AD-A026 636

PROCEEDINGS OF THE CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING (21ST), HELD AT THE WALTER REED ARMY MEDICAL CENTER, WASHINGTON, D.C. ON 22-24 OCTOBER 1975

ARMY MATHEMATICS STEERING COMMITTEE

PREPARED FOR
ARMY RESEARCH OFFICE

MAY 1976

ARO Report 76-2

PROCEEDINGS OF THE TWENTY-FIRST CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by

The Army Mathematics Steering Committee

on Behalf of

THE CHIEF OF RESEARCH, DEVELOPMENT AND ACQUISITION

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO Report 76-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PROCEEDINGS OF THE TWENTY-FIRST CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Army Mathematics Steering Committee on Behalf of the Chief of Research, Development and Acquisition		12. REPORT DATE May 1976 13. NUMBER OF PAGES 502
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) US Army Research Office PO Box 1211 Research Triangle Park, NC 27709		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distrubition unlimited. The findings in this report are not to be construed as official Department of the Army position, unless so designated by other authorized documents.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
This is a technical report resulting from the Twenty-first Conference on the Design of Experiments in Army Research, Development and Testing. It contains most of the papers presented at that meeting. These treat various Army statistical and design problems.

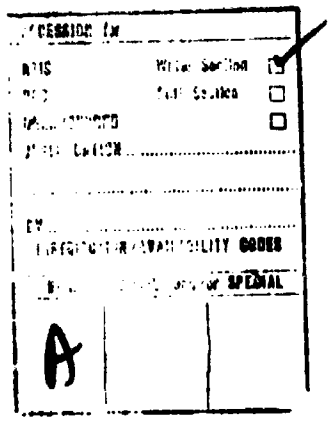
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

rifling configurations
man-machine interface
flight trajectory
nonrandomized clinical trials
measurement models
factorial designs
error variance
quantal response tests
flight suitability
life testing techniques
the exponential distribution
random interval reliability
renewel processes
unknown signal detector
outlier detection
simulation of physiological systems
synergy in drug mixtures
binomial clinical trials
predictive sample reuse
peer evaluations
camouflage via image interpreters
continuous sampling
semi Markov chains
progressively censored sampling

FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



U. S. Army Research Office

Report No. 76-2
May 1976

PROCEEDINGS OF THE TWENTY-FIRST CONFERENCE ON THE DESIGN OF EXPERIMENTS

Sponsored by the Army Mathematics Steering Committee

HOSTS

Walter Reed Army Medical Center

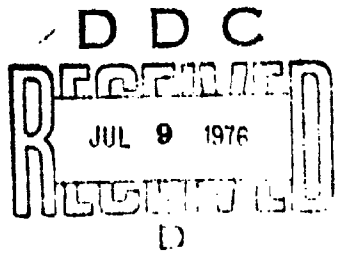
and

Armed Forces Institute of Pathology

22-24 October 1975

Approved for public release; distribution unlimited. The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, North Carolina



FOREWORD

The 21st Conference on the Design of Experiments in Army Research, Development and Testing was held 22-24 October 1975 in Washington, DC. The Conference, which took place at the Walter Reed Medical Complex, had two hosts: the Walter Reed Army Medical Center and the Armed Forces Institute of Pathology. Both hosts furnished excellent conference rooms and meeting rooms for this symposium. Planning for these meetings requires painstaking attention to detail and we are indebted to Dr. Walter D. Foster and Dr. James N. Young, both of the Armed Forces Institute of Pathology, for serving well as Chairmen for Local Arrangments. We are pleased that Major General Robert Bernstein, Commander of the Walter Reed Army Medical Center, opened the Conference and welcomed us. This is not the first meeting to be held at the Walter Reed installation. On each occasion, the reception given us has been excellent, and we look forward to meetings there again in the future.

There were four addresses by invited speakers. Traditionally an attempt is made by the Program Committee to have expository talks on themes somewhat pertinent to the mission of the Army installation at which the annual conference is held. Success along these lines was achieved again. The first address was given by Frederick Mosteller of Harvard University, who spoke on "Success in Social and Medical Experimentation." Dr. Mosteller was given, at his request, two hours to deliver his address. Normally, there would have been five invited addresses, but the length of Professor Mosteller's talk led to four at this meeting. Dr. Mosteller's talk was given at the first morning of the Conference and was followed in the late afternoon by two papers on clinical trials. There has been much in the medical and statistical literature on this topic. Professor Edmund A. Gehan of the University of Texas System Cancer Center spoke on "Non-randomized Clinical Trials" and Professor Paul Meier of the University of Chicago addressed the audience on "Randomized Clinical Trials." On the second day of the Conference, Professor Seymour Geisser of the University of Minnesota gave an invited address on "Predictive Sample Reuse." This was followed on the morning of the last day of the Conference by a talk on "Normality and Disease" given by Professor Edmond A. Murphy of the Johns Hopkins Medical School.

One major purpose of the Conference is to bring together those engaged in scientific work in Army installations with investigators from other government agencies and those from university life. This interaction has been going on successfully since the inception of the program. Statisticians and others in Army installations discuss their work at technical sessions and clinical sessions at each annual conference. For this Conference there were seven technical sessions comprising 24 papers and four clinical sessions. At the clinical sessions a panel of experts responds to problems raised by those in Army installations who have usually given advance manuscript copies to the panelists.

Besides the technical aspects, these sessions provide a source for initiating future collaboration between scientists in Army installations and those in university life.

At the start of this year's opening session, Dr. Walter D. Foster was honored with a Certificate for Achievement for the valuable contributions he made during his twelve years as Chairman of the Probability and Statistics Subcommittee of the Army Mathematics Steering Committee. He was specifically cited for "continuously and vigorously crusading for application of sound statistical principles and methodology to problems in Army research and development."

On the evening of the first day of the Conference, a banquet is held at which the Samuel S. Wilks Memorial Award of the American Statistical Association and the Department of the Army is presented. At this meeting the 11th award was presented by Lester Frankel, President of the ASA, to Dr. Herbert Solomon, Professor of Statistics, Stanford University. The award was made to Dr. Solomon for his significant contributions to statistical methodology and for his outstanding contributions in the application of statistics in the service of the nation.

The Army Mathematics Steering Committee sponsors these meetings on behalf of the Office of the Chief of Research and Development and Acquisition to bring new developments in statistics to Army scientists and engineers and to expose them to thinking that could be profitable to them in the execution of their missions. The Committee has asked that the proceedings of the Conference be published and issued Army-wide and to other scientific communities.

At the beginning of each calendar year the program committee for these conferences is selected and meets in Washington, DC, to suggest areas of interest, to outline a program, and to suggest speakers for the meeting to be held later that year. I would like to express my appreciation to Dr. Frank Grubbs, Program Chairman for this year's committee, and to Dr. Douglas Tang, Chairman of the Subcommittee on Probability and Statistics, Army Mathematics Steering Committee, for their efforts and great help. My thanks also go to other committee members involved in developing this year's program: Drs. David W. Alling, Gary A. Chase, Walter D. Foster, Bernard Harris, J. Stuart Hunter, Clifