - \frac{Y_{i.1}^2}{n_i}\right)$ | $\sigma_e^2 + \sigma_s^2$ |

Remarks: (i) $Y_{1..}$ is the total of all observations on the first sequence, etc.

(ii) the sums of squares for error is obtained from the first period only.

3.2.2 Model II : Without residual/carry over effects

If no residual/carry over effects are assumed then a model for the data would be

$$\begin{aligned}
 y_{ijk} &= \mu + \psi_i + \xi_{ij} + \pi_k + \phi_\ell + \varepsilon_{ijk} & i &= 1,2 \\
 & & j &= 1, \dots, n_i \\
 & & k &= 1,2 \\
 & & \ell &= 1,2
 \end{aligned} \tag{3.2a}$$

where ψ_i = effect of the i th sequence group and the other symbols have the same meaning as for model I.

The analysis of variance for this model is given in Table 3.3a.

Table 3.3a Analysis of Variance for Cross-over Design without Residual Effects (Model II)

| Source of Variation | df | Sum of squares | Expected mean square |
|---------------------|-----|---|---|
| Sequence | 1 | $\frac{1}{2n_1n_2n} (n_2Y_{1..} - n_1Y_{2..})^2$ | $\sigma_e^2 + 2\sigma_s^2 + \frac{2n_1n_2(\psi_1 - \psi_2)^2}{n}$ |
| Subject (seq) | n-2 | $\frac{1}{2} \left(\sum_{j=1}^{n_1} Y_{1j}^2 + \sum_{j=1}^{n_2} Y_{2j}^2 - \frac{Y_{1..}^2}{n_1} - \frac{Y_{2..}^2}{n_2} \right)$ | $\sigma_e^2 + 2\sigma_s^2$ |
| Treatment | 1 | $SST = \frac{1}{2n_1n_2n} (n_2G_1 - n_1G_2)^2$ | $\sigma_e^2 + \frac{2n_1n_2(\phi_1 - \phi_2)^2}{n}$ |
| Period | 1 | $SSP = \frac{1}{2n_1n_2n} (n_2G_1 + n_1G_2)^2$ | $\sigma_e^2 + \frac{2n_1n_2(\pi_1 - \pi_2)^2}{n}$ |
| Error | n-2 | $SSE = \frac{1}{2} \sum_{i=1}^2 \left\{ \left[\sum_{j=1}^{n_i} (Y_{ij1} - Y_{ij2})^2 \right] - \frac{(Y_{i.1} - Y_{i.2})^2}{n_i} \right\}$ | σ_e^2 |

Where $G_1 = Y_{1.1} - Y_{1.2}$
 = difference between the totals for drug N and drug S in first sequence

$G_2 = Y_{2.1} - Y_{2.2}$
 = difference between the totals for drug N and drug S in second sequence

$n = n_1 + n_2$

Remarks: A number of authors use a slightly different model to (3.2a). The term allowing for the effect of the i th sequence group ψ_i is omitted. This simplifies the model which now becomes

$$y_{ijk} = \mu + \xi_{ij} + \pi_k + \phi_l + \varepsilon_{ijk} \tag{3.2b}$$

The ANOVA for this model is obtained from Table 3.3 by combining the sums of squares for sequences and subjects-within-sequence. The resulting sum of squares is called the variation due to subjects. For the sake of simplicity they let $T = \phi_1 = -\phi_2$ (Treatment effect) and $P = \pi_1 = -\pi_2$ (Period effect). This gives $\phi_1 - \phi_2 = 2T$ and $\pi_1 - \pi_2 = 2P$. It is further assumed that $n_1 = n_2$, and let $\sigma_A^2 = \sigma_e^2 + 2\sigma_s^2$.

For completeness, the ANOVA corresponding to these simplifications is given in Table 3.3b.

Table 3.3b Analysis of variance for the cross-over design without Residual Effects. (Model II - with simplifications)

| Source of Variation | df | Sum of Squares | Expected Mean Square |
|---------------------|-------|---|---|
| Formulations | 1 | $SST = 2n\{(Y_{1.1} + Y_{2.2})/2 - Y_{...}\}^2$ | $\sigma_e^2 + 2nT^2$ |
| Periods | 1 | $SSP = 2n(Y_{..1} - Y_{...})^2$ | $\sigma_e^2 + 2nP^2$ |
| Subjects | $n-1$ | $SSS = 2\sum_{i=1}^2 \sum_{j=1}^{n/2} (y_{ij.} - Y_{...})^2$ | $\sigma_A^2 = \sigma_e^2 + 2\sigma_s^2$ |
| Error | $n-2$ | $SSE = \text{By subtraction}$ $= \text{Same as in Table 3.3a}$ | σ_e^2 |

See Steinijans and Diletti (1983), Selwyn et al (1981), Rocke (1984), Flueher et al (1983).

Mendallaz and Mau (1981) use a simplified model but assume that the subject effect is fixed for normally distributed data. For log normally distributed data, they assume that the subject effect is random.

The essence of bioequivalence is the comparison of a new formulation, N, with a standard formulation, S. To aid recognition the following notation will be used:

- 1) μ_N and μ_S represent the population means of the bioavailability parameter for the new and standard formulations respectively.
- 2) $\bar{Y}_N = \bar{Y}_{1.1}$ and $\bar{Y}_S = \bar{Y}_{2.1}$ will represent the sample means of the new and standard formulations respectively.

3.2.3 Problems of the cross-over design

The FDA have discouraged the use of the two period cross-over design where unequivocal evidence of the treatment differences is required. The essential problem with the cross-over design is the estimation of treatment effect in the presence of period interaction, carry over effects or group effects. These three effects are totally confounded and can only be detected relatively inefficiently because they must be tested against the between subject variation. If a treatment by period interaction exists but is not detected the estimated treatment effect will be biased and perhaps meaningless.

Against this argument, many statisticians believe that a carry over effect is inappropriate for bioequivalence trials. A blood sample drawn immediately prior to the drug administration in the second period is regarded as sufficient evidence to demonstrate that the drug has been

eliminated from the system. This however, does not allow for the presence of metabolites of the drug, or for other subtle changes induced by the drug administration in the first period. Clearly, this is still a controversial issue. See Wallenstein and Fisher (1977), Westlake (1979), Huitson et al (1982), Selwyn et al (1981).

3.2.4 Non-parametric analysis of the cross-over design

Koch (1972), Taulbee (1982) and Steinijans and Diletti (1983, 1985) discuss the use of non-parametric methods in the analysis of the two period cross-over design. Tests are provided for the hypothesis $\lambda_1 = \lambda_2$ and $\phi_1 = \phi_2$. Equality of direct effects can be examined by applying the sign test or the Wilcoxon signed rank test. Koch (1972) outlines the steps needed in the analysis. Steinijans and Diletti (1983, 1985) show how to compute non-parametric confidence intervals for the difference or the ratio of treatment means.

3.2.5 Formulating other hypotheses

The classical null hypothesis which the ANOVA address is

$$H_0 : \mu_N = \mu_S$$

Anderson and Hauck (1983) maintain that this hypothesis is inappropriate for two reasons. Firstly, it may lead to the situation where a difference is statistically significant but not clinically meaningful.

The second concerns the logic of an hypothesis test. They argue that in order to demonstrate equivalence, the equivalence hypothesis should be the alternate and not the null hypothesis.

They set up a null and alternate hypothesis as follows

$$H_0 : \mu_N - \mu_S \leq A \quad \text{or} \quad \mu_N - \mu_S \geq B$$

and

$$H_A : A < \mu_N - \mu_S < B$$

and most often $A = -B$.

The test statistic is

$$T = \frac{\bar{Y}_N - \bar{Y}_S - \frac{1}{2}(A+B)}{S(1/n_1 + 1/n_2)^{\frac{1}{2}}}$$

where the \bar{Y} 's are the respective sample means, n_1 and n_2 the group sample sizes (usually $n_1 = n_2$) and S is calculated from the appropriate ANOVA. T then has a non-central t distribution with non-centrality parameter

$$\lambda = \frac{\mu_N - \mu_S - \frac{1}{2}(A+B)}{\sigma(1/n_1 + 1/n_2)^{\frac{1}{2}}}$$

Rocke (1984) compares four procedures that focus on an hypothesis that the true difference $\delta = \mu_N - \mu_S$ is less than some specified tolerance Δ .

$$\text{i.e. } H_0 : -\Delta < \delta < \Delta \quad \text{or} \quad |\delta| < \Delta$$

He defines two formulations to be bioequivalent with tolerance Δ , abbreviated B_Δ , if $|\delta| < \Delta$.

The four procedures, denoted by P_1 , P_2 , P_3 and P_4 , are:

P_1 : construct an ordinary $(1-\alpha)100\%$ confidence interval for δ and conclude B_Δ whenever that confidence interval is contained in $[-\Delta, \Delta]$. Metzler (1974); Kirkwood (1981).

P_2 : construct a confidence interval symmetric around 0 and conclude B_Δ if that interval is contained in $[-\Delta, \Delta]$. Westlake (1972, 1974, 1975, 1976, 1979).

P_3 : use P_1 with a $(1-2\alpha)100\%$ confidence interval. Westlake (1981).

P_4 : Let D be an estimate of δ and $T(x)$ be a one-sided tail area beyond x for the t distribution with ν df. Conclude B_Δ whenever

$$T\{(\Delta - |D|)/s\} - T\{(\Delta + |D|)/s\} \leq \alpha \quad \text{for } |D| \leq \Delta$$

or

$$1 - T\{(\Delta - |D|)/s\} - T\{(\Delta + |D|)/s\} \leq \alpha \quad \text{for } |D| > \Delta$$

He proves the following theorem:

- (i) For any specific instance of D , s , ν and Δ , let p_1 , p_2 , p_3 and p_4 be nominal p-values associated with the test procedures P_1 , P_2 , P_3 and P_4 . Then $p_4 < p_3 < p_2 < p_1$.
- (ii) If each procedure is run with nominal size α , and $\alpha_1(\delta)$, $\alpha_2(\delta)$, $\alpha_3(\delta)$ and $\alpha_4(\delta)$ are the actual probabilities of concluding B_Δ if $E(D) = \delta$, then $\alpha_1(\delta) < \alpha_2(\delta) < \alpha_3(\delta) < \alpha_4(\delta) \leq \alpha$ whenever $|\delta| > \Delta$, with equality if $\delta = \Delta$. Thus P_4 has actual size α whereas the other procedures are conservative and have actual sizes strictly less than the nominal size α .
- (iii) If I_1 , I_2 , I_3 and I_4 are nonequivalence intervals for the four procedures, then $I_4 \subset I_3 \subset I_2 \subset I_1$. Thus P_4 gives the most precise nonequivalence interval.

Mandallaz and Mau (1981) view the bioequivalence problem in terms of the ratio of means $\theta = \mu_N/\mu_S$. Bioequivalence is defined as the condition

$r_1 < \theta < r_2$ ($0 < r_1 < 1 < r_2$). The hypotheses to be considered are:

$H : \theta \in [r_1, r_2]$, the null hypothesis of bioequivalence

$K : \theta \notin [r_1, r_2]$, the alternate hypothesis of no bioequivalence

They compare Westlake's approximate symmetric confidence interval with the well known exact Fieller confidence interval and derive an exact version of Westlake's procedure. They also give a Bayesian interpretation viz. the posterior probability $P[\theta \in (r_1, r_2)]$, to the exact version they have derived. This is achieved using a vague improper prior distribution (cf. Box and Tiao, 1973, §1.3). The data are assumed normal or lognormal.

3.3 Confidence Intervals

Westlake (1972, 1975, 1976, 1979) has argued strongly that establishing a confidence interval is more appropriate than hypothesis testing in bioavailability studies. Others who concur with this opinion are Metzler (1974), Shirley (1976), Steinijans and Diletti (1983) and Anderson and Hauck (1983). Consequently, a number of methods have been derived for establishing a confidence interval. These are

- (i) confidence interval based on the ANOVA
- (ii) Westlake's modification of the ANOVA based confidence interval
- (iii) confidence interval based on the paired t-test
- (iv) nonparametric confidence intervals based on the Wilcoxon signed rank test
- (v) nonparametric confidence intervals based on Pitman's permutation tests.

A brief review of these methods will be given here. The development given here follows that given in Steinijans and Diletti (1983).

For the purpose of computing confidence intervals it will be assumed that carry over effects are of no concern. (If they are present they may be confounded with direct effects, then only the data from the first period can be used.) Using this assumption model II is appropriate, the ANOVA of which is given in Table 3.3a or Table 3.3b depending on the assumed model.

3.3.1 Confidence interval based on the ANOVA

A two-sided $(1-\alpha)100\%$ confidence interval for the expected mean difference

$\delta = \mu_N - \mu_S = \pi_1 - \pi_2$ is calculated as follows:

$$\bar{d} \pm t(n-2; 1 - \frac{\alpha}{2}) \sqrt{2 \text{MSE}/n} \quad (3.3)$$

where $\bar{d} = (\bar{Y}_{1,2} + \bar{Y}_{2,1})/2 - (\bar{Y}_{1,1} - \bar{Y}_{2,2})/2$

= estimated mean difference between the new and standard formulations

MSE = mean square for error in the ANOVA

$t(v, 1 - \frac{\alpha}{2}) = 1 - \frac{\alpha}{2}$ fractile of the t distribution with v df.

An approximate $(1-\alpha)100\%$ confidence interval for the bioavailability ratio

$\theta = \mu_N/\mu_S$ is given by

$$(\bar{Y}_N \pm t(n-2, 1 - \frac{\alpha}{2}) \sqrt{2 \text{MSE}/n}) / \bar{Y}_S \quad (3.4)$$

If $\log(\text{AUC})$ rather than AUC is assumed to follow the normal distribution (Metzler 1974, Westlake 1973, 1976, Steinijans et al 1982) then the confidence limits defined by (3.3) apply to logarithms of the

expected bioavailability ratio. Taking antilogs thus provide $(1-\alpha)100\%$ confidence limits for the bioavailability ratio itself. A point estimate of this ratio is given by the geometric mean of individual ratios.

3.3.2 Westlake's modification of the ANOVA based confidence interval

The confidence interval defined by (3.3) is symmetrical about the estimated mean difference \bar{d} . In order to shift the emphasis from estimation to decision making Westlake (1972, 1976 and 1979) proposed a confidence interval symmetric about 0. This is achieved by selecting two constants k_1 and k_2 that satisfy

$$k_1 + k_2 = 2(\bar{Y}_N - \bar{Y}_S)/\sqrt{2 \text{MSE}/n} \quad (3.5)$$

and

$$\int_{k_1}^{k_2} f_{t(n-2)}(S) dS = 1 - \alpha \quad (3.6)$$

Equation (3.6) ensures that the interval (k_1, k_2) includes $(1-\alpha)100\%$ of the mass of the t distribution.

A $(1-\alpha)100\%$ confidence interval, symmetric around 0, is given by

$$(\bar{d} + k_1\sqrt{2 \text{MSE}/n}, \bar{d} + k_2\sqrt{2 \text{MSE}/n}) \quad (3.7)$$

As before, approximate confidence limits can be obtained for the ratio μ_N/μ_S by

$$(\bar{Y}_N + k_i\sqrt{2 \text{MSE}/n})/\bar{Y}_S \quad i = 1,2 \quad (3.8)$$

These symmetrical confidence intervals have been criticised by both Mantel (1977) and Kirkwood (1981). Mantel gives a few exaggerated (his own adjective) examples that demonstrate the absurdities that might arise

by taking confidence intervals symmetrical about the null value. He considers, for example, what happens when the point estimate of a difference between two means is far from the null value of zero, e.g. the point estimate is 100 ± 3 . The usual confidence interval is approximately (94; 106) while that symmetric about zero is approximately (-105; 105). Alternately, the outcome 5 ± 3 would lead to symmetric limits of (-9,935; 9,935). Yet those same symmetric limits would have arisen had the outcome been -5 ± 3 . Mantel also suggests that for ratios, symmetric intervals are absurd on the grounds that 0,5 is as far from 1 as is 2, 0,1 is as far from 1 as is 10. He also suggests that a problem could arise if the confidence interval for a ratio could not be constructed. Since the lower bound is 0 the largest possible interval is (0,2). Kirkwood makes similar remarks.

Westlake (1977, 1981) defends his symmetrical intervals on the grounds that they were intended solely for use in bioequivalence assessment and were essentially proposed as a decision making device, i.e. two formulations are declared bioequivalent if the confidence interval lies wholly inside the bioequivalence specification interval which is usually symmetric. Rocke (1984) has established that, as a decision making device, these symmetrical intervals are superior to the standard intervals in the sense that has already been discussed in section 3.2.5. Hence both Mantel's and Kirkwood's objections are largely irrelevant in bioequivalence assessment.

3.3.3 Confidence intervals based on the paired t-test

If no period effect is assumed the model (3.2b) for the ANOVA reduces to

$$y_{ijk} = \mu + \xi_{ij} + \phi_l + \varepsilon_{ijk} \quad (3.9)$$

Intra-individual differences $d_{1j} = y_{1j2} - y_{1j1}$ and $d_{2j} = y_{2j1} - y_{2j2}$ in sequences 1 and 2 respectively, have expectation and variance

$$E(d_{ij}) = \mu_N - \mu_S$$

$$V(d_{ij}) = 2\sigma_e^2$$

Estimating the variance $V(d_{ij})$ by

$$S_d^2 = \sum_{i=1}^2 \sum_{j=1}^{n/2} (d_{ij} - \bar{d})^2 / (n-1)$$

$(1-\alpha)100\%$ confidence limits are given by

$$\bar{d} \pm t(n-1; 1-\alpha/2) S_d / \sqrt{n}$$

As before, approximate $(1-\alpha)100\%$ confidence limits for the ratio μ_N/μ_S are given by

$$\{\bar{Y}_N \pm t(n-1; 1-\alpha/2) S_d / \sqrt{n}\} / \bar{Y}_S \quad (3.10)$$

The procedure for log-transformed data is straightforward. Notice that the paired t-test is closely related to the ANOVA. The estimate of $V(d_{ij})$ can be obtained as follows

$$\hat{\text{Var}}(d_{ij}) = (SSE + SSP) / (n-1)$$

where SSE and SSP are obtained from Tables 3.3a or 3.3b.

3.3.4 Nonparametric confidence intervals based on Wilcoxon's signed rank tests

Wilcoxon's signed rank test is the nonparametric analogue of the paired t-test. The test is based on the assumption that no period effect is

present. Intra-individual differences are denoted by d_j , $j = 1, \dots, n$ irrespective of the sequence of administration.

The model is

$$d_i = \delta + \varepsilon_i \quad i = 1, \dots, n \quad (3.11)$$

where the ε 's are random error terms and δ is the expected difference between formulations. The distributional assumptions on the error term are that the ε 's are independent and symmetrically distributed about 0 from a continuous distribution. See Hollander and Wolfe (1973, pp. 26-33) and Steinijans and Diletti (1983, 1985).

To form a nonparametric confidence interval the $n(n+1)/2$ arithmetic Walsh averages a_{ij} are computed

$$a_{ij} = (d_i + d_j)/2 \quad i \leq j; \quad j = 1, \dots, n \quad (3.12)$$

Let $a(1), \dots, a(n(n+1)/2)$ denote their ordered values. The $(1-\alpha)100\%$ confidence interval (L, U) is given by

$$L = a(C_\alpha) ; \quad U = a(n(n+1)/2 + 1 - C_\alpha)$$

where $C_\alpha = n(n+1)/2 + 1 - t(\alpha/2; n)$

$t(\alpha/2; n)$ is the critical point of the Wilcoxon sum of positive ranks.

A table of values of C_α and $n(n+1)/2 - C_\alpha$ is given in Table 2 of Hollander and Wolfe (1973 pp.269-271).

This method may be modified for ratios instead of differences by taking logarithms of the AUC. The details are given in Steinijans and Diletti (1983).

3.3.5 Nonparametric confidence intervals based on Pitman's permutation test

The model for the permutation test is again given in (3.11) but with the distinction that the ε 's need not necessarily come from a continuous distribution. The distribution is still assumed symmetric though.

Under the hypothesis $\delta = 0$, the 2^n permutations of signs produce a discrete uniform distribution with point mass 2^{-n} . The $(1-\alpha)100\%$ confidence limits are computed as follows:

(i) Let $\{i_1, \dots, i_m\}$ denote a nonempty subset of the index set $\{1, \dots, n\}$. There are $2^n - 1$ such subsets; n have only 1 element i.e. $m = 1$, $\binom{n}{2}$ have two elements i.e. $m = 2$, etc.

(ii) Let A be the set of all $2^n - 1$ arithmetic averages of observed differences d_i ($i = 1, \dots, n$):

$$A = \left\{ M^{-1} \sum_{k=1}^M d_{i_k} : \{i_1, \dots, i_m\} \subset \{1, \dots, n\} \right\}$$

(iii) Let $a(1), \dots, a(2^n - 1)$ denote the ordered elements of A .

(iv) The $(1-\alpha)100\%$ confidence limits (L, U) are given by

$$L = a(k_\alpha) \quad U = a(2^n - k_\alpha)$$

where k_α is chosen such that $k_\alpha / 2^n \leq \alpha / 2$

This procedure can be modified to ratios in a similar manner to the Wilcoxon procedure.

3.4 Bayesian approach to bioequivalence

The Bayesian approach to bioequivalence assessment has been derived for both the difference $\theta = \mu_N - \mu_S$ (Selwyn et al, 1981) and the ratio

$\theta = \mu_N/\mu_S$ (Mandallaz and Mau 1981, Fluehler et al 1983) of the formulation means. In both cases the two period cross-over trial is assumed.

The Bayesian criterion adopted for bioequivalence is a high posterior probability that the parameter θ , either the difference or ratio of formulation means, lies inside a specified interval.

Selwyn and Hall (1984) extend the Bayesian methodology to other designs such as the Latin square and a design where the formulations are administered simultaneously. The simultaneous design is achieved by tagging a radio active isotope to the new formulation and administering both formulations simultaneously. The concentration of tagged and untagged drug are measured simultaneously by the combined use of gas chromatography and mass spectrometry of the blood samples.

Section 3.4.1 considers the Bayesian approach for the difference between formulation means without carry over effects, section 3.4.2 includes a carry over effect and section 3.4.3 considers the ratio of formulation means without carry over effect.

3.4.1 Difference between formulation means without carry over effect

Selwyn et al (1981) assume a two period cross-over design and, initially, that no carry over effect is present. The model for the data is therefore given by equation (3.2b). For simplicity they let

$$T = \phi_1 = -\phi_2 \quad (\text{Treatment effect}), \quad P = \pi_1 = -\pi_2 \quad (\text{Period effect}) \quad \text{and} \\ \sigma_A^2 = \sigma_e^2 + 2\sigma_S^2.$$

The criterion adopted for bioequivalence is a high posterior probability that the difference in formulation means is less than some fraction, say k , of the mean of the standard. With the above notation $\mu_S = \mu + T$ and $\mu_N = \mu - T$. The criterion is

$$|\mu_N - \mu_S| < k \mu_S \quad (3.13)$$

After substituting for μ_S and μ_N into (3.13) the criterion for bioequivalence translates into

$$|\mu_N - \mu_S| < k \mu_S$$

$$\therefore |\mu - T - \mu - T| < k(\mu+T)$$

$$\therefore |2T| < k(\mu+T)$$

$$\text{or } k_1\mu < T < k_2\mu \quad (3.14)$$

where $k_1 = -k/(2+k)$ and $k_2 = k/(2-k)$

The posterior density of μ and T is computed and then this density is integrated over the wedge shaped region defined by (3.14) to obtain the posterior probability. If the posterior probability is sufficiently high the formulations are considered bioequivalent. Non-informative priors are used and the components of the likelihood are assumed to be normal or chi-squared.

They test the sensitivity of the posterior probability on the prior by considering four different non-informative priors

$$\begin{aligned}
p_1(\mu, P, T, \sigma_e^2, \sigma_A^2) &\propto \sigma_e^{-2} \sigma_A^{-2} \\
p_2(\mu, P, T, \sigma_e^2, \sigma_A^2) &\propto \sigma_e^{-2} \sigma_A^{-2} \quad \text{for } \sigma_A^2 > \sigma_e^2 \\
p_3(\mu, P, T, \sigma_e^2, \sigma_A^2) &\propto \sigma_e^{-1} \sigma_A^{-1} \\
p_4(\mu, P, T, \sigma_e^2, \sigma_A^2) &\propto \sigma_e^{-1} \sigma_A^{-1} \quad \text{for } \sigma_A^2 > \sigma_e^2
\end{aligned} \tag{3.15}$$

They give the following discussion on these priors:

The priors p_1 and p_2 are obtained from Jeffrey's (1966)* rule.

Prior p_1 results in a joint posterior density for μ and T , which is the product of t-densities and thus conforms most closely to standard practice. Prior p_2 incorporates the knowledge that $\sigma_A^2 > \sigma_e^2$. However the difference is only expected to matter if the F ratio for subjects is small. Priors p_3 and p_4 are flatter than p_1 and p_2 and hence lead to longer tailed posterior densities for μ and T .

Denoting the sum of squares due to subjects and error by SSS and SSE respectively the log likelihood is

$$\begin{aligned}
\ln L = & -n \ln \sigma_A - n \ln \sigma_e - n(\hat{\mu} - \mu)^2 / \sigma_A^2 - n(\hat{P} - P)^2 / \sigma_e^2 - \\
& n(\hat{T} - T)^2 / \sigma_e^2 - \text{SSS} / 2\sigma_A^2 - \text{SSE} / 2\sigma_e^2
\end{aligned} \tag{3.16}$$

Combining the prior p_1 with the likelihood and integrating out P from the posterior density yields

$$p_1(\mu, T, \sigma_e^2, \sigma_A^2 | Y) \propto \sigma_e^{-N-1} \sigma_A^{-N-1} \exp\{-\frac{1}{2}(Q_1 / \sigma_A^2 + Q_2 / \sigma_e^2)\} \tag{3.17}$$

where $Q_1 = 2n(\hat{\mu} - \mu)^2 + \text{SSS}$

$Q_2 = 2n(\hat{T} - T)^2 + \text{SSE}$

*Jeffrey's Rule for obtaining non-informative prior densities (Box and Tiao, 1973, p.54): The prior distribution for a set of parameters is taken to be proportional to the squareroot of the determinant of the information matrix.

By applying (A2.1.2) of Box and Tiao (1973) they show that the posterior density of μ and T based on prior p_1 is

$$p_1(\mu, T | \underline{Y}) = n(n-2) \{ \pi (SSS \cdot SSE)^{\frac{1}{2}} \}^{-1} \{ 1 + 2n(\hat{\mu} - \mu)^2 / SSS \}^{-n/2} \times \\ \{ 1 + 2n(\hat{T} - T)^2 / SSE \}^{-(n-1)/2} \quad (3.18)$$

The posterior probability for assessing bioequivalence is obtained by integrating (3.18) over the region defined by (3.14). This integration will have to be done numerically. Selwyn et al (1981) recommend the Gauss-Hermite procedure. They also suggest that this will be the most difficult aspect of the Bayesian method for the practising statistician.

Let C denote the constraint $\sigma_A^2 > \sigma_e^2$. Then the posterior densities that correspond to priors p_2 and p_4 are related to the unconstrained priors p_1 and p_3 as follows (see 1.5.5 of Box and Tiao 1973)

$$p_2(\mu, T | \underline{Y}) = p_1(\mu, T | \underline{Y}) P_1[C | \mu, T, \underline{Y}] / P_1[C | \underline{Y}] \quad (3.19)$$

and

$$p_4(\mu, T | \underline{Y}) = p_3(\mu, T | \underline{Y}) P_3[C | \mu, T, \underline{Y}] / P_3[C | \underline{Y}] \quad (3.20)$$

The probabilities on the right hand side of (3.19) and (3.20) are computed using p_1 and p_3 respectively. These turn out to be (details are given in Selwyn et al (1981) and Box and Tiao (1973))

$$P_1(C | \underline{Y}) = P[F_{n-1, n-2} < (n-2) SSS / \{(n-1)SSE\}] \quad (3.21)$$

and

$$P_1(C | \mu, T, \underline{Y}) = P[F_{n, n-1} < (n-1) Q_1 / nQ_2] \quad (3.22)$$

Similarly

$$p_3(\mu, T | \underline{Y}) = n(n-3) \{ \pi(SSS \cdot SSE)^{\frac{1}{2}} \}^{-1} \{ 1 + 2n(\hat{\mu} - \mu)^2 / SSS \}^{-(n-1)/2} \times \\ \{ 1 + 2n(\hat{T} - T)^2 / SSE \}^{-n/2+1} \quad (3.23)$$

and

$$P_3[C | \underline{Y}] = P[F_{n-2, n-3} < (n-3)SSS / \{(n-2)SSE\}] \quad (3.24)$$

$$P_3[C | \mu, T, \underline{Y}] = P[F_{n-1, n-2} < (n-2)Q / \{(n-1)Q_2\}] \quad (3.25)$$

3.4.2 Difference between formulation means with carry over effects

Selwyn et al (1981) incorporate a carry over effect into the Bayesian framework by putting a normal prior on the carry over. Letting R denote the carry over of the standard formulation, with prior density centred at 0 with standard deviation σ_R , the joint prior becomes

$$p(\mu, P, T, R, \sigma_e^2, \sigma_A^2) \propto \sigma_e^{-2} \sigma_A^{-2} \exp(-R^2 / 2\sigma_R^2) \quad (3.26)$$

σ_R^2 can be thought of as reflecting one's prior belief in the plausibility of a carry over effect. A σ_R close to zero indicates a strong prior belief of no carry over effect while a large σ_R corresponds to a suspected presence of a carry over effect.

They derive the conditional posterior density of μ and T given R and \underline{Y} to be

$$p(\mu, T | R, \underline{Y}) = n(n-2) \{ \pi(D \cdot SSE)^{\frac{1}{2}} \}^{-1} \{ 1 + 2n(\hat{\mu} - \mu)^2 / D \}^{-n/2} \times \\ \{ 1 + 2n(T - E)^2 / SSE \}^{-(n-1)/2} \quad (3.27)$$

where $D = SSQ + n(R - \hat{R})^2 / 2$

$$E = \hat{T} + R/2$$

SSQ = sums of squares for subjects within sequences

The posterior conditional density for R given \underline{Y} is

$$p(R|\underline{Y}) \propto [1+n(R-\hat{R})^2/\{2(n-2)SSQ\}]^{-(n-1)/2} \exp(-R^2/2\sigma_R^2) \quad (3.28)$$

3.4.3 Ratio of formulation means

Once again, the two period cross-over design is assumed. The data are assumed to be normally or log normally distributed and no carry over effect is allowed. Mandallaz and Mau (1981) derive the theory while Fluehler et al (1983) give a practical discussion via an example.

The condition of bioequivalence is accepted if the posterior probability that the ratio $\theta = \mu_N/\mu_S$ of formulation means belonging to the interval (r_1, r_2) is sufficiently large.

i.e. if $P[\theta \in (r_1, r_2)]$ is large.

For normally distributed data this posterior probability can be computed by

$$P[\theta \in (r_1, r_2)] = \int_A^B t_{n-2}(\lambda) d\lambda \quad (3.29)$$

where $t_{n-2}(\lambda)$ is the density of a t-distribution with $n-2$ df

$$A = (\hat{\theta} - r_1)n^{1/2}/(CV(1+r_1^2))^{1/2}$$

$$B = (\hat{\theta} - r_2)n^{1/2}/(CV(1+r_2^2))^{1/2}$$

$$\hat{\theta} = \bar{Y}_N/\bar{Y}_S$$

$$CV = (\text{Error mean square})^{1/2}/\bar{Y}_S$$

For log normally distributed data a similar formula applies except that \bar{Y}_N and \bar{Y}_S will denote the means of the log-transformed data and the limits of integration are given by

$$A = (\bar{Y}_N - \bar{Y}_S - \ln(r_1))n^{\frac{1}{2}}/(2^{\frac{1}{2}}S)$$

$$B = (\bar{Y}_N - \bar{Y}_S - \ln(r_2))n^{\frac{1}{2}}/(2^{\frac{1}{2}}S)$$

The derivation is given in Mandallaz and Mau (1981).

3.5 Discussion

Perhaps the most pressing question at the end of this review is: how does a pharmacologist draw useful information from a bioequivalence trial? Does he resort to frequentist methods or to Bayesian methods? Does he test hypotheses or compute confidence intervals? Should he work with the difference of two means or their ratio or perhaps even with individual ratios?

There are no clear-cut answers. The choice between frequentist and Bayesian approach is a continuing controversy, see for example Geertsema (1983). Ultimately this choice may be largely a matter of individual preference, but it should be noted that the Bayesian methods offered so far depend entirely upon the assumption that the bioavailability estimates are normally distributed. While this may be a reasonable assumption for AUC it may not be for CMAX or TMAX.

In the assessment of bioequivalence the pharmacologist must make a decision as to whether or not two substances are for all practical purposes equivalent, and also must be able to give some indication of the amount of variation there is in his estimate.

Thus it would seem that both hypothesis testing and an estimate plus confidence interval approach seem appropriate.

In the frequentist framework, for decision purposes it would appear that the methods of Anderson and Hauck (1983) and Rocke (1984) are most appropriate since the null hypothesis tested is that of non-equivalence rather than equivalence. These tests also depend on the assumption of normality or log-normality.

In giving a confidence interval again one has the choice between normal theory and non-parametric intervals. Here, as in the Bayesian approach, the normality assumption for AUC may be reasonable but not for CMAX or TMAX. Thus the non-parametric techniques offered by Steinijans and Diletti (1983, 1985) would be more appropriate and for consistency of analysis it would seem best to apply them to all three quantities.

In practice assessment of bioequivalence is essentially a multivariate problem since AUC, TMAX and CMAX are interdependent, and a decision would rarely be made on the basis of one of these alone. Classical multivariate statistical techniques depend heavily upon the assumption of joint normality of the variables and as remarked earlier with such diverse quantities as AUC, TMAX and CMAX, this assumption is difficult to justify. The pragmatic approach has been to ignore the interdependence and apply univariate procedures to each parameter individually. This affects significance levels and coverage probabilities.

Another problem in assessing bioequivalence is the definition of bioequivalence itself. Although this is a clinical problem and not a statistical one, it is inevitable that the available statistical machinery will influence the clinical definition, perhaps even dictate it. Certainly if pharmacologists wish to make use of statistics to assess

bioequivalence they are forced to make use of the available procedures and to formulate their problems conformable with these procedures.

3.6 Recommendations

In order to remedy some of the problems outlined in the previous section, we propose that Efron's (1979) bootstrap procedure should be used to assess bioequivalence. The bootstrap is extremely simple and versatile. Because of this versatility it will allow a clinical definition of bioequivalence that reflects the clinical requirements and does not depend on the available statistical procedures.

The bootstrap frees one from simple distributional assumptions and tractable mathematics that is the hallmark of traditional statistical procedures. It achieves this at the expense of computational effort. However, with the advent of modern micro computers and the low cost of computing this is no drawback at all.

An overview of the bootstrap method is given in Chapter 4 and its application to bioequivalence assessment is discussed in Chapter 5.

CHAPTER 4

BOOTSTRAP METHOD

4.1 Introduction

In 1979 an article entitled "Bootstrap Methods : Another look at the Jackknife" appeared in the Annals of Statistics. The author was Bradley Efron of Stanford University. In that article he outlined a statistical method that is extremely simply yet extremely versatile. He named the method "the bootstrap."

The bootstrap provides a non-parametric method for the following familiar problem : Given a random sample $\underline{X} = (X_1, \dots, X_n)'$ of size n from an unknown probability distribution F estimate the sampling distribution of some prespecified random variable $T(\underline{X}, F)$ on the basis of the observed data \underline{x} (Efron 1979(a), Singh (1981)).

Efron's bootstrap has proved to be a powerful and popular tool among statisticians. This is evident by the increasing number of journal articles on both theoretical and applied aspects of bootstrapping. The random variable $T(\underline{X}, F)$ is often of the following form : Let $\theta(F)$ be a functional which is to be estimated and let $\hat{\theta}(\underline{X})$ be an estimator of $\theta(F)$. Then we would define

$$\begin{aligned} T(\underline{X}, F) &= \hat{\theta}(\underline{X}) - \theta(F) \\ &= \text{estimator} - \text{parameter} \\ &= \hat{\theta} - \theta \end{aligned}$$

If $\hat{\theta}$ is an unbiased estimator of θ then expectation of T will be zero. Knowledge of the true distribution of T would enable one to determine confidence intervals for θ .

Alternatively, let $g : \mathbb{R} \rightarrow \mathbb{R}$ be monotonic, then T can have the form

$$T(\underline{X}, F) = g\{\hat{\theta}(\underline{X})\} - g\{\theta(F)\}$$

An example of a random variable $T(\underline{X}, F)$ is the arc tanh transformation of the correlation coefficient that Fisher proposed. Let ρ denote the population correlation coefficient and r the sample correlation coefficient. Fisher considered the random variable

$$\begin{aligned} T(\underline{X}, F) &= \text{arc tanh } r - \text{arc tanh } \rho \\ &= g\{\hat{\theta}(\underline{X})\} - g\{\theta(F)\} \end{aligned}$$

Here $g(t) = \text{arc tanh } (t)$ is a monotonic function. This example is discussed in more detail in section 4.11.

Another common choice for the random variable T is the "t-statistic" form

$$T(\underline{X}, F) = \frac{g\{\hat{\theta}(\underline{X})\} - g\{\theta(F)\}}{s\{\theta(F)\}}$$

Here $s\{\theta(F)\}$ is meant to serve as a scale factor. The standardised sample mean is of this form, with g being the identity function.

A useful and convenient estimator for $\theta(F)$ is

$$\hat{\theta}(\underline{X}) = \theta(F_n^X)$$

where F_n^X is the e.d.f. (empirical distribution function) of the random sample X_1, \dots, X_n . If $\theta(F)$ is a sufficiently smooth real valued functional then $\theta(F_n^X)$ is an asymptotically optimal estimate of $\theta(F)$ in the locally asymptotic minimax sense (Beran (1982)). This is however by no means the only choice for $\hat{\theta}(X)$. Any estimator may be used. Of course, the choice of estimator will affect the sampling distribution of T .

4.2 The Bootstrap Approximation

The following formal description of the bootstrap is taken from Singh (1981) : "Let $\{X_1, \dots, X_n\}$ be a random sample of size n from a population with distribution F and let $T(X_1, \dots, X_n; F)$ be a specified random variable of interest, possibly depending on the unknown distribution F . Let F_n^X denote the e.d.f. of $\{X_1, \dots, X_n\}$, i.e. the distribution that puts mass $1/n$ at each of the points X_1, \dots, X_n . The bootstrap method is to approximate the sampling distribution of $T(X_1, \dots, X_n; F)$ under F by that of $T(Y_1, \dots, Y_n; F_n^X)$ under F_n^X where $\{Y_1, \dots, Y_n\}$ denotes a random sample of size n from F_n^X ."

H will denote the sampling distribution of $T(X_1, \dots, X_n; F)$ under F ; H_b will denote the distribution of $T(Y_1, \dots, Y_n; F_n^X)$ under F_n^X and will be called the bootstrap distribution.

The bootstrap distribution H_b , of $T(Y_1, \dots, Y_n; F_n^X)$ under F_n^X , is necessarily discrete. There are, in fact, n^n ways of drawing a random sample of size n from F_n^X . Each of these has equal probability

mass of n^{-n} . Therefore, in principle at least, it is possible to systematically obtain the bootstrap distribution H_b . The difficulty is that n^n soon becomes impossibly large.

The principal methods to approximate H_b are:

Method 1. Direct theoretical calculation. This is usually very difficult in all but the simplest of examples. Efron (1979a) derives the bootstrap distribution of the median.

Method 2. Monte Carlo approximation. This is the method most frequently employed and makes the bootstrap easy to achieve using a computer. The Monte Carlo algorithm is described later in this section. See also Efron (1979a).

Method 3. An Edgeworth series approximation. This method has been used to derive asymptotic results for the bootstrap for special cases of the random variable T . The standardised sample mean is one such case although many other cases have also been considered. Bickel and Freedman (1980, 1981), Singh (1981), Beran (1982), Efron (1984a). This approximation is discussed later in this section.

The Monte Carlo algorithm for approximating the bootstrap distribution is:

(i) Construct F_n , the empirical distribution function of the observations x_1, \dots, x_n . $F_n(t) = n^{-1} \sum_{i=1}^n I(x_i \leq t)$ where I is the indicator function.

(ii) Draw a bootstrap sample Y_1, \dots, Y_n by independent random sampling from F_n . In other words make n random draws with replacement from $\{x_1, x_2, \dots, x_n\}$. Compute T for this sample.

(iii) Do step (ii) some large number "B" of times to obtain the values

$$T_1, \dots, T_B.$$

The distribution, denoted by H_B^* , that puts mass $1/B$ at each value T_1, \dots, T_B is the Monte Carlo approximation to the bootstrap distribution H_B of $T(Y_1, \dots, Y_n; F_n)$ under F_n . Note that as $B \rightarrow \infty$ the approximation, by definition, becomes exact. It is conventional not to distinguish between the bootstrap distribution H_B and the Monte Carlo approximation H_B^* , calling both "the bootstrap distribution" (Efron 1981a). We shall not follow this convention.

In using the Monte Carlo approximation one must consider a suitable choice of B . The choice of B will depend to a large extent on those aspects of the bootstrap distribution H_B in which one is interested. Efron (1981a) suggests that B in the range 50-200 is adequate for estimating standard errors but that larger values are needed for confidence interval calculations. For calculating a $1-2\alpha$ confidence interval with $\alpha = 0,16$ for a correlation coefficient he uses $B = 1000$. Of course, as α decreases so B will need to increase. However one will need to consider each application of the bootstrap on its own merits. General rules are not yet available.

The Edgeworth series expansion may be described as follows: The true sampling distribution $H(x)$ of $T(\underline{X}, F)$ under F is a function of the sample size n , the functional form of T and the distribution F . If H tends to the standard normal distribution as $n \rightarrow \infty$ then for large n one can approximate the sampling distribution $H(x)$ by the standard normal distribution $\Phi(x)$. An example of such a random variable is the standardized sample mean. Many others, especially those that depend on

sums of variates, like the moments also tend to normality under a central limit effect. However, this approximation by the standard normal distribution effectively uses only the first two moments. Remembering that some random variables tend to normality much more rapidly than others there may be advantage in utilising higher moments or cumulants.

In these circumstances the Edgeworth series may provide an approximation to $H(x)$ which is an improvement over $\Phi(x)$. The Edgeworth series achieves this improvement by making use of the first four moments or cumulants of $H(x)$.

Suppose that $H(x)$ has zero mean and unit standard deviation while its skewness γ_1 and excess γ_2 are (γ_1, γ_2 are the 3rd and 4th cumulants (Kendall and Stuart (1977) p.88) :

$$\gamma_1 = \mu_3/\sigma^3 \quad \text{or} \quad \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$$

$$\gamma_2 = \mu_4/\sigma^4 - 3 \quad \text{or} \quad \gamma_2 = \mu_4/\mu_2^2 - 3$$

The Edgeworth approximation to $H(x)$ up to terms of order n^{-1} is (see Cox and Hinkley (1974) p.464).

$$G(x) = \Phi(x) - \phi(x) \left\{ \frac{\gamma_1}{6\sqrt{n}} H_2(x) + \frac{\gamma_2}{24n} H_3(x) + \frac{\gamma_1^2}{72n} H_5(x) \right\}$$

where $H_r(x)$ denotes the r th degree Hermite polynomial defined by

$$H_r(x) = (-1)^r e^{\frac{1}{2}x^2} D^r(e^{-\frac{1}{2}x^2})$$

Here $D = \frac{d}{dx}$ is the differential operator and $\Phi(x)$ and $\phi(x)$ are, respectively, the standard normal distribution and density functions.

The three Hermite polynomials of interest are (Kendall and Stuart (1977) p.167) .

$$H_2(x) = x^2 - 1$$

$$H_3(x) = x^3 - 3x$$

$$H_5(x) = x^5 - 10x^3 + 15x$$

In order for $G(x)$ to be of practical value γ_1 and γ_2 need to be expressed in terms of the parent population F . The r th moment of T , that is the r th moment of its sampling distribution $H(x)$ is (Kendall and Stuart (1977) p.243).

$$E(T^r) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T^r dF(x_1) \dots dF(x_n)$$

This will express moments of H in terms of moments of F . The Edgeworth approximation to the bootstrap distribution $H_b(x)$ is obtained by replacing F with F_n^X . In this case γ_1 and γ_2 are the skewness and excess of $H_b(x)$ and are expressed in terms of the c.d.f. F_n^X . We shall denote the Edgeworth approximation to the bootstrap distribution by $H_e(x)$.

4.3 Standard Errors

How does one use the bootstrap to estimate the standard error of a statistic $T = T(X_1, \dots, X_n; F)$? Our development will closely follow that of Efron (1981a). The true standard error of T is a function of the sample size n , the functional form of T and the distribution F . We denote this standard error by $\sigma(n, T, F)$.

$$\sigma = \sigma(n, T, F) = [\text{VAR}_F T(X_1, \dots, X_n; F)]^{\frac{1}{2}} \quad (4.1)$$

The only unknown in this expression is F which we estimate using the e.d.f. F_n^X . The bootstrap estimate of σ , denoted by $\hat{\sigma}_b$, is

$$\hat{\sigma}_b = \sigma(n, T, F_n^X) = [\text{VAR}_{F_n^X} T(Y_1, \dots, Y_n; F_n^X)]^{\frac{1}{2}} \quad (4.2)$$

For most statistics $\hat{\sigma}_b$ cannot be expressed in closed form. However it is easily approximated by Monte Carlo methods : Given the Monte Carlo values T_1, \dots, T_B $\hat{\sigma}_b$ is approximated by

$$\hat{\sigma}_b^* = \left[\left\{ \sum_{i=1}^B (T_i - T_{\cdot})^2 \right\} / (B-1) \right]^{\frac{1}{2}} \quad (4.3)$$

$$T_{\cdot} = B^{-1} \sum_{i=1}^B T_i \quad (4.4)$$

Notice that we are using an approximation chain here : $\hat{\sigma}_b^*$ is the Monte Carlo approximation to $\hat{\sigma}_b$ which in turn is the bootstrap approximation to σ .

$$\text{i.e. } \hat{\sigma}_b^* \xrightarrow[B]{\infty} \hat{\sigma}_b \xrightarrow[n]{\infty} \sigma \quad (4.5)$$

4.4 Confidence intervals

Suppose that our problem is to construct a confidence interval for the functional $\theta(F)$ using the bootstrap method. By a confidence interval we mean lower and upper points $L = L(\underline{X})$ and $U = U(\underline{X})$ such that $P[L < \theta < U] = 1-2\alpha$. Our discussion will be confined to central confidence intervals, i.e. intervals (L, U) such that $P[\theta < L] = P[\theta > U] = \alpha$.

Tibshirani (1984) discusses five methods for constructing a central confidence interval for θ using the bootstrap. These are: the pivotal, generalised pivotal, percentile, bias-corrected percentile and the tilting methods. Tibshirani's paper expands the methods proposed by Efron (1981a). We will restrict attention to the first four methods listed above.

4.5 The Simple Pivotal Interval

Suppose that the random variable of interest has the form

$$T(\underline{X}, F) = \hat{\theta}(\underline{X}) - \theta(F)$$

or
$$T = \hat{\theta} - \theta \sim H$$

The bootstrap approximation to this is

$$T(\underline{Y}, F_n^X) = \hat{\theta}(\underline{Y}) - \theta(F_n^X)$$

$$T_b = \hat{\theta}^* - \hat{\theta} \sim H_b$$

In order to construct a confidence interval for θ we make two assumptions:

A1 : the distribution H does not involve θ .

In other words we assume that

$$T = \hat{\theta} - \theta \text{ is pivotal.}$$

A2 : the bootstrap distribution H_b closely approximates the distribution H .

Assumption A2 is based on the premise that if F_n is close to F , the bootstrap distribution of $\hat{\theta}^* - \theta$ will be close to $\hat{\theta} - \theta$, as long as $\theta(\cdot)$ is a reasonably smooth functional.

Under A1 and A2 we have

$$1-2\alpha = P[H^{-1}(\alpha) < \hat{\theta} - \theta < H^{-1}(1-\alpha)]$$

$$\therefore 1-2\alpha = P[\hat{\theta} - H^{-1}(1-\alpha) < \theta < \hat{\theta} - H^{-1}(\alpha)] \quad (4.6)$$

where $H^{-1}(k)$ denotes the k th percentile of H and $\hat{\theta}$ is the value obtained from the original sample.

Since H is unknown we approximate $H^{-1}(\cdot)$ by $H_b^{-1}(\cdot)$.

$$1-2\alpha = P[\hat{\theta} - H_b^{-1}(1-\alpha) < \theta < \hat{\theta} - H_b^{-1}(\alpha)]$$

$$\text{so } (\hat{\theta} - H_b^{-1}(1-\alpha), \hat{\theta} - H_b^{-1}(\alpha)) \quad (4.7)$$

is a central $(1-2\alpha)100\%$ Bootstrap Pivotal interval.

Of course as has been pointed out previously, one seldom has the bootstrap distribution H_b but rather the Monte Carlo approximation H_b^* . Therefore in practice a $(1-2\alpha)100\%$ Bootstrap Pivotal interval is

$$(\hat{\theta} - H_b^{*-1}(1-\alpha), \hat{\theta} - H_b^{*-1}(\alpha)) \quad (4.8)$$

Notice again the approximation chain

$$(\hat{\theta} - H_b^{*-1}(1-\alpha), \hat{\theta} - H_b^{*-1}(\alpha)) \quad (4.9)$$

$B \downarrow \infty$

$$(\hat{\theta} - H_b^{-1}(1-\alpha), \hat{\theta} - H_b^{-1}(\alpha)) \quad (4.10)$$

$n \downarrow \infty$

$$(\hat{\theta} - H^{-1}(1-\alpha), \hat{\theta} - H^{-1}(\alpha)) \quad (4.11)$$

Interval (4.8) can be expressed in a different manner as follows:

Let $\hat{CDF}(t)$ be the Monte Carlo approximation to the bootstrap distribution of $\hat{\theta}^*$

$$\text{i.e. } \hat{CDF}(t) = H_b^*(t - \hat{\theta}) = \frac{\#(\hat{\theta}^* \leq t)}{B} \quad \forall t \in \mathbb{R}$$

$$\text{and } H_b^{*-1}(\alpha) = \hat{CDF}^{-1}(\alpha) - \hat{\theta}. \quad \forall 0 < \alpha < 1$$

Substituting this into (4.8) the bootstrap pivotal interval becomes

$$(2\hat{\theta} - \hat{CDF}^{-1}(1-\alpha), 2\hat{\theta} - \hat{CDF}^{-1}(\alpha)) \quad (4.12)$$

Example: Consider a random sample X_1, \dots, X_n from the $N(\mu, 1)$ distribution and let $\hat{\theta} = \sqrt{n} \bar{X}$ and $\theta = \sqrt{n} \theta$.

Here $T = \sqrt{n}(\hat{\theta} - \theta)$ is pivotal with true distribution $H(x)$ which is equal to the standard normal distribution $\Phi(x)$. The usual $(1-2\alpha)100\%$ confidence interval is

$$(\bar{X} - H^{-1}(1-\alpha); \bar{X} - H^{-1}(\alpha))$$

or, since in this case $H(x) = \Phi(x)$

$$(\bar{X} - \Phi^{-1}(1-\alpha), \bar{X} - \Phi^{-1}(\alpha)).$$

However this confidence interval may be expressed in terms of the distribution of $\hat{\theta} = \sqrt{n} \bar{X}$ which is $N(\mu, 1)$, say $C(x)$. We have

$$C(x - \theta) = \Phi(x) = H(x)$$

and hence

$$\begin{aligned} H^{-1}(\alpha) &= \Phi^{-1}(\alpha) = C^{-1}(\alpha) - \theta \\ &\cong C^{-1}(\alpha) - \bar{X}. \end{aligned}$$

Substituting this into the above interval, we obtain

$$(2\bar{X} - C^{-1}(1-\alpha), 2\bar{X} - C^{-1}(\alpha))$$

If the parent distribution is not known and T is pivotal, we replace the true distribution $H(x)$ by its bootstrap estimate, either H_b or H_b^* and replace C by $\hat{C}DF$. The two intervals are then

$$(\bar{X} - H_b^{*-1}(1-\alpha), \bar{X} - H_b^{*-1}(\alpha))$$

and

$$(2\bar{X} - \hat{C}DF^{-1}(1-\alpha), 2\bar{X} - \hat{C}DF^{-1}(\alpha))$$

4.6 Generalised Pivotal Intervals

For the simple pivotal we assumed the following form for the random variable of interest

$$T = T(\underline{X}, F) = \hat{\theta} - \theta$$

We now suppose that T is an arbitrary, but known, function of $\hat{\theta}$ and θ , say $t(\hat{\theta}, \theta)$. We also require that $t(\hat{\theta}, \theta)$ be monotone in θ . Let the inverse of $t(\cdot, \cdot)$ with respect to the second argument be $t_2^{-1}(\cdot)$. Again we assume that random variable

$$T(\underline{X}, F) = t(\hat{\theta}, \theta) \sim H \quad (4.13)$$

is pivotal and that the bootstrap distribution of

$$T(Y, F_n^X) = t(\hat{\theta}^*, \hat{\theta}) \sim H_b \quad (4.14)$$

closely approximates the true distribution H .

From (4.13) and the pivotal assumption we have

$$1-2\alpha = P[H^{-1}(\alpha) < t(\hat{\theta}, \theta) < H^{-1}(1-\alpha)]$$

If $t(\cdot, \cdot)$ is monotone increasing in θ then

$$1-2\alpha = P[t_2^{-1}\{H^{-1}(\alpha)\} < \theta < t_2^{-1}\{H^{-1}(1-\alpha)\}] \quad (4.15)$$

Approximating $H^{-1}(\cdot)$ by $H_b^{-1}(\cdot)$ or its Monte Carlo approximation $H_b^{*-1}(\cdot)$ we obtain the generalised bootstrap pivotal interval

$$(t_2^{-1}\{H_b^{*-1}(\alpha)\}, t_2^{-1}\{H_b^{*-1}(1-\alpha)\}) \quad (4.16)$$

If $t(\cdot, \cdot)$ is monotone decreasing in θ then the interval is

$$(t_2^{-1}\{H_b^{*-1}(1-\alpha)\}, t_2^{-1}\{H_b^{*-1}(\alpha)\}) \quad (4.17)$$

4.7 Percentile intervals

If in addition to assumptions A1 and A2 we can further assume that

A3: H is symmetric around 0

$$\text{i.e. } H(-t) = 1 - H(t) \quad \forall t \in \mathbb{R}$$

$$\text{and } H^{-1}(1-\alpha) = -H^{-1}(\alpha) \quad \forall 0 < \alpha < 1$$

then (4.12) becomes

$$(\hat{CDF}^{-1}(\alpha), \hat{CDF}^{-1}(1-\alpha)) \quad (4.18)$$

To see this note that if H is symmetric around zero, then H_b should also be close to symmetric around zero hence also H_b^* . Therefore \hat{CDF} is (close to) symmetric around $\hat{\theta}$.

$$\therefore \frac{1}{2}[\hat{CDF}^{-1}(\alpha) + \hat{CDF}^{-1}(1-\alpha)] = \hat{\theta}$$

$$\therefore \hat{CDF}^{-1}(1-\alpha) = 2\hat{\theta} - \hat{CDF}^{-1}(\alpha)$$

$$\text{and } \hat{CDF}^{-1}(\alpha) = 2\hat{\theta} - \hat{CDF}^{-1}(1-\alpha)$$

4.8 Generalisation of the Percentile Interval

Let us suppose that a symmetric pivot exists on some other scale, i.e.

$$A4: g(\theta) - g(\hat{\theta}) \sim H$$

and

$$A5: g(\hat{\theta}^*) - g(\hat{\theta}) \sim H_b$$

with H symmetric around zero and $g(\cdot)$ an unknown, monotone increasing function, then as for (4.15) we get a central confidence interval for $g(\theta)$:

$$(G_b^{-1}(\alpha), G_b^{-1}(1-\alpha)) \quad (4.19)$$

where G_b is the bootstrap distribution of $g(\hat{\theta}^*)$. Transforming back to the θ scale gives a confidence interval for θ :

$$(g^{-1}(G^{-1}(\alpha)) ; g^{-1}(G^{-1}(1-\alpha))) \quad (4.20)$$

$$= (\widehat{\text{CDF}}^{-1}(\alpha) ; \widehat{\text{CDF}}^{-1}(1-\alpha)) \quad (4.21)$$

which is again a percentile interval. A similar argument applies if $g(\cdot)$ is monotonic decreasing. Thus the percentile interval has the correct coverage if a symmetric pivotal exists on any scale. The real benefit is that we do not have to know $g(\cdot)$ because the resultant interval (4.21) does not depend on $g(\cdot)$.

4.9 Bias-Corrected Percentile Intervals

If the distribution H in assumption A4 is symmetric around some value $\mu \neq 0$, the percentile interval will be biased and may not have the correct coverage. This would happen if $\hat{\theta}$ was a biased estimator of θ . It is possible to estimate μ and derive a corrected interval provided that we are willing to assume a parametric form for H .

Let H be a symmetric distribution that belongs to a symmetric location-scale family, say $H(x|\mu, \sigma) = H_0(\frac{x-\mu}{\sigma})$. Initially, we consider the case with $\sigma = 1$. We assume that a pivot exists on some scale and that

$$g(\hat{\theta}) - g(\theta) \sim H_0(x-\mu) \quad (4.21)$$

and

$$g(\hat{\theta}^*) - g(\hat{\theta}) \sim H_b(x-\mu) \quad (4.22)$$

where g is some monotonic function.

We can estimate μ by noting that

$$P[g(\hat{\theta}) - g(\theta) < 0] = H_0(-\mu) \quad (4.23)$$

and

$$P[g(\hat{\theta}^*) - g(\hat{\theta}) < 0] = H_b(-\mu) \quad (4.24)$$

But, since g is monotonic

$$P[g(\hat{\theta}^*) < g(\hat{\theta})] = P[\hat{\theta}^* < \hat{\theta}] = \text{CDF}(\hat{\theta}) \quad (4.25)$$

Hence, from (4.24) and (4.25) we have

$$\mu = -H_b^{-1}\{\text{CDF}(\hat{\theta})\} \quad (4.26)$$

since by assumption $H_b \cong H_0$ we estimate μ by

$$b = -H_0^{-1}\{\text{CDF}(\hat{\theta})\}. \quad (4.27)$$

In order to derive a confidence interval for θ we consider the distribution of $g(\theta) - g(\hat{\theta})$:

By assumption

$$\begin{aligned} g(\hat{\theta}) - g(\theta) &\sim H(\mu) \\ \Leftrightarrow g(\hat{\theta}) - g(\theta) - \mu &\sim H_0 \\ \Leftrightarrow g(\theta) - g(\hat{\theta}) + \mu &\sim H_0 \quad \because H_0 \text{ symmetric about zero} \\ \Leftrightarrow g(\theta) - g(\hat{\theta}) &\sim H_0 - \mu \end{aligned}$$

Define a distribution function D^g as follows:

$$\begin{aligned} D^g(g(t)) &= P[g(\theta) - g(\hat{\theta}) < g(t) - g(\hat{\theta})] \\ &= P[g(\theta) - g(\hat{\theta}) + \mu < g(t) - g(\hat{\theta}) + \mu] \\ &= H_0(g(t) - g(\hat{\theta}) + \mu) \end{aligned} \quad (4.28)$$

Note that the quantiles of D^g would define a confidence interval for $g(\theta) - g(\hat{\theta})$, hence also for $g(\theta)$ and for θ . We shall see that g does not need to be known.

Next we consider the bootstrap distribution and its Monte Carlo approximation.

$$\begin{aligned}
 \hat{CDF}(t) &\cong P[\hat{\theta}^* \leq t] \\
 &= P[g(\hat{\theta}^*) \leq g(t)] \\
 &= P[g(\hat{\theta}^*) - g(\hat{\theta}) - \mu \leq g(t) - g(\hat{\theta}) - \mu] \\
 &= H_b\{g(t) - g(\hat{\theta}) - \mu\} \\
 &\cong H_0\{g(t) - g(\hat{\theta}) - \mu\}
 \end{aligned} \tag{4.29}$$

Solving (4.29) for $g(t) - g(\hat{\theta})$ we get

$$g(t) - g(\hat{\theta}) \cong H_0^{-1}\{\hat{CDF}(t)\} + \mu \tag{4.30}$$

Substituting (4.30) for $g(t) - g(\hat{\theta})$ and (4.27) for μ into (4.28), we obtain the approximation

$$D^g(g(t)) \cong H_0[H_0^{-1}\{\hat{CDF}(t)\} + 2b] \tag{4.31}$$

Setting (4.31) equal to α and $1-\alpha$ and solving for t we get a $1-2\alpha$ confidence interval (L,U) for θ where

$$L = \hat{CDF}^{-1}[H_0[H_0^{-1}(\alpha) + 2H_0^{-1}\{\hat{CDF}(\hat{\theta})\}]] \tag{4.32}$$

$$U = \hat{CDF}^{-1}[H_0[H_0^{-1}(1-\alpha) + 2H_0^{-1}\{\hat{CDF}(\hat{\theta})\}]] \tag{4.33}$$

If $\sigma \neq 1$ and we repeat the above derivation, we get

$$b = \frac{\mu}{\sigma} = -H_0^{-1}\{\hat{CDF}(\hat{\theta})\}$$

and we obtain the same interval as defined by (4.32) and (4.33)

Note that when $b = 0$, the bias corrected percentile interval reduces to the percentile interval.

Tibshirani (1984) has compared the amount of bias correction (that is $H_0^{-1}[\alpha) + 2H_0^{-1}\{\widehat{CDF}(\widehat{\theta})\}]$) for the normal, logistic and Cauchy distributions when $\alpha = 0.05$ and for various choices of $H_0^{-1}\{\widehat{CDF}(\widehat{\theta})\}$. He concludes that the choice of symmetrical distribution H_0 appears to make little difference.

4.10 Discussion of Bootstrap confidence intervals

Four methods of constructing bootstrap confidence intervals have been derived : the pivotal, the generalised pivotal, the percentile and the bias-corrected percentile. Each of these methods differ in their assumptions.

In order to construct a pivotal interval we had to specify the exact form of the pivot but nothing was assumed about the pivotal distribution. In order to build a percentile interval we did not have to know the exact form of the pivot, g , but we did assume that the pivotal distribution was symmetric about zero. For the bias-corrected percentile interval we assumed a parametric distribution, symmetric about some point μ .

In order to check whether or not a random variable $t(\widehat{\theta}, \theta)$ is a pivotal quantity Hinkley (1983) has suggested that one should "bootstrap the bootstrap." Specifically, let $Z_1 = (Y_1^*, \dots, Y_n^*)$ be a random sample of size n from the distribution F_n and let G_n^1 be the e.d.f. of Z_1 . Suppose now that one has M such random samples Z_1, \dots, Z_M with corresponding c.d.f.'s $G_n^1, G_n^2, \dots, G_n^M$. If one now performs a separate

bootstrap on each of these distributions one will have a check on the distribution of $t(\hat{\theta}, \theta)$ for a variety of values of $\theta_i = \theta(G_n^i)$. Denote the bootstrap distribution of $t(\hat{\theta}, \theta)$ under G_n^i by H_{bi} and its Monte Carlo approximation by H_{bi}^* . If $t(\hat{\theta}, \theta)$ is indeed a pivotal random variable then we expect the distribution of $t(\hat{\theta}, \theta)$ under G_n^i to be identical for every G_n^i i.e. we expect $H_{b1}, H_{b2}, \dots, H_{bM}$ all to be identical. A plot of quantiles of H_{bi} or, perhaps H_{bi}^* vs θ_i should produce lines of constant height. A trend upward or downward or a change of spread would be symptomatic of a non-pivotal quantity.

4.11 Examples

Tibshirani (1984) gives three examples to illustrate the theory. We shall expand on his discussion. In each case, the data are assumed to be Gaussian. We also cite an example from Efron (1984a).

Example 1 The Mean : Let $\theta = E(X)$, $\hat{\theta} = \bar{X} = \theta(F_n^X)$. Suppose the variance σ^2 is known. Since $n^{1/2}(\hat{\theta} - \theta)/\sigma$ is pivotal with symmetric distribution the bootstrap pivotal interval and the percentile interval will both apply. We expect both methods to yield similar results, each with approximately the correct coverage.

Example 2 The Correlation Coefficient : Let $\theta = \rho$ be the population correlation coefficient and $\hat{\theta} = r = \theta(F_n^X)$ the sample correlation coefficient. By the familiar arc tanh transformation due to Fisher we know that the r.v. $T = \text{arc tanh } r - \text{arc tanh } \rho$ is approximately $N(\theta/(2(n-1)); 1/(n-3))$. Here T is biased but has a symmetric, parametric pivotal distribution about $\mu = \theta/(2(n-3))$. Hence the bias corrected pivotal interval using the normal family should yield limits

with the correct coverage. The uncorrected percentile interval, Efron's initial example, is expected to be biased (Efron, 1979a, 1983).

Example 3 The Variance : Let $\theta = E(X-E(X))^2 = \int (x-\mu)^2 dF(x)$.

This provides an interesting example where one may want to choose $\hat{\theta}$ different from $\theta(F_n^X) = \Sigma(X_i - \bar{X})^2/n$ which is a biased estimate of θ .

We shall compare the two familiar estimators

$$\hat{\theta}_1 = \theta(F_n^X) = \Sigma(X_i - \bar{X})^2/n \quad E(\hat{\theta}_1) = (n-1)\theta/n$$

and

$$\hat{\theta}_2 = \Sigma(X_i - \bar{X})^2/(n-1) \quad E(\hat{\theta}_2) = \theta$$

Case 1 : $\hat{\theta}_1 = \Sigma(X_i - \bar{X})^2/n$

It is well known that the random variable

$$t_1(\hat{\theta}, \theta) = n\hat{\theta}_1/\theta \sim \chi^2(n-1) = H$$

Since $\chi^2(n-1)$, or H , does not depend on θ it follows that t_1 is a pivotal quantity. Hence we expect the generalised pivotal interval to have the correct coverage. Denoting the Monte Carlo approximation to the bootstrap distribution of $t_1(\hat{\theta}_1, \theta)$ by H_B^* and the inverse of $t_1(.,.)$ w.r.t. the second argument by $t_2^{-1}(.)$ the generalized percentile interval is obtained using the interval (4.17)

$$(t_2^{-1}\{H_B^{*-1}(1-\alpha)\}, t_2^{-1}\{H_B^{*-1}(\alpha)\}) \quad (4.34)$$

Note that since the true sampling distribution H_1 of $t_1(\hat{\theta}_1, \theta)$ is not symmetric one would not expect the percentile or bias corrected percentile interval to give the correct coverage unless of course a symmetric pivotal exists on some other scale. Notice also that one could apply the bootstrap to the random variable

$\ln(t_1/n) = \ln \hat{\theta} - \ln \theta \sim \ln \chi^2$. Again the generalised pivotal

interval should give the correct coverage, with the obvious adjustments.

Case 2 : $\hat{\theta} = \Sigma(X_i - \bar{X})/(n-1)$

Here we have

$$t_2(\hat{\theta}, \theta) = (n-1)\hat{\theta}/\theta \sim \chi^2(n-1) \sim H_1$$

Like t_1 , this random variable t_2 is pivotal.

By an analogous argument the generalised pivotal interval is appropriate. Denoting the inverse of $t_2(\cdot, \cdot)$ w.r.t. the second argument by $t_2^{-1}(\cdot)$, the required interval is

$$(t_2^{-1}\{H_b^{*-1}(1-\alpha)\}, t_2^{-1}\{H_b^{*-1}(\alpha)\}) \quad (4.35)$$

The bootstrap could be applied to the random variable

$$\ln(t_2/(n-1)) = \ln \hat{\theta}_2 - \ln \theta.$$

We expect both (4.34) and (4.35) to yield the correct coverage.

Shenker (1985) gives a discussion of this example and shows that the percentile type intervals do not give satisfactory results.

Example 4 Ratio Estimation : This example is taken from Efron (1984) and has strong relevance to the problem of bioequivalence. He discusses sampling from a bivariate normal population with mean vector $\underline{\mu} = (\mu_1, \mu_2)$ and identity covariance matrix i.e. $N(\underline{\mu}, I)$. He shows that the bootstrap bias corrected percentile interval for $\theta = \mu_2/\mu_1$ using the maximum likelihood estimate $\hat{\theta} = \bar{X}_2/\bar{X}_1$ for θ agree closely with the intervals obtained from the exact Fieller (1954) distribution. He also shows that the uncorrected bootstrap percentile interval gives the Creasy (1954) fiducial solution.

4.12 Asymptotic accuracy of the bootstrap

In order to indicate dependence on the sample size n , we shall for the present discussion denote $H(x)$ by $H(x,n)$, $H_b(x)$ by $H_b(x,n)$ and $H_e(x)$ by $H_e(x,n)$.

How good is the bootstrap approximation? This question cannot be answered in general but various special cases have been considered in detail. Although the bootstrap would probably only be used in practice when the sampling distribution could not be derived analytically it is important to check the behaviour of the bootstrap in situations which are simple enough to be handled analytically. Efron (1979a) gives a number of examples in which the bootstrap works. He also establishes that the method works for a general class of statistics when the sample space is finite. Singh (1981) gives a detailed account of the bootstrap in the case of the standardised sample mean and sample quantile.

Bickel and Freedman (1981) show that the bootstrap method works for means; for pivotal quantities of the "t-statistic" sort; and their multivariate extensions; U-statistics and other von Mises functionals; the empirical process; the quantile process; and Trimmed means and Winsorised variances. They also give examples where the bootstrap fails, for instance, when estimating θ from variables uniformly distributed over the interval $[0,\theta]$. Beran (1982) establishes that $H_b(x,n)$ is asymptotically minimax; the loss function being any bounded monotone increasing function of a certain norm on the scaled difference $n^{\frac{1}{2}}\{H(x,n) - H_b(x,n)\}$. He also establishes that the estimated first order Edgeworth expansion $H_e(x,n)$ is also asymptotically minimax and is equivalent to $H_b(x,n)$ up to terms of order $n^{-\frac{1}{2}}$. By comparison the straight forward normal approximation, with estimated variance, is

usually not asymptotically minimax, because of bias. Efron (1984a) discusses the use of the bootstrap in setting confidence intervals for a real valued function θ of an unknown parameter vector η when sampling from the family of densities $f(x; \eta)$. He considers sampling from a bivariate normal distribution $N(x; \eta, I)$, $\eta = (\eta_1, \eta_2)$ and setting confidence limits for $\theta = t(\eta)$ a well behaved function. As examples he considers $\theta_1 = \eta_2/\eta_1$; $\theta_2 = \|\eta\|$ and $\theta_3 = \eta_1 \eta_2$. Within this class of problems the bootstrap bias corrected percentile interval improves on the asymptotic normal approximation. For the more complicated problem of setting a confidence interval for θ , having observed $\bar{y} \sim \theta \chi_1^2/g$, the bootstrap bias corrected percentile interval gives only a partial improvement over the asymptotic normal approximation. An interesting point that Efron (1984a) mentions; when sampling from a parametric family the estimated Edgeworth expansion \hat{H}_e is equivalent to using $B = \infty$ for the Monte Carlo estimate H_b^* .

Singh (1981) considers the following basic cases of $T(\underline{X}, F)$; $\bar{X}_n - \mu$, $(\bar{X}_n - \mu)/\sigma$ and $F_n^{-1}(t) - F^{-1}(t)$ where $\bar{X}_n = \sum X_i/n$, $\mu = E_F(X)$, $0 < \sigma^2 = V_F(X)$ and F_n^{-1} and F^{-1} are the right continuous versions of F_n and F respectively, at some fixed value $t \in (0, 1)$.

The essence of his first theorem is as follows: the bootstrap approximation to the distribution of $n^{1/2}(\bar{X}_n - \mu)/\sigma$ is better than the approximation by the limiting normal distribution, provided that the underlying distribution is non-lattice. He establishes that the difference in accuracy between the two approximations decreases with decreasing skewness of the underlying distribution and is non-existent for symmetric distributions.

His second theorem considers the bootstrap and normal approximations to the distribution of $n^{\frac{1}{2}}(F_n^{-1}(t) - F^{-1}(t))$. If $F'(F^{-1}(t))$ is known exactly then the normal approximation is better. If this quantity is not known, as is usually the case, then the bootstrap approximation is as good as the normal approximation.

Referring to the definition of the bootstrap given in section 4.2 we introduce the following notation : $\bar{Y}_n = \Sigma Y_i/n$, $S_n^2 = \Sigma (X_i - \bar{X})^2/n$, $\mu_3 = E_F(X-\mu)^3$, $\hat{\mu}_3 = \Sigma (X_i - \bar{X})^3/n$ and $\rho = E_F|X-\mu|^3$. P and P^* denote probabilities under F and F_n^X ; E and E^* denote expectations under F and F_n^X , respectively. The norm $\| \cdot \|$ under consideration is the sup-norm $\sup_{x \in \mathbb{R}} | \cdot |$.

We now state the theorems. The proofs are given in Singh (1981).

Theorem 1

A : If $EX^2 < \infty$, then for $T = n^{\frac{1}{2}}(\bar{X}_n - \mu)$

$$\|H(x,n) - H_b(x,n)\| \rightarrow 0 \quad \text{a.s.} \quad (4.36)$$

B : If $EX^4 < \infty$, then for

$$T = n^{\frac{1}{2}}(\bar{X}_n - \mu) \quad \text{and} \quad T_b = n^{\frac{1}{2}}(\bar{Y}_n - \bar{X}_n)$$

$$\begin{aligned} & \text{Lim sup}_{n \rightarrow \infty} n^{\frac{1}{2}}(\log \log n)^{-\frac{1}{2}} \|H(x,n) - H_b(x,n)\| \\ &= (2\sigma^2(2\pi e)^{\frac{1}{2}})^{-1} (2V_F((X-\mu)^2))^{\frac{1}{2}} \quad \text{a.s.} \end{aligned} \quad (4.37)$$

C : If $E|X|^3 < \infty$, then for

$$T = n^{\frac{1}{2}}(\bar{X}_n - \mu)/\sigma \quad \text{and} \quad T_b = n^{\frac{1}{2}}(\bar{Y}_n - \bar{X}_n)/S_n$$

$$\text{Lim sup}_{n \rightarrow \infty} \rho \sigma^{-3} n^{\frac{1}{2}} \|H(x,n) - H_b(x,n)\| \leq 2K \quad \text{a.s.} \quad (4.38)$$

where K is the universal appearing in the Berry-Esséen bound.

D : If $E|X|^3 < \infty$ and F is non-lattice, then for $T = n^{\frac{1}{2}}(\bar{X}_n - \mu)/\sigma$
and $T_b = n^{\frac{1}{2}}(\bar{Y}_n - \bar{X}_n)/S_n$

$$H_b(x, n) = \Phi(x) + \{\mu_3(1-x^2)/(6\sigma^3 n^{\frac{1}{2}})\}\phi(x) + o(n^{-\frac{1}{2}}) \quad (4.39)$$

or

$$H_b(x, n) = H_e(x, n) + o(n^{-\frac{1}{2}})$$

uniformly in x a.s. where $\Phi(x)$ and $\phi(x)$ are the standard normal distribution function and density respectively; therefore in this case

$$n^{\frac{1}{2}}\|H(x, n) - H_b(x, n)\| \rightarrow 0 \quad \text{a.s.} \quad (4.40)$$

E : If $E|X|^3 < \infty$ and F is lattice with span h then for

$$T = n^{\frac{1}{2}}(\bar{X}_n - \mu)/\sigma \quad \text{and} \quad T_b = n^{\frac{1}{2}}(\bar{Y}_n - \bar{X}_n)/S_n$$

$$H_b(x, n) = \Phi(x) + \{\mu_3(1-x^2)/(6\sigma^2 n^{\frac{1}{2}})\}\phi(x) +$$

$$\{h/(\sigma n^{\frac{1}{2}})\}g(n^{\frac{1}{2}}s_n x/h)\phi(x) + o(n^{-\frac{1}{2}}) \quad (4.41)$$

uniformly in x a.s. where $g(t) = [y] - y + \frac{1}{2} \quad \forall t \in \mathbb{R}$.

Also, in this case

$$\text{Lim sup}_{n \rightarrow \infty} n^{\frac{1}{2}}\|H(x, n) - H_b(x, n)\| = h(2\pi\sigma^2)^{-\frac{1}{2}} \quad \text{a.s.} \quad (4.42)$$

Theorem 2 : If F has bounded second derivative in a neighbourhood of $F^{-1}(t)$ and $F'(F^{-1}(t)) > 0$, then a.s. for $T = n^{\frac{1}{2}}(F_n^{-1}(t) - F^{-1}(t))$

$$\text{and} \quad T_b = n^{\frac{1}{2}}(G_n^{-1}(t) - F_n^{-1}(t))$$

$$\text{Lim sup}_{n \rightarrow \infty} n^{\frac{1}{4}}(\log \log n)^{-\frac{1}{2}}\|H(x, n) - H_b(x, n)\| = K_{t, F} \quad (4.43)$$

a constant depending upon t and F only. Here $G_n(t)$ is the e.d.f. of Y_1, \dots, Y_n .

Remarks

- 1) Parts A and B of Theorem 1 establish the uniform convergence to zero of the normed distance between the actual distribution H of $n^{\frac{1}{2}}(\bar{X}-\mu)$ and the bootstrap approximation H_b of it.
- 2) Parts C, D and E discuss the same convergence problem for the distribution of $n^{\frac{1}{2}}(\bar{X}-\mu)/\sigma$. Part D establishes that for non-lattice distributions the bootstrap has the edge over the asymptotic normal approximation. Part E establishes that this convergence is not valid for distributions defined on a lattice. However Part E does show the effect of rounding data on the bootstrap approximation. In equation (4.42) one may use the quantity h as the rounding error in the sampled values.
- 3) The Edgeworth expansion given in (4.39) demonstrates why the bootstrap approximation has the edge over the limiting normal distribution if the sampling is from a skew distribution F . This expansion also supplies an alternative to the Monte Carlo method of approximating the bootstrap distribution. One would need to estimate μ_3 and σ^2 using the sample values X_1, \dots, X_n .
- 4) The bootstrap distribution $H_b(x)$ can be expanded up to as many terms as one wants provided that the Cramer conditions are imposed on the distribution F . Singh (1981) uses the three term expansion to establish that

$$\|H(x,n) - H_b(x,n)\| = O\{n^{-1}(\log \log n)^{\frac{1}{2}}\}$$

provided that $E|X|^6 < \infty$ and the Cramer condition about F holds.

The Cramer conditions are given in Kendall and Stuart (1977) vol. 1 p.173. They give the warning as do Cox and Hinkley (1974) that this series is only useful in cases of moderate skewness.

- 5) Theorem 2 establishes the consistency of the bootstrap approximation of $n^{1/2}(F_n^{-1}(t) - F^{-1}(t))$ and provides the exact rate at which the normed distance between the two distributions converges to zero.

4.13 Conditions for Bootstrapping

Beran (1982) gives a counter example to demonstrate that the bootstrap is not foolproof, even for statistics $\{T_n\}$ whose asymptotic distribution is normal. He says that asymptotic optimality, or even consistency of the bootstrap estimate H_b can only be expected if the sampling distribution $H(x)$ depends smoothly on F .

Bickel and Freedman (1981) give two counter examples for which the bootstrap does not work. They devise the following rule which they term "rough". The bootstrap will work provided that

- a) $T_n(X_1, \dots, X_n; G)$ tends weakly to a limit law L_G whenever X_1, \dots, X_n are i.i.d. with distribution G , for all G in a "neighbour" of F into which F_n falls with probability 1,
- b) the convergence in (a) is uniform, and
- c) the function $G \rightarrow L_G$ is continuous.

CHAPTER 5

BOOTSTRAP APPROACH TO BIOEQUIVALENCE

5.1 Introduction

Given the data AUC, TMAX and CMAX from a bioavailability study this chapter considers what the bootstrap has to offer in the problem of assessing bioequivalence.

The bootstrap may be used in two different ways to assess bioequivalence:

- (i) by computing confidence intervals, and/or
- (ii) by estimating the relative frequency that the bioequivalence specifications would be met if the bioequivalence trial were repeated indefinitely. We shall call this relative frequency the Index of Concordance.

The bootstrap offers a new look at bioequivalence assessment : it releases the definition of bioequivalence from its previous dependence on available statistical procedures. Secondly it frees one from simple distributional assumptions and tractable mathematics.

5.2 Assessing bioequivalence using intervals

In this section we shall suppose that bioequivalence is to be assessed on the basis of a single parameter, which we denote by θ . This assumption is made in order to allow a comparison with the procedures described in Chapter 3. Further, we conform with Rocke's (1984) method of declaring two formulations bioequivalent; which may be described as follows:

- 1) define a specification interval, (a,b) say, into which θ should fall in order that the two formulations be considered bioequivalent
- 2) compute a central $(1-2\alpha)100\%$ confidence interval, say (L,U) for θ
- 3) declare bioequivalence, with level $(1-2\alpha)100\%$, if the confidence interval (L,U) is wholly contained in the specification interval (a,b) i.e. $(L,U) \subset (a,b)$.

The parameter θ may be any one of three familiar cases that are commonly discussed in the literature.

Case 1: parameter $\theta_1 = \mu_N - \mu_S =$ difference between population means
 estimator $\hat{\theta}_1 = \bar{Y}_N - \bar{Y}_S =$ difference between sample means

The random variable to be bootstrapped is

$$T_1 = \hat{\theta}_1 - \theta_1 \quad (5.1)$$

Case 2: parameter $\theta_2 = \mu_N/\mu_S =$ ratio of population means
 estimator $\hat{\theta}_2 = \bar{Y}_N/\bar{Y}_S =$ ratio of sample means

The random variable to be bootstrapped is

$$T_2 = \hat{\theta}_2 - \theta_2 \quad (5.2)$$

Case 3: parameter $\theta_3 =$ geometric mean of the population of individual ratios

estimator $\hat{\theta}_3 = \left(\prod_{i=1}^n r_i \right)^{1/n}$
 $=$ geometric mean of observed individual ratios

(Steinijans and Diletti 1985)

The random variable to be bootstrapped is

$$T_3 = \hat{\theta}_3 - \theta_3 \quad (5.3)$$

However, the bootstrap will work for other choices of θ and is not restricted to the three cases given here.

5.3 Assumptions of bootstrap interval

In Chapter 4, four methods of constructing bootstrap intervals are derived. These are:

- | | | |
|--------------------------------|--|-----------------------------------|
| (i) pivotal | $T = \hat{\theta} - \theta \sim H$ | pivotal |
| (ii) generalised pivotal | $T = t(\hat{\theta}, \theta) \sim H$ | pivotal |
| (iii) percentile | $T = g(\hat{\theta}) - g(\theta) \sim H$ | symmetric pivot about 0 |
| (iv) bias corrected percentile | $T = g(\hat{\theta}) - g(\theta) \sim H$ | symmetric location scale pivot |

Each of these methods differ in their assumptions:

- (i) For pivotal and generalised pivotal intervals : the assumption is that the sampling distribution, H , is pivotal i.e. does not depend on the unknown parameter θ . But we do not assume anything about the form of H . However, the exact form of $t(\hat{\theta}, \theta)$ must be known in order to construct the interval.
- (ii) For percentile intervals : we assume that a symmetric pivot exists on some scale, i.e. we assume the existence of a monotonic function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that the sampling distribution of $g(\hat{\theta}) - g(\theta) \sim H$ is symmetric.
- (iii) For bias corrected percentile intervals : we assume that on some scale a symmetric pivot exists that belongs to a location scale family, i.e. $g(\hat{\theta}) - g(\theta) \sim H(x; \mu, \sigma) = H_0\left(\frac{x-\mu}{\sigma}\right)$

The advantage of the percentile and bias corrected percentile intervals is that we do not need to know the form of g at all, but only assume its existence (cf. Chapter 4).

5.4 Choosing an interval

Shenker (1985) has demonstrated that bootstrap intervals should not be applied blindly. Some consideration should be given to the bootstrap assumption (mentioned above). The following criterion has been adopted for selecting a bootstrap interval:

Use the interval that will give the correct answer if the data are normally/lognormally distributed.

The idea is that, should the data be normal/lognormal, then little will be lost by applying the bootstrap instead of the parametric procedures. Should the data not be normal/lognormal then the bootstrap, by virtue of its robustness, should have the advantage.

Although a criterion has been adopted, diagnostic methods for checking on the assumptions is an area that needs further research.

For the three random variables given in (5.1), (5.2) and (5.3) that are appropriate for bioequivalence which bootstrap intervals should be used? Each will be discussed in turn.

$$\begin{aligned} \text{For (5.1) : } T_1 &= \hat{\theta}_1 - \theta_1 \\ &= (\bar{Y}_N - \bar{Y}_S) - (\mu_N - \mu_S) \end{aligned}$$

If the data are normal then T_1 will also be normal and will not depend on θ_1 . Hence the distribution of T_1 is a symmetric pivotal location-scale distribution. It seems likely that any bootstrap interval would work well for this random variable. If the data are not normal and come from a skewed distribution then the bias-corrected interval is recommended.

$$\begin{aligned} \text{For (5.2) : } T_2 &= \hat{\theta}_2 - \theta_2 \\ &= \bar{Y}_N / \bar{Y}_S - \mu_N / \mu_S \end{aligned}$$

Efron (1984a, 1985) and Efron and Tibshirani (1985) have considered this problem in detail under the heading of Ratio Estimation. Efron has shown that under the assumption of normality the bias corrected percentile interval approximates closely the exact Fieller solution, while the percentile interval that does not correct for bias gives the Creasy (1954) fiducial solution. Therefore the bias corrected interval is recommended.

$$\begin{aligned} \text{For (5.3) : } T_3 &= \hat{\theta}_3 - \theta_3 \\ &= (\prod_{i=1}^n r_i)^{1/n} - \theta_3 \end{aligned}$$

The motivation for this choice of $\hat{\theta}$ is the assumption that the log-transformed data are normally distributed. Hence under the assumption of log normality the quantity

$$\begin{aligned}
 T_4 &= \log \hat{\theta}_3 - \log \theta_3 \\
 &= g(\hat{\theta}_3) - g(\theta_3)
 \end{aligned}$$

follows a normal distribution that does not depend on θ . This is a symmetric location-scale pivotal quantity on the log scale. If the data is not lognormal then to allow for possible bias the bias-corrected interval is recommended.

5.5 Algorithm for computing the bias corrected interval

As before, for the purpose of computing confidence intervals it will be assumed that carry over effects are of no concern. Effectively this means that the data may be viewed either as though it were the result of a simple two sample experiment comparing a standard formulation (S) with a new formulation (N). Alternatively the data may be viewed as though coming from a matched-pairs experiment comparing S with N.

To describe the bootstrap algorithm a simplified notation for the data will be adopted:

y_{Sj} = response of the j th individual to the standard formulation

y_{Nj} = response of the j th individual to the new formulation

$$x_j = (y_{Nj} \ y_{Sj}) \quad j = 1, \dots, n$$

Also, let

$$\bar{y}_N = \sum_{j=1}^n y_{Nj} / n$$

= sample mean for new formulation

and

$$\bar{y}_S = \sum_{j=1}^n y_{Sj} / n$$

= sample mean for standard formulation

and

$$r_j = y_{Nj}/y_{Sj} \quad j = 1, \dots, n$$

= ratio of responses to new and standard formulations for
jth individual.

The algorithm for generating a bias-corrected percentile interval for $\theta = \theta_1$ or θ_2 or θ_3 is described below. Note that the Monte Carlo approximation to the bootstrap distribution is used.

Step 1 The random mechanism generating the data is estimated by the e.d.f.

$$F_n : \text{mass } 1/n \text{ at each observation } x_1, \dots, x_n.$$

Recall that x_j is the vector (y_{Nj}, y_{Sj}) .

Step 2 Obtain a random sample of size n (with replacement) from F_n to give x_1^*, \dots, x_n^* . Call this a bootstrap sample. Compute $\hat{\theta}^*$ for this sample.

Step 3 Repeat Step 2 B times to give the bootstrap values $\hat{\theta}^*(1), \dots, \hat{\theta}^*(B)$.

Step 4 Compute the function $\hat{CDF}(t)$

$$\hat{CDF}(t) = \#(\hat{\theta}^*(i) \leq t)/B$$

= proportion of bootstrap values less than or equal to t .

This can be most easily done by ordering the values $\hat{\theta}^*(1), \dots, \hat{\theta}^*(B)$ from smallest to largest and then computing the "less-than" cumulative percentage frequency.

Step 5 Compute $\widehat{CDF}(\widehat{\theta})$ = proportion of bootstrap values less than the original estimate.

Step 6 Compute $z_0 = \Phi^{-1}(\widehat{CDF}(\widehat{\theta}))$ where Φ is the standard normal cumulative distribution function.

Step 7 Finally compute the quantities

$$L = \widehat{CDF}^{-1}(\Phi(\Phi^{-1}(\alpha) + 2z_0))$$

and

$$U = \widehat{CDF}^{-1}(\Phi(\Phi^{-1}(1-\alpha) + 2z_0))$$

The interval (L,U) is a bias-corrected percentile interval for θ .

Notice that if $\widehat{CDF} = \frac{1}{2}$, i.e. exactly half the bootstrap values $\widehat{\theta}^*$ are less than the original sample estimate $\widehat{\theta}$, then $z_0 = 0$ and the interval endpoints become

$$L = \widehat{CDF}(\alpha) \quad U = \widehat{CDF}(1-\alpha)$$

the α and $1-\alpha$ percentiles of the bootstrap distribution \widehat{CDF} .

If $\widehat{CDF} \neq \frac{1}{2}$ then the term z_0 compensates for the bias of $\widehat{\theta}$ as an estimator of θ .

In Steps 6 and 7 of the bootstrap algorithm it is necessary to evaluate both Φ and Φ^{-1} , the cumulative standard normal and its inverse. These cannot be computed in closed form but good approximations are given below.

Given z compute $\Phi(z)$ $-\infty < z < \infty$

Step 1 : Compute $x = 1/(1+a|z|)$ $a = 0,2316419$

Step 2 : Compute $R(z) = \phi(z)(b_1x+b_2x^2+b_3x^3+b_4x^4+b_5x^5)$

where $\phi(z) = (2\pi)^{-\frac{1}{2}}e^{-z^2/2}$
 = standard normal density

$$b_1 = 0,319381530$$

$$b_2 = -0,356563782$$

$$b_3 = 1,781477937$$

$$b_4 = -1,821255978$$

$$b_5 = 1,330274429$$

Step 3 : $\Phi(z) = \begin{cases} R(z) & \text{if } z < 0 \\ 1-R(z) & \text{if } z > 0 \end{cases}$

Given α compute $\Phi^{-1}(\alpha)$ $0 < \alpha < 1$

Step 1 : Compute $t = \begin{cases} (-2 \ln (1-\alpha))^{\frac{1}{2}} & \text{if } 0,5 < \alpha < 1 \\ (-2 \ln \alpha)^{\frac{1}{2}} & \text{if } 0 < \alpha \leq 0,5 \end{cases}$

Step 2 : Compute $z = t - \frac{c_0+c_1t+c_2t^2}{1+d_1t+d_2t^2+d_3t^3}$

where $c_0 = 2,515517$ $d_1 = 1,432788$

$$c_1 = 0,802853$$
 $d_2 = 0,189269$

$$c_2 = 0,010328$$
 $d_3 = 0,001308$

Step 3 : $\Phi^{-1}(\alpha) = \begin{cases} -z & \text{if } 0 < \alpha \leq 0,5 \\ z & \text{if } 0,5 < \alpha < 1 \end{cases}$

Reference: Abramowitz and Stegun (1970) Handbook of Mathematical Functions p.932-933.

5.6 Example of bootstrap intervals

To demonstrate the bootstrap intervals we take an example from Steinijans and Diletti (1983). The objective of the experiment from which the data are taken was to investigate the influence of food intake on the bio-availability of theophylline from a sustained-release aminophylline preparation. The data are given in Table 5.1.

Table 5.1 : Area under the concentration/time curve, AUC, after administration of 385,6 mg theophylline in a sustained release preparation under reference condition (fasted) and test condition (standard breakfast)

| Subject | AUC(mg/l h) | | |
|----------------|-------------|-------|-------|
| | Reference | Test | Ratio |
| 1 | 136,0 | 135,7 | 1,00 |
| 2 | 152,6 | 155,3 | 1,02 |
| 3 | 123,1 | 148,9 | 1,21 |
| 4 | 77,0 | 81,2 | 1,05 |
| 5 | 115,7 | 139,2 | 1,20 |
| 6 | 72,0 | 91,7 | 1,27 |
| 7 | 116,4 | 118,7 | 1,02 |
| 8 | 151,1 | 133,2 | 0,88 |
| 9 | 118,9 | 115,6 | 0,97 |
| 10 | 156,1 | 150,3 | 0,96 |
| 11 | 222,4 | 223,9 | 1,01 |
| 12 | 158,1 | 154,1 | 0,97 |
| Geometric mean | 127,7 | 133,1 | 1,04 |

For the data in Table 5.1 Steinijans and Diletti (1983) derive confidence intervals for what they call the "bioavailability ratio" using the eight different procedures listed in Table 5.2. Table 5.2 is a reproduction of Table 6 in Steinijans and Diletti (1983), except that the bootstrap bias-corrected intervals have been added. The bootstrap limits have been computed by choosing $B = 1000$ in Step 3 of the bootstrap algorithm.

Notice that there are two ways of computing a "bioavailability ratio":

- (i) ratio of two arithmetic means. This computation would be used if the bioavailability ratio referred to the ratio of two expectations i.e. $\theta = \mu_N/\mu_S$. (Used for the first three intervals in Table 5.2.)
- (ii) geometric mean of individual ratios = ratio of geometric means. Steinijans and Diletti (1983). This estimator makes use of the matched pairs nature of the data.
(Used for the last four intervals in Table 5.2.)

These two statistics do not estimate the same theoretical quantity. The choice will of course depend on the definition of bioequivalence.

Steinijans and Diletti (1983, 1985) have argued convincingly in favour of distribution free methods for computing confidence intervals for bioavailability parameters. Especially so in the case of skewed or bimodal distributions. They recommend Tukey's procedure based on the Wilcoxon signed rank test where the assumption is that each error term* comes from a continuous distribution (not necessarily the same one) symmetrical about zero.

The bootstrap does not require the assumption of symmetry but only that a transformation to symmetry exists (cf. Chapter 4) and has the further

*The model for the Wilcoxon signed rank test is given by $d_j = \delta + \epsilon_j$. See equation (3.11).

Table 5.2 Point estimate and 95% confidence limits of bioavailability ratio for data given in Table 5.1

| | Statistical method | Point estimate | 95% confidence limits | Exact level of confidence |
|--|---|----------------|-----------------------|---------------------------|
| Normal distribution | Paired t-test | 1,03 | 0,97 ; 1,09 | |
| | ANOVA | 1,03 | 0,97 ; 1,10 | |
| | Westlake | | 0,92 ; 1,08 | > 0,95 |
| Lognormal distribution | Paired t-test | 1,04 | 0,97 ; 1,12 | |
| | ANOVA | 1,04 | 0,97 ; 1,12 | |
| | Westlake | | 0,89 ; 1,11 | > 0,95 |
| Distribution-free (nonparametric) ratio analysis | Signed rank test (Tukey) | 1,02 | 0,97 ; 1,11 | 0,9575 |
| | Pitman's permutation test | 1,04 | 0,97 ; 1,12 | 0,9502 |
| Distribution free | Bootstrap (Ratio of individual readings) | 1,04 | 0,98 ; 1,10 | |
| Distribution free | Bootstrap (Ratio of arithmetic means) | 1,03 | 0,98 ; 1,09 | |

advantage that it mimics reality. It does this by treating the sample as though it were the population and then repeatedly drawing random samples from this population. For each sample the statistic is computed. The cumulative distribution $\hat{C}DF(t)$ referred to in Step 4 of the bootstrap algorithm is built up from these values. Figure 5.1 gives a comparison of the cumulative distributions $\hat{C}DF(t)$ (bootstrap) and $G(t)$ (Tukey) for the data in Table 5.1.

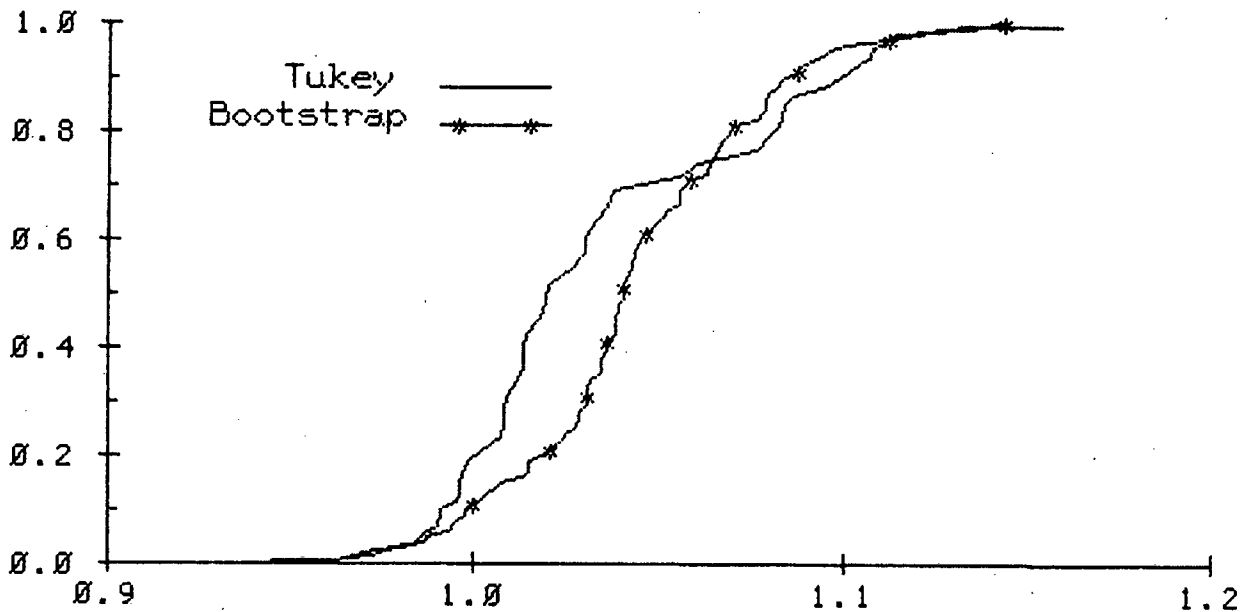


Figure 5.1 : Cumulative probability distributions for geometric mean of individual ratios for Tukey and bootstrap methods
The bootstrap curve is a Monte Carlo approximation using
 $B = 100$

5.7 Wilcoxon and Pitman procedures as resampling plans

Efron (1982) defined a resampling procedure as a generic term for all methods which evaluate $\hat{\theta}$ at reweighted versions of the e.d.f. F_n . He described a resampling procedure as follows:

"For simplicity consider a functional statistic $\hat{\theta} = \theta(F_n)$. The data x_1, \dots, x_n are thought of as observed and fixed in what follows. A resampling vector

$$\underline{P}^* = (P_1^*, P_2^*, \dots, P_n^*)$$

is any vector on the n -dimensional simplex

$$S_n = \{\underline{P}^* : P_\lambda^* \geq 0, \sum P_\lambda^* = 1\}$$

in other words, any probability vector. Corresponding to each \underline{P}^* is a reweighted e.d.f. $F_n^* = F_n(\underline{P}^*)$

$$F_n^* : \text{mass } P_\lambda^* \text{ on } x_\lambda, \quad \lambda = 1, \dots, n,$$

and a resampled value of $\hat{\theta}$, say $\hat{\theta}^*$,

$$\hat{\theta}^* = \theta(F_n(\underline{P}^*)) = \hat{\theta}(\underline{P}^*)$$

Some of the resampling vectors play special roles in the bootstrap and jackknife theories. In particular,

$$P^0 = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

corresponds to F_n itself i.e. $F_n(P^0) = F_n$, and the observed value of the statistic $\hat{\theta} = \hat{\theta}(P^0)$. The jackknife considers the vectors

$$P_{(i)} = \left(\frac{1}{n-1}, \frac{1}{n-1}, \dots, 0, \dots, \frac{1}{n-1}\right) \quad (0 \text{ in } i\text{th place})$$

with corresponding values $\hat{\theta}_{(i)}$ of the statistic $i = 1, \dots, n$. The bootstrap considers all P^* vectors of the form M^*/n , M^* having non-negative integer coordinates adding to n ."

The nonparametric Wilcoxon procedure considers all Walsh averages of the differences $d_i = y_{Ni} - y_{Si}$, $i = 1, \dots, n$. The Walsh averages a_{ij} are defined by

$$a_{ij} = (d_i + d_j)/2$$

This corresponds to a resampling vector with $\frac{1}{2}$ in the i th and j th places

$$P_{ij} = (0, \dots, \frac{1}{2}, 0, \dots, \frac{1}{2}, 0, \dots, 0) \quad i \neq j$$

and a resampling vector with 1 in the i th place

$$P_{ii} = (0, \dots, 1, \dots, 0) \quad i = j$$

It is clear that the Wilcoxon procedure may be viewed as a resampling plan.

It should further be noted that if n is even then every Walsh average is also a bootstrap point. This is evident from the identity

$$\begin{aligned} a_{ij} &= (d_i + d_j)/2 \\ &= \frac{\text{n/2 terms}}{(d_i + d_i + \dots + d_i) + \frac{\text{n/2 terms}}{(d_j + d_j + \dots + d_j)}} n \end{aligned}$$

This corresponds to a bootstrap resampling vector with M^* having $n/2$ in the i th and j th places

$$M^* = (0, \dots, n/2, 0, \dots, n/2, 0, \dots, 0)$$

Pitman's procedure may also be viewed as a resampling plan. Referring to the notation in Chapter 3, consider any one of the 2^{n-1} Pitman averages. Suppose that it is the point

$$a = \sum_{i_m} d_{i_m} / M$$

where the index set is $\{i_1, \dots, i_M\} \subset \{1, \dots, n\}$.

The resampling vector corresponding to this point a has 1's in positions i_1, \dots, i_M and zero elsewhere, rescaled by M^{-1} , i.e.

$$\underline{p}^* = M^{-1}(0, \dots, 1, \dots, 1, \dots, 1, \dots, 1, \dots, 0)$$

A final point worth noting is the comparison of the number of point or probability atoms in each of the nonparametric procedures. This is given in Table 5.3.

Table 5.3 : Comparison of the number of points in the support of the Wilcoxon, Pitman's and bootstrap distributions

| Method | General n | n = 12 |
|-----------|-------------------|-----------|
| Wilcoxon | $n(n+1)/2$ | 78 |
| Pitman's | 2^{n-1} | 4 095 |
| Bootstrap | $\binom{2n-1}{n}$ | 1 352 078 |

This comparison between bootstrap distribution and certain other nonparametric distributions needs further development. Efron (1981a, 1982a) has compared a number of nonparametric procedures in some detail.

5.8 Index of Concordance

Essentially bioequivalence is a multivariate problem, yet in practice a univariate procedure is applied to each bioavailability parameter individually. A possible reason for this is that most useful multivariate procedures rest heavily on the assumption of joint normality, which hardly seems feasible for the bioavailability parameters.

It is possible to obtain a bivariate or even multivariate bootstrap distribution and this might offer hope of deriving multivariate bootstrap confidence regions. However it is not yet clear how this should be done.

To overcome this difficulty we propose a new measure of bioequivalence which we shall call the Index of Concordance. The Index of Concordance I_n for the population is the relative frequency with which the bioequivalence specification is met if applied to a group of n individuals at random.

In principle, to compute the index of concordance one would need to take every possible sample of size n from the population, perform the clinical trial on each group, and then compute I_n

$$I_n = \text{Index of Concordance} \\ = \frac{\text{number of times specifications were met}}{\text{number of clinical trials performed}}$$

Clearly I_n is a number lying between 0 and 1. Values of I_n lying close to 1 indicate that the bioequivalence specifications are met in most cases while values of I_n lying close to 0 indicate that the bioequivalence specifications are seldom met.

The bootstrap provides a very simple method for estimating I_n . The bootstrap uses the sample values as though they were the population. The bootstrap estimate of I_n is obtained by taking random samples of size n , over and over again, say B times from a population with distribution F_n , the e.d.f. (A random sample of size n is obtained by sampling with replacement from the n original data values.) The bootstrap estimate of the index of concordance is then

$$\hat{I}_n = \text{number of times specifications were met}/B$$

The index of concordance has the advantage of being extremely easy to compute for any number of parameters jointly, and the problem of inter-

pretation does not arise. It has the further advantage of being easy to understand by those with a limited statistical background. The concepts underlying hypothesis testing, confidence intervals and posterior probabilities are not easily communicated to non-statisticians.

5.9 Example

To illustrate the index of concordance we use the data from Fluehler et al (1983). The data relates to a comparison of a slow release formulation (New) against a standard formulation with the aim of producing markedly lower peak concentrations. The bioequivalence specification region is

$$\text{AUC} : 0,8 < \theta_1 < 1,2$$

and

$$\text{C}_{\text{MAX}} : \theta_2 < 0,6$$

For AUC they use the ratio of means as a statistic while for C_{MAX} they use the geometric mean of individual ratios. This is because they assume AUC to be normally distributed and C_{MAX} to be lognormally distributed. For the sake of comparison we shall use the same statistics. The data are given in Table 5.4.

For this data the estimated joint index of concordance, based on a bootstrap with $B = 1000$, is

$$\hat{I}_n = \hat{I}_n[0,8 < \theta_1 < 1,2; \theta_2 < 0,6] = 0,8480$$

Table 5.4 : Comparative bioavailability data from Fluehler *et al* (1983). Comparison of a slow release formulation (New) against a standard

| Number of Subjects | Standard Formulation | | | | New Formulation | | | |
|--------------------|----------------------|--------|------------------|---------------------------|-----------------|--------|------------------|---------------------------|
| | Period | AUC | C _{MAX} | t_n (C _{MAX}) | Period | AUC | C _{MAX} | t_n (C _{MAX}) |
| 1 | 2 | 114,57 | 296,11 | 5,6907 | 1 | 115,21 | 67,97 | 4,2190 |
| 2 | 1 | 98,17 | 146,69 | 4,9883 | 2 | 106,60 | 92,63 | 4,5286 |
| 3 | 1 | 121,87 | 259,37 | 5,5583 | 2 | 129,70 | 97,75 | 4,5824 |
| 4 | 2 | 30,20 | 197,36 | 5,2850 | 1 | 52,85 | 196,53 | 5,2808 |
| 5 | 2 | 131,51 | 281,37 | 5,6397 | 1 | 59,42 | 59,71 | 4,0895 |
| 6 | 1 | 104,17 | 179,14 | 5,1882 | 2 | 152,76 | 54,99 | 4,0072 |
| 7 | 1 | 71,54 | 251,37 | 5,5269 | 2 | 31,24 | 93,11 | 4,5337 |
| 8 | 2 | 71,98 | 233,29 | 5,4523 | 1 | 108,22 | 109,26 | 4,6938 |
| 9 | 2 | 78,83 | 173,61 | 5,1568 | 1 | 82,05 | 152,18 | 5,0251 |
| 10 | 1 | 140,48 | 227,56 | 5,4274 | 2 | 101,10 | 177,09 | 5,1767 |
| 11 | 2 | 75,27 | 211,85 | 5,3559 | 1 | 58,72 | 100,70 | 4,6121 |
| 12 | 1 | 111,56 | 225,71 | 5,4192 | 2 | 83,27 | 172,22 | 5,1488 |
| Mean | - | 98,35 | 223,62 | 5,3907 | - | 90,10 | 114,51 | 4,6581 |

Fluehler et al (1983) compute the marginal posterior probabilities but do not compute the joint posterior probability. As a comparison the marginal indices of concordance have been computed. The results are summarised in Table 5.5.

Table 5.5 : Comparison of marginal posterior probabilities and marginal indices of concordance for the Fluehler et al (1983) data

| Event | Marginal Posterior Probability | Marginal Index of Concordance |
|------------------------|--------------------------------|-------------------------------|
| $0,8 < \theta_1 < 1,2$ | 0,846 | 0,898 |
| $\theta_2 < 0,6$ | 0,906 | 0,947 |

Notice that since \hat{I}_n is simply a proportion the variance of \hat{I}_n is approximately

$$\text{Var}(\hat{I}_n) \simeq \hat{I}_n(1-\hat{I}_n)/B$$

5.10 Conclusions

The bootstrap has much to offer in dealing with the problem of bio-equivalence. Some of the advantages are:

- (i) It frees the clinician to define bioequivalence in a manner that will reflect clinical requirements and not depend on the available statistical procedures or on his knowledge of statistics.
- (ii) The bootstrap can be easily described to the statistical layman.

- (iii) For univariate bioequivalence assessment bootstrap confidence intervals can be obtained for almost any definition of a bioequivalence parameter.
- (iv) For univariate or multivariate bioequivalence assessment, the index of concordance can be computed with equal ease.
- (v) The basic bootstrap makes no assumptions about the functional form of the random mechanism F generating the data. However, should one want to make assumptions about F these can be easily accommodated by using for example, a parametric bootstrap. See Efron (1982). In this sense the bootstrap encompasses parametric methods.

CHAPTER 6

BOOTSTRAPPING THE TOTAL RESPONSE TO A
BIOAVAILABILITY STUDY6.1 Introduction

The bootstrap is a procedure for estimating the sampling distribution, or some functional thereof, of a specified random variable. This is achieved by resampling the data in a suitable way.

This idea of resampling the data may be applied to compartmental modeling of bioavailability studies. The technical difficulties associated with compartmental models include

- (i) appropriateness of model
- (ii) correlation between errors
- (iii) unequal variances at different time points
- (iv) possible instability of coefficient estimates.

If a compartmental model has been fitted to the data by some statistical procedure, then in addition to the fitted values, there is a set of residuals. Assumptions have been placed on the residuals explicitly or implicitly by the fitting procedure. If the residuals are resampled, preserving this stochastic structure, a distribution is generated using the model's own assumptions.

Assuming the model and estimated parameters to be correct, resampling the residuals produces pseudo-data which mimics a new set of experiments.

Using the pseudo-data it is possible to focus on any aspect of the experiment that may be of particular interest. For example the pseudo-data could be used to "predict" the possible behaviour of a given subject on a number of future occasions. Alternatively it may be used to "observe" the joint distribution of the derived parameters such as AUC, TMAX and CMAX, either within subjects or between subjects.

This pseudo-data may be a useful guide to planning future investigations or even deciding if further investigation is necessary.

Although similar results could be obtained using the model equation and attaching an error generated from a $N(0, \sigma^2)$ distribution the advantage in using the experimental observed residuals is that they contain the stochastic variability that is inherent in the data.

We shall illustrate some of these possibilities with a simple example.

6.2 Data, model and assumptions

For the data in Table 2.1 (Button (1979)) we shall suppose that the appropriate model is the sum of two exponentials. Denoting the observation at time t_j on the i th individual by y_{ij} $i = 1, \dots, 6$ and $j = 1, \dots, 18$ the model is

$$y_{ij} = A_i \left(e^{-b_{1i} t_j} - e^{-b_{2i} t_j} \right) + \epsilon_{ij} \quad \begin{array}{l} i = 1, \dots, 6 \\ j = 1, \dots, 18 \end{array} \quad (6.1)$$

The errors ϵ_{ij} are assumed to be independent. However the usual assumption of constant variance does not seem appropriate.

The parameters (A_i, b_{1i}, b_{2i}) $i = 1, \dots, 6$ will be estimated using ordinary least squares. This choice of using ordinary least squares in favour of weighted least squares is motivated largely by the lack of an obvious choice of weighting factors. A discussion of the choice of weighting factors is given in Chapter 2.

The bioavailability parameters AUC, TMAX and CMAX are nonlinear functions of the model parameters (A, b_1, b_2) (cf. Chapter 1). For $i = 1, \dots, 6$ we have

$$\text{AUC}_i = A_i(1/b_{1i} - 1/b_{2i})$$

$$\text{TMAX}_i = (\ln b_{2i} - \ln b_{1i})/b_{2i} - b_{1i})$$

$$\begin{aligned} \text{CMAX}_i &= A_i \left(e^{-b_{1i} \text{TMAX}_i} - e^{-b_{2i} \text{TMAX}_i} \right) \\ &= A_i (b_{2i}/b_{1i})^{b_{1i}/(b_{1i}-b_{2i})} - (b_{2i}/b_{1i})^{b_{2i}/(b_{1i}-b_{2i})} \end{aligned}$$

It is well known that the maximum likelihood estimator for the vector (A, b_1, b_2) is asymptotically multivariate normal with variance-covariance matrix $\sigma^2(F'F)^{-1}$, where F is the design matrix. For details and notation see Box and Lucas (1959). However we do not know how closely the sample behaviour is to the asymptotic behaviour and we know even less about the distribution of the derived parameters AUC, TMAX and CMAX. It therefore seems desirable to adopt a nonparametric approach. The bootstrap procedure was chosen because, according to Efron (1981), of all nonparametric procedures, "*the bootstrap performs notably best.*"

In addition to estimating the variance-covariance structure from a practical viewpoint, an 'estimate' or display of the distribution of possible values

of AUC, TMAX and CMAX would be of great interest. After all, the drug is to be given to an individual and it would be of interest to know where he/she might lie on the distribution. In fact the quantiles of these distributions are probably of more interest than the mean and variance. However, bioavailability parameters are usually estimated from experiments based on only a few subjects, typically five to ten. Is there a way in which we can obtain some idea of the possible variability? It seems that the bootstrap can help here in the following ways:

- (i) Give insight into the way in which the pharmacological parameters behave in a population.
- (ii) Indicate the extreme instances.
- (iii) Give a graphical indication of the distribution of these parameters. This could not be done with only six original observations.

6.3 The bootstrap algorithm

The bootstrap algorithm devised for this problem is as follows:

1. Fit the model given by equation (6.1) to each of the six data sets given in Table 2.1.
2. For each of the data sets store the vector of fitted values as well as the vector of residuals.
3. Generate a 'bootstrap curve' consisting of 18 points. (The method for doing this is described below.)
4. Fit model (6.1) to the bootstrap curve.
5. Repeat steps 3 and 4 a total of 250 times.

The details of the bootstrap are:

(a) Fit all curves using ordinary least squares.

(b) The method of generating a bootstrap curve is as follows:

Select at random a vector of fitted values from the six available.

This vector consists of 18 points and to each of these we must add

an error term. Starting with the first point we choose at random

one of the six 'first' residuals. With the second point we choose

one of the six 'second' residuals. We repeat this for all 18 points.

Call this set the bootstrap parameters.

This algorithm is based on the following assumptions.

(i) model (6.1) is correct

(ii) errors are independent

(iii) unequal variance at different sampling times.

The total bootstrap distribution consists of 6×6^{18} 'bootstrap curves' or pseudo-data. We have used the Monte Carlo method to approximate this distribution (cf. Chapter 4).

6.4 Results

6.4.1 Original data

Model (6.1) was fitted to each of the six horses in Table 2.1 to give

$(A_i, b_{1i}, b_{2i}) \quad i = 1, \dots, 6$. These estimates were then used to compute

$(AUC_i, TMAX_i, CMAX_i) \quad i = 1, \dots, 6$. The results are given in Table 6.1.

The fit was found to be reasonable but not very good. There was slight evidence of patterned residuals, the pattern being to overestimate the first few points, underestimate the peak and overestimate the tail points.

Table 6.1 Estimated parameters for the data in Table 2.1

| | HORSE | | | | | | Mean | SD |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | | |
| A | 22,54 | 22,54 | 21,78 | 21,18 | 21,57 | 20,00 | 21,60 | 0,953 |
| b_1 | 0,0686 | 0,0600 | 0,0588 | 0,0606 | 0,0562 | 0,0592 | 0,0606 | 0,0042 |
| b_2 | 1,242 | 2,905 | 4,295 | 3,087 | 5,931 | 5,646 | 3,848 | 1,786 |
| AUC | 310 | 368 | 365 | 343 | 380 | 334 | 351,5 | 25,03 |
| TMAX | 2,47 | 1,36 | 1,01 | 1,30 | 0,79 | 0,82 | 1,2917 | 0,624 |
| CMAX | 18,0 | 20,3 | 20,2 | 19,2 | 20,4 | 18,9 | 19,50 | 0,963 |

Marginal histograms of AUC, TMAX and CMAX are given in Figure 6.1.

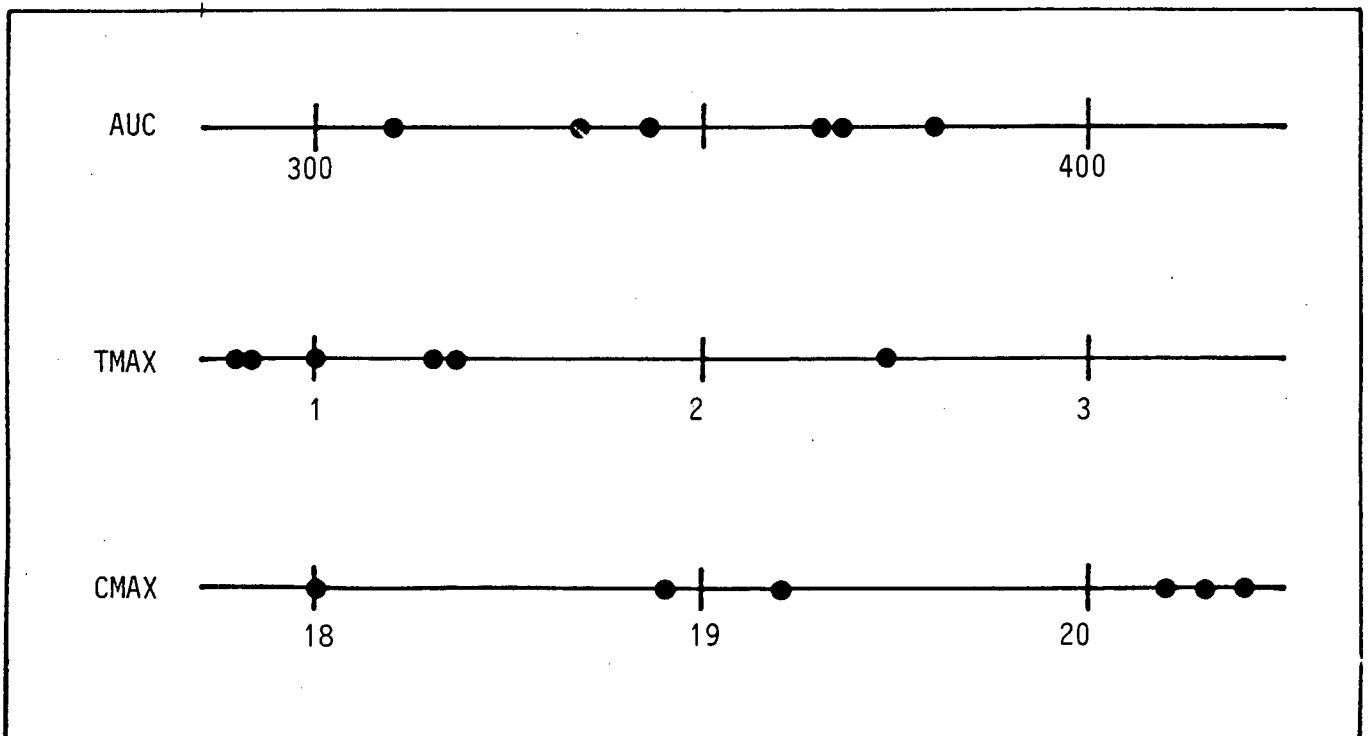


Figure 6.1 Marginal histograms for AUC, TMAX and CMAX based on the data in Table 6.1.

6.4.2 Bootstrap distribution of (A, b_1, b_2) and $(AUC, TMAX, CMAX)$

Model (6.1) was fitted to each of the 250 'bootstrap curves' to give $(A_j^*, b_{1j}^*, b_{2j}^*)$ $j = 1, \dots, 250$. These estimates were then used to compute $(AUC_j^*, TMAX_j^*, CMAX_j^*)$ $j = 1, \dots, 250$. As with the original sample, the fit was found to be reasonable but not very good. There was slight evidence of patterned residuals, the pattern being similar to that in the original data. The serial correlations for the residuals tended to be smaller for the bootstrap curves than for the original data.

The distribution putting mass $1/250$ at each vector will be called the bootstrap distribution. Although, of course, it is actually the Monte Carlo approximation to the bootstrap distribution.

The marginal bootstrap distributions of AUC, TMAX and CMAX proved to be most interesting. Especially those of TMAX and CMAX, both of which are bi-modal.

The bootstrap identified two groups for TMAX: A large group of 'normal' responders (about 82%) and a small group of 'slow' responders (about 18%). Perhaps this is to be expected because the first horse in Table 2.1 was a slow responder. Similar comments apply to the marginal bootstrap distribution of CMAX. In contrast, the marginal bootstrap distribution of AUC was unimodal and even close to normality.

However, what is most significant, and surprising, is that no such grouping was evident in the marginal bootstrap distributions of A, b_1 and b_2 . These distributions were unimodal.

The remainder of this section consists of tables and figures that compare and summarise the results for the original sample, the bootstrap and some asymptotic results.

Table 6.2 gives a comparison of the means and standard deviations for the original sample, the bootstrap and, where possible, asymptotic results.

Table 6.2 : Comparison of means and standard deviations for the original sample and the bootstrap

| | | Original sample n = 6 | Bootstrap n = 250 | Asymptotic |
|----------------|------|--------------------------|----------------------|------------|
| A | Mean | 21,60 | 21,67 | 21,42 |
| | SD | 0,95 | 0,97 | 0,62 |
| b ₁ | Mean | 0,0606 | 0,0600 | 0,0588 |
| | SD | 0,0042 | 0,0040 | 0,0048 |
| b ₂ | Mean | 3,85 | 3,55 | 3,32 |
| | SD | 1,79 | 1,38 | 0,32 |
| AUC | Mean | 352 | 352 | - |
| | SD | 25 | 22 | - |
| TMAX | Mean | 1,29 | 1,32 | - |
| | SD | 0,62 | 0,47 | - |
| CMAX | Mean | 19,51 | 19,62 | - |
| | SD | 0,96 | 1,18 | - |

Notice how the bootstrap results agree with the original sample whilst the asymptotic standard deviations of A and b₂ are markedly smaller.

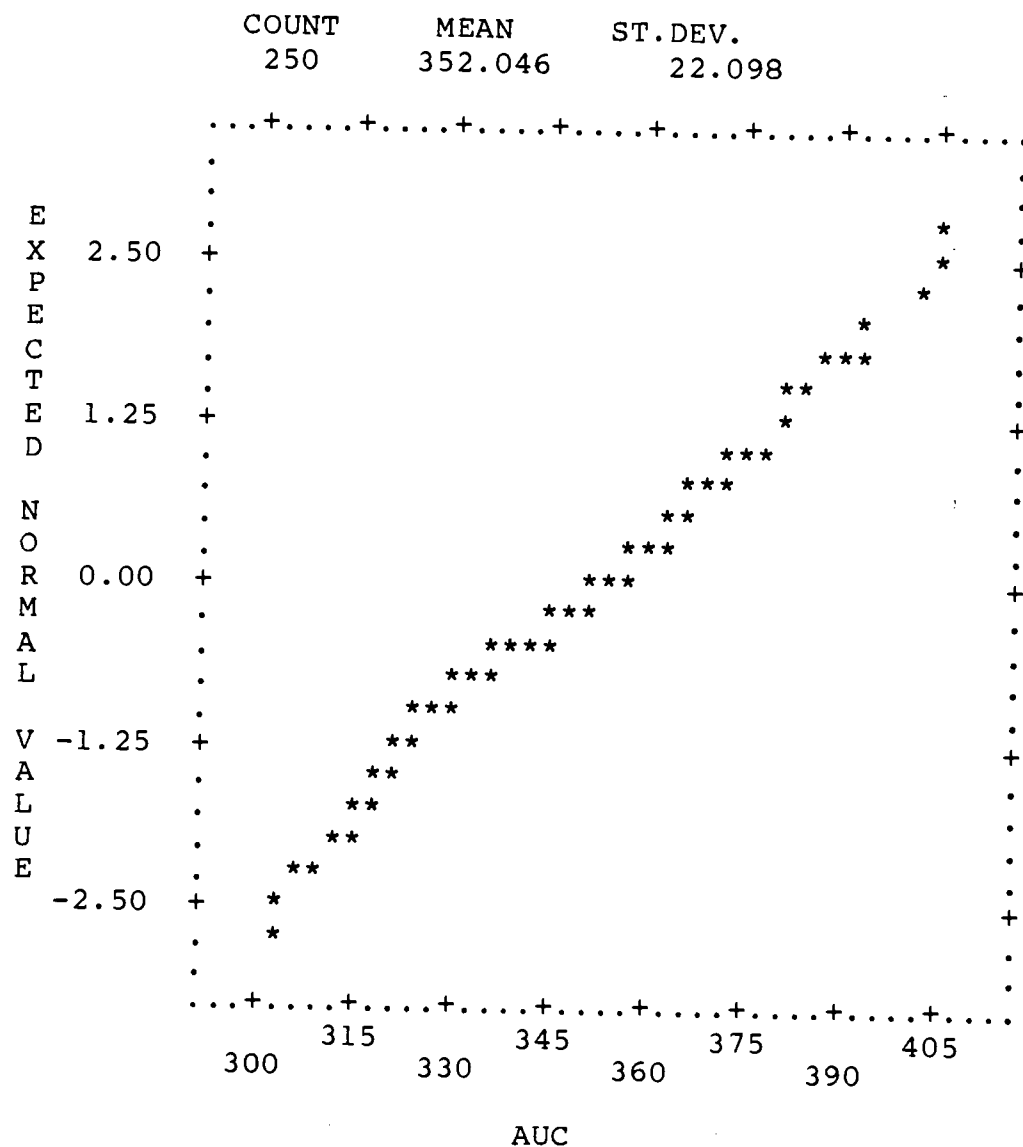


Figure 6.2 Normal probability plot of AUC

Figure 6.2 indicates that the marginal bootstrap distribution of AUC is close to normal with perhaps slightly heavier tails.

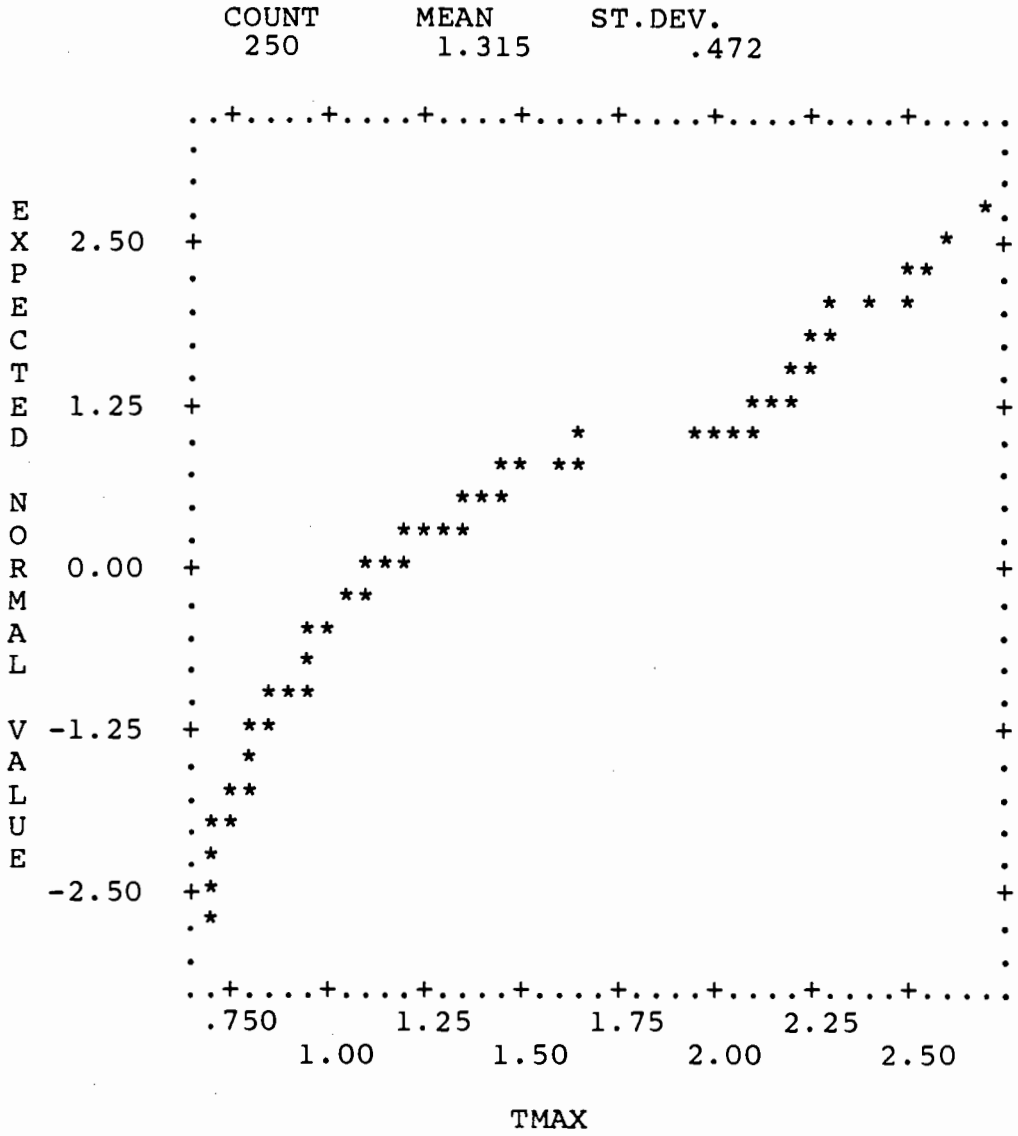
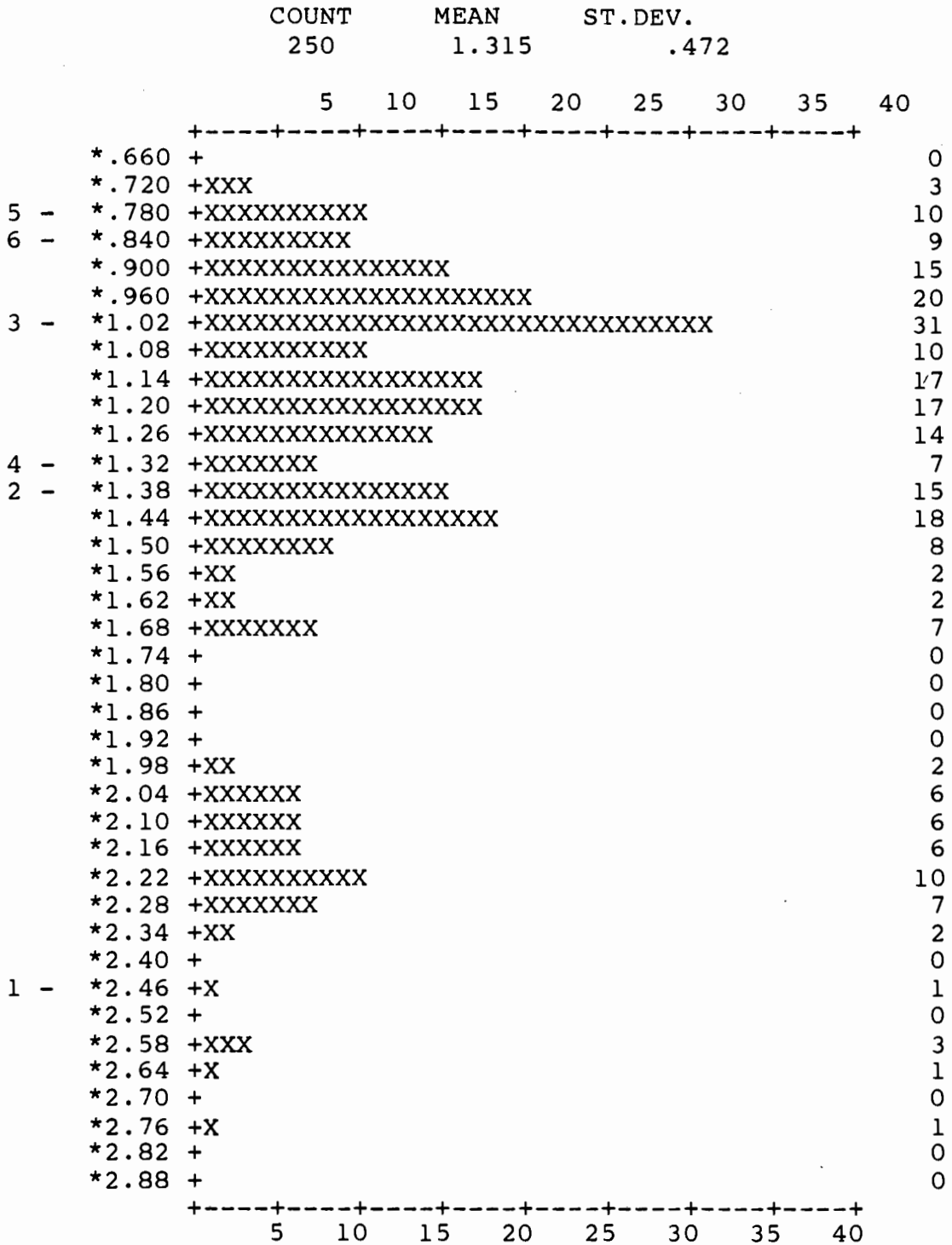


Figure 6.3 Normal probability plot of TMAX

The two groups in this plot correspond to the 'normal' responders in the lower left hand corner and the 'slow' responders in the upper right hand corner. The two groups indicate the bimodal nature of the marginal bootstrap distribution of TMAX which is distinctly non-normal. The histogram in Figure 6.4 displays this clearly.

Figure 6.4 Histogram of TMAX

Bi-modal marginal bootstrap distribution of TMAX. The numbers on the extreme left of the figure indicate where the original sample values lay on the distribution. Notice how the bootstrap has produced even more extreme cases than the original sample. The gap between '1' and '2' noted in the original sample did not close in the bootstrap distribution.

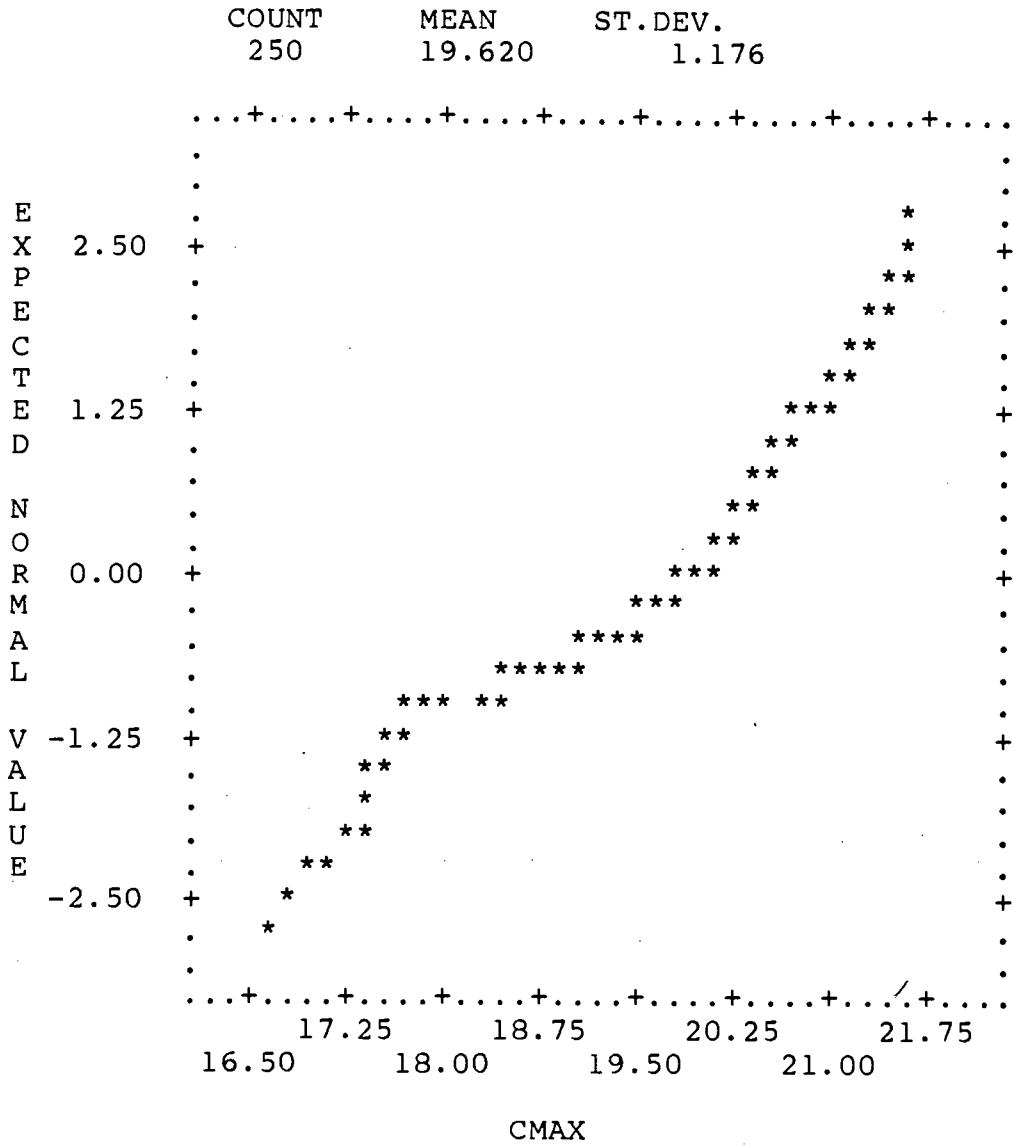


Figure 6.5 Normal probability plot of CMAX

This indicates a bi-modal marginal bootstrap distribution for CMAX. Again the histogram illustrates the bi-modal nature more effectively than the normal probability plot. The histogram is given in Figure 6.6.

6.13

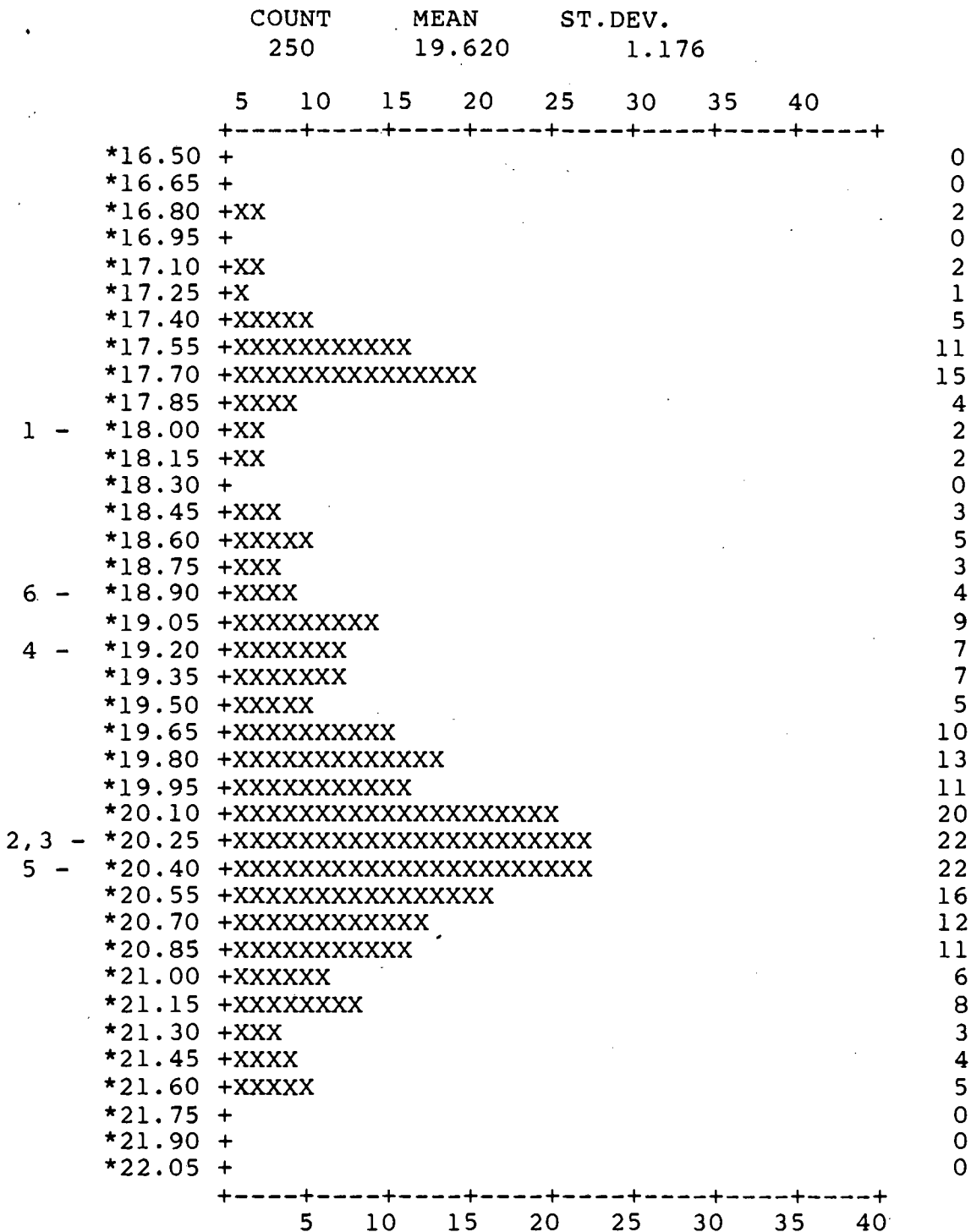


Figure 6.6 Histogram of CMAX

Bi-modal marginal bootstrap distribution of CAMX. The numbers of the extreme left of the figure indicate where the original horses lay on the distribution. Notice again, as for TMAX, the bootstrap has produced more extreme cases than noted in the original sample, and the gap between '1' and '6' noted in the original sample did not close in the bootstrap distribution.

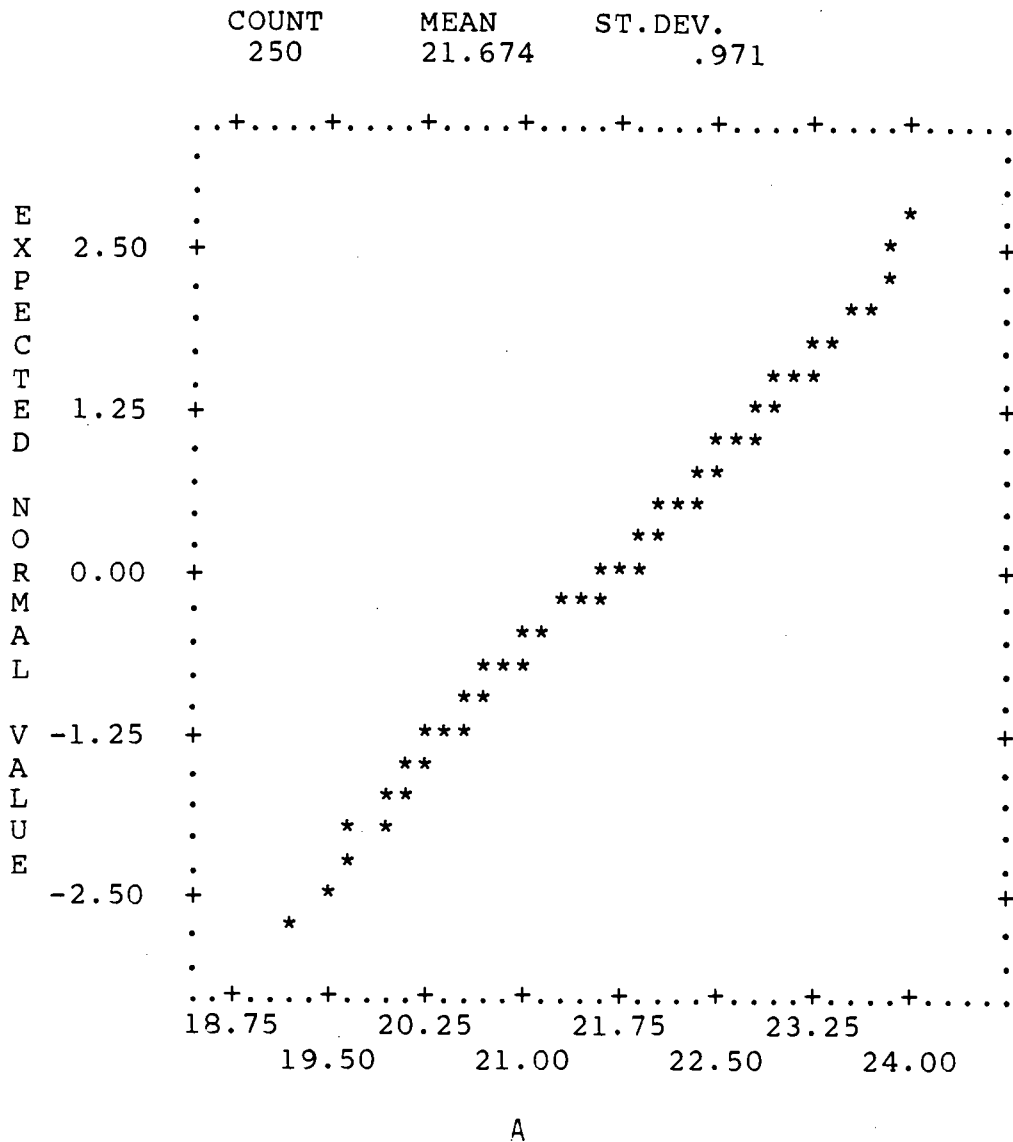


Figure 6.7 Normal probability of plot of A

This plot indicates that the marginal bootstrap distribution of A is close of normality.

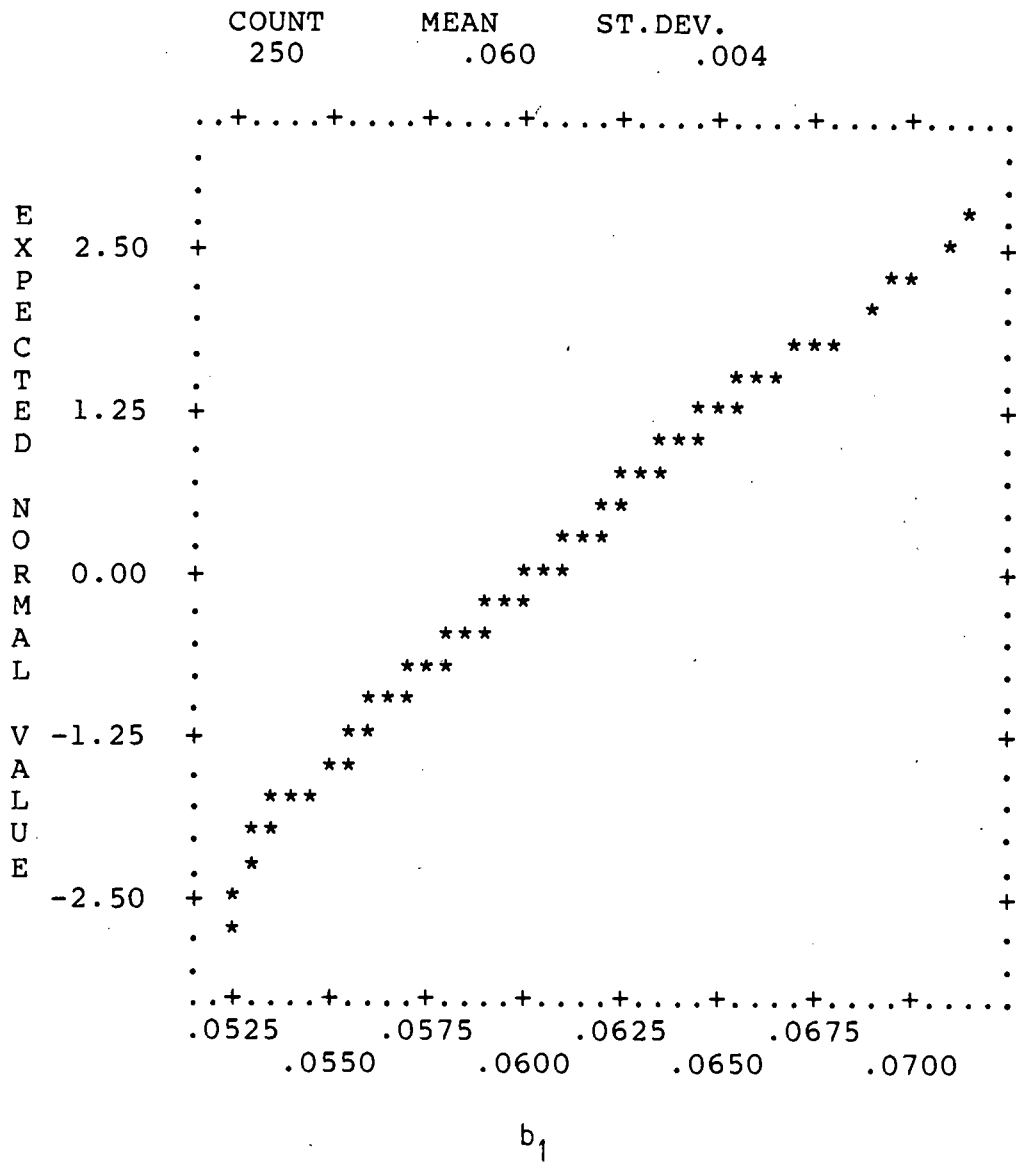


Figure 6.8 Normal probability plot of b_1

This plot indicates that the left hand tail of the marginal bootstrap distribution of b_1 is heavier than found in the normal distribution.

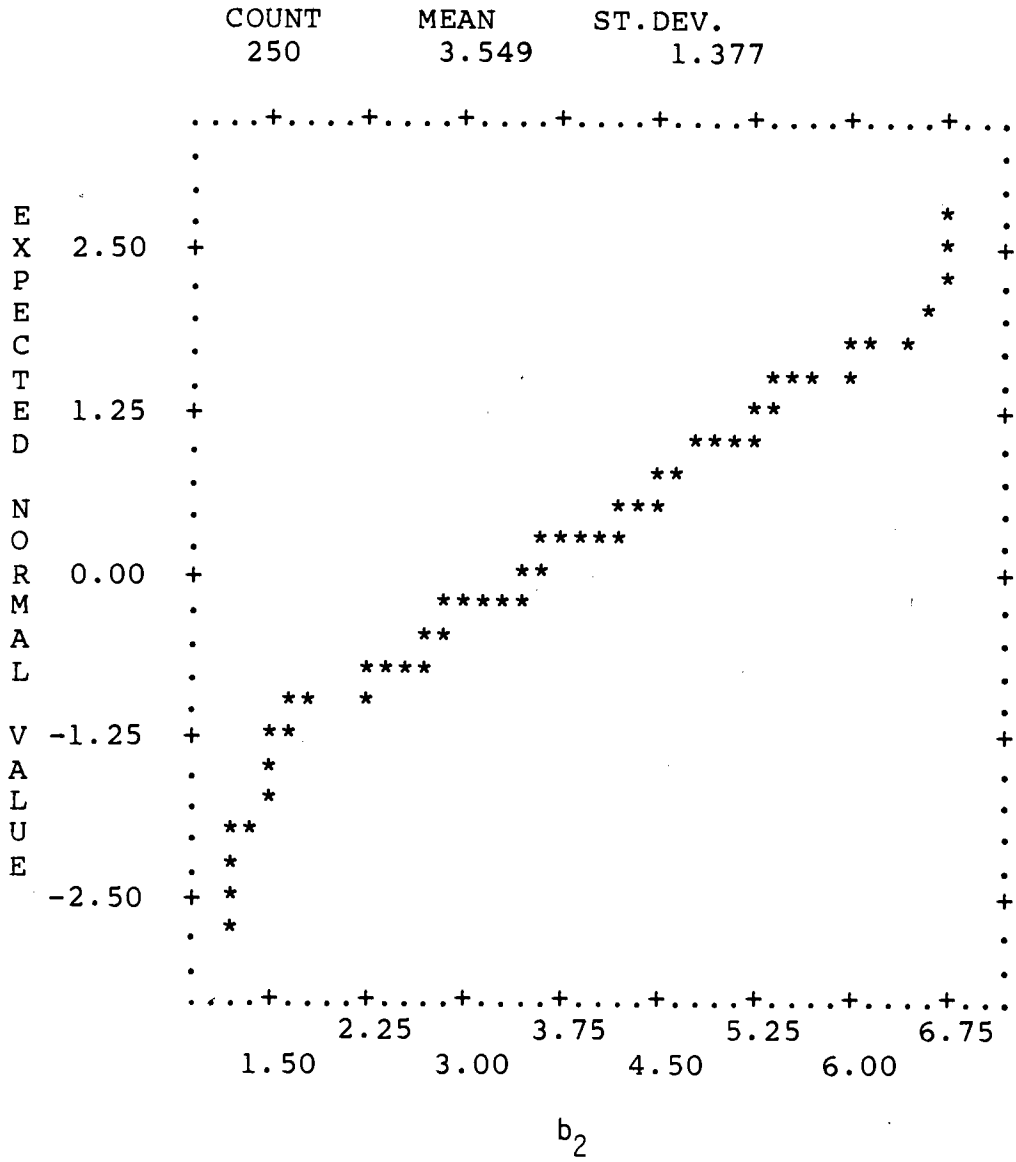


Figure 6.9 Normal probability plot of b_2

This plot indicates that both tails of the marginal bootstrap distribution of b_2 are heavier than those of the normal distribution. However the distribution does appear to be symmetric.

Table 6.3 : Bootstrap correlation matrix

| | A | b_1 | b_2 |
|-------|---|-------|-------|
| A | 1 | 0,40 | -0,07 |
| b_1 | | 1 | -0,13 |
| b_2 | | | 1 |

Table 6.4 : Asymptotic correlation matrix

| | A | b_1 | b_2 |
|-------|---|-------|-------|
| A | 1 | -0,74 | 0,64 |
| b_1 | | 1 | -0,44 |
| b_2 | | | 1 |

The asymptotic correlations are markedly different to the bootstrap correlations. The asymptotic correlations are larger in magnitude than those for the bootstrap. While the correlation between A and b_1 is positive for the bootstrap values (0,4) it is strongly negative (-0,74) in the asymptotic correlation matrix. The correlation between A and b_2 is in the bootstrap (-0,07) but is given as 0,64 in the asymptotic matrix.

6.5 Conclusions

By utilizing the assumptions that are built into the statistical model and the Monte Carlo method, it has been possible to estimate the distribution of the bioavailability parameters AUC, TMAX and CMAX. The

assumption of normality or lognormality has been replaced with more appropriate but more complex assumptions and prodigious calculation. Using this method a detailed graphic picture has been obtained of the way in which the important parameters AUC, TMAX and CMAX might behave in a population. Note particularly how it emphasises the unusual cases. Asymptotic results based on normal theory give smoothed estimates and little indication of sampling variability.

Under special circumstances, (normal distribution of AUC, TMAX and CMAX) all information about their distribution can be derived from a knowledge of the mean and variance. Good classical statistics would allow us to draw inference from samples as small as six. However we have obtained the derived parameters AUC etc., in a complex manner and there is no guarantee that the distribution of these quantities is normal. In fact, as we have seen, the bootstrap indicates that these quantities are not normally distributed although AUC is nearly so.

The bootstrap has utilised the evidence available in the sample in a manner that summary statistics are not able to do. It has shown that although the multivariate normal distribution might serve as an approximate model for the vector (A, b_1, b_2) it is by no means a good model for the derived parameters.

The objective of the experiment is to measure the derived parameters whereas statistical theory focuses on estimation and asymptotic normality of A, b_1 and b_2 . It is no easy task to obtain analytic results for the distribution of AUC, TMAX and CMAX. However, in the pharmacologic literature it is often assumed that the normal or log-normal distribution provides a good model for the marginal distributions

of these parameters. The results of this chapter indicate that for TMAX and CMAX such an assumption may well be erroneous.

In assessing the effectiveness of a drug one would like to know more than can be given by a few summary statistics.

e.g. Percentiles

Skewness

Construct tolerance intervals

Bi-variate plots of the derived parameters

Are the distributions symmetric?

Are there any unusual cases? The method given here gives answers to these questions without recourse to normal theory.

Other statistical aspects of bioavailability depend to a greater or lesser extent on the distribution of the derived parameters. For example comparative bioavailability is concerned with the comparison of two drug formulations and depends on these distributions.

B I B L I O G R A P H Y

- AGHA, M. (1971). A direct method for fitting linear combinations of exponentials. Biometrics, 27, 399-413.
- ANDERSON, S. and HAUCK, W.W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. Commun. Statist. Theor. Meth., 12, 23, 2663-2692.
- BABU, G.J. and SINGH, K. (1983). Inference on means using the bootstrap. Ann. Statist., 11, 3, 999-1103.
- BERAN, R. (1982). Estimated sampling distributions : The bootstrap and competitors. Ann. Statist., 10, 1, 212-225.
- BERAN, R. (1984a). Bootstrap methods in statistics. Jber. d. Dt. Math.Verein, 86, 14-30.
- BERAN, R. (1984b). Jackknife approximations to bootstrap estimates. Ann. Statist. 12, 1, 101-118.
- BICKEL, P.J. and FREEDMAN, D.A. (1980). On Edgeworth expansions for the bootstrap. Technical Report, Department of Statistics, University of California, Berkeley.
- BICKEL, P.J. and FREEDMAN, D.A. (1981). Some asymptotic theory for the bootstrap. Ann. Statist., 9, 6, 1196-1217.
- BOX, G.E.P. and LUCAS, H.L. (1959). Design of experiments in non-linear situations. Biometrika, 46, 77-90.
- BOX, G.E.P. and TIAO, G.C. (1962). A further look at robustness via Bayes' theorem. Biometrika, 49, 419-432.
- BOX, G.E.P. and TIAO, G.C. (1964). A note on criterion robustness and inference robustness. Biometrika, 51, 169-173.
- BOX, G.E.P. and TIAO, G.C. (1973). Bayesian inference in statistical analysis. Reading, Massachusetts : Addison-Wesley.

- BOXENBAUM, H.G., RIEGELMAN, S., ELASHOFF, R.M. (1974). Statistical estimation in pharmacokinetics. Journal of Pharmacokinetics and Biopharmaceutics, 2, 123-148.
- BRUCE, I. (1984). Drug levels in the body. Mathematical Gazette,
- BUTTON, C. (1979). Pharmacokinetics, Bioavailability, and dosage regimens of Digoxin in the horse and dog. Unpublished Ph.D. thesis, Graduate College of Texas, A & M University.
- CORNFIELD, J. (1969). The Bayesian outlook and its application. Biometrics, 25, 617-657.
- CORNELL, R.G. (1962). A method for fitting linear combinations of exponentials. Biometrics, 18, 104-113.
- COX, D.R. and HINKLEY, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.
- CURRIE, D.J. (1982). Estimating Michaelis-Menton parameters : bias, variance and experimental design. Biometrics, 38, 907-919.
- DELLA CORTE, M., BURICCHI, L. and ROMANO, S. (1974). On the fitting of linear combinations of exponentials. Biometrics, 30, 367-369.
- DIACONIS, P. and EFRON, B. (1983). Computer intensive methods in statistics. Technical Report No. 83, Division of Biostatistics, Stanford University.
- DIXON, W.J. (ed.) (1982). BMDP Biomedical Computer Programs, University of California Press, Berkeley.
- DUNFIELD, L.G. and READ, J.F. (1972). Determination of reaction rates by use of cubic spline interpolation. J. Phys. Chem., 57, 2178-2183.
- EFRON, B. (1979a). Bootstrap methods : Another look at the jackknife. Annals. Statist., 7, 1-26.
- EFRON, B. (1979b). Computers and the theory of statistics : Thinking the unthinkable. SIAM Review, 21, 4, 460-480.
- EFRON, B. (1981a). Nonparametric standard errors and confidence intervals (with discussion). Canadian J. of Statist., 9, 139-172.

- EFRON, B. (1981b). Censored data and the bootstrap. J. Amer. Statist. Assoc., 76, 374, 312-319.
- EFRON, B. (1982a). The jackknife, the bootstrap and other resampling plans. SIAM-NSF, CBMS No. 38.
- EFRON, B. (1982b). Estimating the error rate of a prediction rule : Improvements on cross-validation. Technical Report No. 78, Stanford University, Division of Biostatistics.
- EFRON, B., and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife and cross-validation. The American Statistician, Feb. 1983, 37, 1, 36-48.
- EFRON, B. (1984a). Bootstrap confidence intervals for parametric problems. Technical Report No. 90, Stanford University, Division of Biostatistics.
- EFRON, B. (1984b). Better bootstrap confidence intervals. L.C.S. Technical Report No. 14, Department of Statistics, Stanford University.
- EFRON, B. (1984c). Comparing non-nested linear models. J. Amer. Statist. Assoc., 79, 388, 791-803.
- EFRON, B. (1985). Bootstrap confidence intervals for a class of parametric problems. Biometrika, 72, 1, 45-58.
- EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. Ann. Statist., 9, 3, 586-596.
- EFRON, B. and TIBSHIRANI, R.J. (1985). The bootstrap method for assessing statistical accuracy. Technical Report No. 101, Stanford University, Division of Biostatistics.
- FAIRWEATHER, W.R. (1977). Investigating relationships between in vivo and in vitro pharmacological variables for the purpose of prediction. Journal of Pharmacokinetics and Biopharmaceutics, 5, 405-418.
- FLUEHLER, H., GRIEVE, A.P., MANDALLAZ, D., MAU, J. and MOSER, H.A. (1983). Bayesian approach to bioequivalence assessment : an example. Journal of Pharmaceutical Sciences, 72, 10, 1178-1181.

- FLUEHLER, H., HIRTZ, J. and MOSER, H.A. (1981). An aid to decision making in bioequivalence assessment. Journal of Pharmacokinetics and Biopharmaceutics, 9, 235-243.
- FREEDMAN, D.A. (1981). Bootstrapping regression models. Ann. Statist., 9,6, 1218-1228.
- FREEDMAN, D.A. and PETERS, S.C. (1984). Bootstrapping a regression equation : some empirical results. J. of Amer. Statist. Assoc., 79, 385, 97-106.
- FRESEN, J.L. and FRESEN, J.W. (1985). Estimating the parameter in the Pauling equation. Technical Report No. 2, Department of Statistics, University of the Western Cape.
- FRESEN, J.L. and JURITZ, J.M. (1985). A note on Foss's method of obtaining initial estimates for exponential curve fitting by numerical integration. Technical Report No. 2, Department of Statistics, University of the Western Cape.
- FRESEN, J.L. and FRESEN, J.W. (1985). Estimating the parameter in the Pauling equation. To appear in The Journal of Applied Statistics.
- FOSS, S.D. (1969). A method for obtaining initial estimates of the parameters in exponential curve fitting. Biometrics, 25, 580-584.
- FOSS, S.D. (1970). A method of exponential curve fitting by numerical integration. Biometrics, 26, 815-821.
- GARDNER, G.D., GARDNER, J.C., LAUSH, G. and MEINKE, W.W. (1959). Method of analysis of multicompartment exponential decay curves. Journal of Chemistry and Physics, 31, 978-986.
- GEERTSEMA, J.C. (1983). Recent views in the foundational controversy in statistics. S. Afr. Statist. Jour., 17, 2, 121-146.
- GONIN, R. and MONEY, A.H. (1985a). Non linear L_p -norm estimation. I. On the choice of the exponent, p , where the errors are additive. Commun. Statist. Theory and Methods, 14, 4, 827-840.
- GONIN, R. and MONEY, A.H. (1985b). Non linear L_p -norm estimation. II. The asymptotic distribution of the exponent, p , as a function of sample kurtosis. Commun. Statist. Theory and Methods, 14, 4, 841-850.

- GREENBLATT, D.T., SMITH, T.W., KOCH-WESER, J. (1984). Bioavailability of Drugs : The Digoxin Dilemma. Handbook of Clinical Pharmacokinetics. ADIS Health Science Press. Ed. Gibaldi, M. and Prescott, L., New York.
- GRIZZLE, J.E. (1965). The two period change over design and its use in clinical trials. Biometrics, 21, 467-480. Correction: Biometrics, 30, 727.
- HILLS, M. and ARMITAGE, P. (1979). The two period cross-over clinical trial. British Journal of Clinical Pharmacology, 8, 7-20.
- HINKLEY, D.V. (1983). How to use the bootstrap for interval estimation. Presented at the annual conference of the South African Statistical Association, Durban, 1983.
- HOLLANDER, M. and WOLFE, D.A. (1973). Nonparametric statistical methods. Wiley, New York.
- HUITSON, A., POLONIECKI, J., HEWS, R. and BARKER, N. (1982). A review of cross-over trials. The Statistician, 31, 1, 71-80.
- JEFFREYS, H. (1966). Theory of Probability (3rd ed.). Oxford : Clarendon, 1961 and 1966.
- JURITZ, J.M., GONIN, R. AND BRIDLE, T. (1983). A comparison of immediate release theophylline drug preparations in comparative bioavailability studies. Technical report, Institute of Biostatistics, Tygerberg.
- KENDALL, M.G. and STUART, A. (1977). The Advanced Theory of Statistics, (4th ed.) Vol. 1. Hafner, New York.
- KIRKWOOD, T.B.L. (1981). Bioequivalence testing : A need to rethink (Reader Reaction). Biometrics, 37, 589-591.
- KOCH, G.G. (1972). The use of nonparametric methods in the statistical analysis of the two period change over design. Biometrics, 28, 577-584.
- LANCZOS, C. (1957). Applied analysis. London, Pitman.
- LEVY, G. (1972). Bioavailability limits. CMA Journal, 107, 722.
- MANDALLAZ, D. and JOCHEN, M. (1981). Comparison of different methods for decision making in bioequivalence assessment. Biometrics, 37, 2, 213-222.
- MANDALLAZ, D. and MAU, J. (1981). Comparison of different methods for decision making in bioequivalence assessment. Biometrics, 37, 213-222.

- MANTEL, N. (1977). Do we want confidence intervals symmetric about the null value? Biometrics, Dec. 1977, 759-760.
- MCCABE, G.P. (Jr.)(1978). Evaluation of regression coefficient estimates using α -acceptability. Technometrics, 20, 2, 131-139.
- MELANDER, A. (1984). Influence of food on the bioavailability of drugs. Handbook of Clinical Pharmacokinetics. ADIS Health Science Press. Ed. Gibaldi, M. and Prescott, L., New York.
- METZLER, C.M. (1974). Bioavailability - a problem in equivalence. Biometrics, 30, 309-317.
- MILLER, R.G. (1974a). The jackknife - a review. Biometrika, 61, 1-15.
- MILLER, R.G. (1974b). An unbalanced jackknife. Ann. Statist., 2, 880-891.
- POCOCK, S.J. (1982). When not to rely on the central limit theorem - an example from absenteeism data. Commun. Statist. Theor. Meth., 11, 19, 2169-2179.
- POLONIECKI, J. and DANIEL, D. (1981). Further analysis of the Hills and Armitage Enuresis Data. The Statistician, 30, 3, 225.
- RHODDA, B.E., SAMPSON, C.B. and SMITH, D.W. (1975). The one compartment open model : some aspects of parameter estimation. Applied Statistics, 24, 3, 309-318.
- ROCKE, C.M. (1984). On testing for bioequivalence. Biometrics, 40, 225-230.
- ROYCE SADLER, D. (1975). Numerical methods for nonlinear regression. University of Queensland Press, St Lucia, Queensland.
- RUBIN, D.B. (1981). The Bayesian bootstrap. Ann. Statist., 9, 1, 130-134.
- SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. JASA, 80, 390, 360-361.
- SELWYN, M.R., DEMPSTER, A.P., HALL, N.R. (1981). A bayesian approach to bioequivalence for the 2x2 change over design. Biometrics, 37, 11-21.

- SEN, P.K. (1971). Robust statistical procedures in problems of linear regression with special reference to quantitative bioassays. Rev. Int. Statist. Inst., 39, 21-38.
- SHAH, B.K. (1973). Obtaining preliminary estimates to fit two-term exponential model to blood concentration data. J. Pharm. Sciences, 62, 1208-1209.
- SHIRLEY, E. (1976). The use of confidence intervals in biopharmaceutics. Journal of Pharmacy and Pharmacology, 28, 312-313.
- SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Statist., 9, 1187-1195.
- SMITH, M.R., COHN-SFETCU, S. and BUCKMASTER, H.A. (1976). Decomposition of multicomponent exponential decays by spectral analysis techniques. Technometrics, 18, 467-486.
- STEINIJANS, V.W. (1975). Pharmakokinetische Compartment - Modelle. Byk Gulden Pharmazeutika, Konstanz.
- STEINIJANS, V.W. and DILETTI, E. (1983). Statistical analysis of bio-availability studies : Parametric and non-parametric confidence intervals. Eur. J. Clin. Pharmacol., 24, 127-136.
- STEINIJANS, V.W. and DILETTI, E. (1985). Generalisation of distribution-free confidence intervals for bioavailability ratios. Eur. J. Clin. Pharmacol., 28, 85-88.
- STEYN, H.S. (1980). Fitting linear combinations of exponential decays by using Fourier analysis. COMPSTAT 1980. Physica-Verlag, 615-621.
- STEYN, H.S. and VAN WYK, J.W.J. (1977). Some methods for fitting compartment models to data. Wetenskaplike bydraes van die PU vir CHO, Reeks B : Natuurwetenskappe Nr. 91.
- TIBSHIRANI, R.J. (1984). Bootstrap confidence intervals. Technical Report No. 91, Stanford University, Division of Biostatistics.
- WAGNER, J.G. (1975). Fundamentals of Clinical Pharmacokinetics (2nd ed.) Drug Intelligence Publications, Inc., Hamilton, Illinois.

- WALLENSTEIN, S. and FISHER, A.C. (1977). The analysis of the two-period repeated measurements cross-over design with application to clinical trials. Biometrics, 33, 261-269.
- WESTLAKE, W.J. (1972). Use of confidence intervals in the analysis of comparative bioavailability trials. Journal of Pharmaceutical Sciences, 61, 1340-1341.
- WESTLAKE, W.J. (1974). The use of balanced incomplete block designs in comparative bioavailability trials. Biometrics, 30, 319-327.
- WESTLAKE, W.J. (1975). Statistical problems in interpreting comparative bioavailability trials. Int. J. Clin. Pharmacol. 11, 2, 342-345.
- WESTLAKE, W.J. (1976). Symmetrical confidence intervals for bioequivalence trials. Biometrics, 32, 741-744.
- WESTLAKE, W.J. (1977). Reply to letter to the Editor from N. Mantel. Biometrics, 33, 760.
- WESTLAKE, W.J. (1979). Statistical aspects of comparative bioavailability trials. Biometrics, 35, 273-280.
- WESTLAKE, W.J. (1981). Bioequivalence testing - a need to rethink (Reader reaction response). Biometrics, 37, 580-594.
- YEH, K.C. and KWAN, K.C. (1978). A comparison of numerical integrating algorithms by trapezoidal, Legrange and spline approximation. J. Pharm. Biopharm., 6, 1, 79-98.

The following computer program will compute bootstrap bias corrected confidence intervals for data from a bioequivalence trial. Confidence intervals are computed for three common measures of bioequivalence:

- 1) difference between formulation means - THETA1
- 2) ratio of formulation means - THETA2
- 3) geometric mean of individual ratios - THETA3.

The response may be either AUC, TMAX, CMAX or even some other bioavailability parameter of interest.

It is assumed that there are 'N' subjects, each of whom has two responses - for the new and the standard formulation. The data are read in with a maximum of ten on a line in free format. The data for the standard formulation are entered first and then the data for the new formulation are entered.

The size of the bootstrap (=M), the number of subjects (=N) and the required coverage probability (=C) are treated as parameters and are defined in the first three lines of the program. These are currently set at the values B = 1000, N = 12 and C = 0.95.

Example of input format: data from Steinijans and Diletti(1983)

```
136.0 152.6 123.1 77.0 115.7 72.0 116.4 151.1 118.9
156.1 222.4 158.1 (Standard data)
135.7 155.3 148.9 81.2 139.2 91.7 118.7 133.2 115.6
150.3 223.9 154.1 (New data)
```

The program listing is given below.

```
PARAMETER M=1000
PARAMETER N=12
PARAMETER C=0.95
```

```
C M IS THE NO OF TIMES WE ARE GOING TO BOOTSTRAP
C N IS THE NO OF SUBJECTS
C THIS MUST BE CHANGED FOR DIFFERENT EXAMPLES
C C IS THE REQUIRED COVERAGE PROBABILITY
```

```
DIMENSION AUC(2,N),AGUC(2,N),THETA1(M),THETA2(M),
DIMENSION XU(3),XL(3),ANSL(3),ANSU(3),THETA3(M),CDF(3)
```

```
C AUC IS THE AREA UNDER THE CURVE
C AGUC IS THE VECTOR GENERATED AFTER USING URAND
C THETA1,THETA2,THETA3 ARE ESTIMATORS
C CDF IS THE CUMULATIVE DISTRIBUTION FUNCTION
C THREE PARAMETERS.
```

```
INTEGER NO(N),LANS(3),UANS(3)
C NO IS AN ARRAY THAT CONTAINS THE GENERATED NOS
REAL MEAN(2),INVPHI(3),INALPH(2),LBOUND(3),UBOUND(3)
```

```

100  FORMAT(
110  FORMAT(15X,F6.2,18X,F6.2/)
120  FORMAT(4X,'DIFFERENCE OF MEANS = ',F6.2,/,/,4X,
    & 'RATIO OF MEANS = ',F6.2,/,/,4X,
    & 'GEOMETRIC MEAN OF RATIOS = ',F6.2)
130  FORMAT(1H0,'MEAN(1) = ',F6.2,5X,'MEAN(2) = ',F6.2,/,1H0,
    & 'THETA1(',I3,') = ',F6.2,2X,'THETA2(',I3,') = ',F6.2,
    & 2X,'THETA3(',I3,') = ',F6.2)
140  FORMAT(1H1,20X,'ORIGINAL VALUES OF DATA',/,20X,23('-'),
    & /,/,10X,'STANDARD FORMULATION',4X,'NEW FORMULATION',/,/)
150  FORMAT(6X,'MEAN = ',F6.2,13X,'MEAN = ',F6.2,/)

```

```

READ(5,100)((AUC(I,J),J=1,N),I=1,2)
WRITE(6,140)
WRITE(6,110)((AUC(I,J),I=1,2),J=1,N)

```

```

C-----
C CALCULATE THE MEANS OF THE STANDARD AND NEW FORMULATIONS
C-----

```

```

DO 10 I=1,2
SUM=0.0
DO 20 J=1,N
SUM=AUC(I,J)+SUM
20 CONTINUE
MEAN(I)=SUM/N
10 CONTINUE
WRITE(6,150)(MEAN(I),I=1,2)

```

```

C-----
C CALCULATE THE DIFFERENCE OF MEANS,RATIO OF MEANS AND
C GEOMETRIC MEAN OF RATIOS
C-----

```

```

THET1=MEAN(2)-MEAN(1)
THET2=MEAN(2)/MEAN(1)
TOTALS=1.0
TOTALN=1.0
DO 90 I=1,12
TOTALS=AUC(1,I)*TOTALS
TOTALN=AUC(2,I)*TOTALN
90 CONTINUE
THET3=(TOTALN/TOTALS)**(1.0/N)
WRITE(6,120)THET1,THET2,THET3

```

```

C-----
C GENERATE M NEW AUC(1) VECTORS AND M NEW AUC(2) VECTORS
C-----

```

```

ISEED=0
DO 40 J=1,M
ISEED=ISEED+20
DO 30 I=1,N
K=INT(URAND(ISEED)*N+1)
C K IS A RANDOM NO BETWEEN 1 AND 12
NO(I)=K
30 CONTINUE

DO 50 I=1,2
DO 60 L=1,12
AGUC(I,L)=AUC(I,NO(L))
60 CONTINUE

```

50 CONTINUE

C-----
 C CALCULATE THE MEANS, DIFFERENCE OF MEANS, RATIO OF MEANS AND
 C GEOMETRIC MEAN OF RATIOS M TIMES
 C-----

```

      DO 70 I=1,2
      SUM=0.0
      DO 80 L=1,N
      SUM=AGUC(I,L)+SUM
80    CONTINUE
      MEAN(I)=SUM/N
70    CONTINUE
      THETA1(J)=MEAN(2)-MEAN(1)
      THETA2(J)=MEAN(2)/MEAN(1)
      TOTALS=1.0
      TOTALN=1.0
      DO 200 I=1,12
      TOTALS=AGUC(1,I)*TOTALS
      TOTALN=AGUC(2,I)*TOTALN
200   CONTINUE
      THETA3(J)=(TOTALN/TOTALS)**(1.0/N)
40    CONTINUE
  
```

C-----
 C SORT THE ARRAYS
 C-----

```

      CALL SORT(THETA1,M)
      CALL SORT(THETA2,M)
      CALL SORT(THETA3,M)

      WRITE(6,501)(THETA1(I),THETA2(I),THETA3(I),I,I=1,M)
501   FORMAT(3(5X,F6.2),5X,I6)
  
```

C-----
 C CALCULATE THE CUMULATIVE DISTRIBUTION FUNCTIONS
 C-----

```

      CALL FUNCT(THETA1,THET1,M,CDF(1))
      CALL FUNCT(THETA2,THET2,M,CDF(2))
      CALL FUNCT(THETA3,THET3,M,CDF(3))

C     WRITE(6,170)(CDF(I),I=1,3)
170   FORMAT(3(5X,F6.2))
  
```

C-----
 C CALCULATE THE CONFIDENCE INTERVALS
 C-----

```

      CO=2.515517
      C1=0.802853
      C2=0.010328
      D1=1.432788
      D2=0.189269
      D3=0.001308
      ALPHA=(1.0-C)/2.0
      ALPH=1-ALPHA

      CALL INV(CDF(1),CO,C1,C2,D1,D2,D3,INVPHI(1))
      CALL INV(CDF(2),CO,C1,C2,D1,D2,D3,INVPHI(2))
      CALL INV(CDF(3),CO,C1,C2,D1,D2,D3,INVPHI(3))
  
```

```

      CALL INV(ALPHA,CO,C1,C2,D1,D2,D3,INALPH(1))
      CALL INV(ALPH,CO,C1,C2,D1,D2,D3,INALPH(2))
C     WRITE(6,180)(INVPHI(I),I=1,3),(INALPH(II),II=1,2)
180   FORMAT(2X,F8.4)

      DO 230 I=1,3
        XL(I)=INALPH(1)+2*INVPHI(I)
        XU(I)=INALPH(2)+2*INVPHI(I)
230   CONTINUE

      A=0.2316419
      PI=3.141593
      B1=0.319381530
      B2=-0.356563782
      B3=1.781477937
      B4=-1.821255978
      B5=1.330274429

      DO 240 I=1,3
        CALL PHI(XL(I),A,PI,B1,B2,B3,B4,B5,ANSL(I))
        CALL PHI(XU(I),A,PI,B1,B2,B3,B4,B5,ANSU(I))
240   CONTINUE

C     WRITE(6,250)(XL(I),XU(I),ANSL(I),ANSU(I),I=1,3)
250   FORMAT(4(5X,F8.4))

      DO 270 I=1,3
        LANS(I)=INT(ANSL(I)*M)
        UANS(I)=INT(ANSU(I)*M)
270   CONTINUE

      LBOUND(1)=THETA1(LANS(1))
      LBOUND(2)=THETA2(LANS(2))
      LBOUND(3)=THETA3(LANS(3))
      UBOUND(1)=THETA1(UANS(1))
      UBOUND(2)=THETA2(UANS(2))
      UBOUND(3)=THETA3(UANS(3))

280   WRITE(6,280)(LBOUND(I),I=1,3),(UBOUND(I),I=1,3)
      &     FORMAT(1H1,20X,'CONFIDENCE INTERVALS',/,21X,20('-'),/,/,
      &     26X,'THETA1',10X,'THETA2',10X,'THETA3',/,/,
      &     5X,'LOWER BOUND',10X,F6.2,10X,F6.2,10X,F6.2,/,/,
      &     5X,'UPPER BOUND',10X,F6.2,10X,F6.2,10X,F6.2)

      STOP
      INCLUDE UCT*ASCII.URAND
      END
      SUBROUTINE FUNCT(THETA,THET,LEN,CDF)
      DIMENSION THETA(LEN)
      DO 270 I=1,LEN
        IF(THETA(I).GT.THET)THEN
          CDF=(I-1)*1.0/LEN
          GO TO 300
        END IF
270   CONTINUE
300   RETURN
      END

      SUBROUTINE SORT(ARRAY,LENGTH)

```

```

DIMENSION ARRAY(LENGTH)
DO 210 I=1,LENGTH-1
DO 220 L=I+1,LENGTH
  IF(ARRAY(I).GT.ARRAY(L))THEN
    TEMP=ARRAY(I)
    ARRAY(I)=ARRAY(L)
    ARRAY(L)=TEMP
  END IF
220 CONTINUE
210 CONTINUE
RETURN
END

SUBROUTINE INV(CDF,CO,C1,C2,D1,D2,D3,INVPHI)
REAL INVPHI
IF(CDF.GT.0.5.AND.CDF.LT.1.0)THEN
  T=(-2.0*ALOG(1.0-CDF))**(1.0/2.0)
ELSE
  T=(-2.0*ALOG(CDF))**(1.0/2.0)
END IF

Z=T-((CO+C1*T+C2*(T*T))/(1+D1*T+D2*(T*T)+D3*(T*T*T)))
IF(CDF.GT.0.0.AND.CDF.LE.0.5)THEN
  INVPHI=-Z
ELSE
  INVPHI=Z
END IF
RETURN
END

SUBROUTINE PHI(T,A,PI,B1,B2,B3,B4,B5,ANS)
X=1.0/(1.0+A*ABS(T))
Y=((2.0*PI)**(-0.5))*(EXP(-(T*T)/2.0))
R=Y*(B1*X+B2*(X*X)+B3*(X*X*X)+B4*(X*X*X*X)+B5*(X*X*X*X*X))
IF(T.LT.0.0)THEN
  ANS=R
ELSE
  ANS=1-R
END IF
RETURN
END

```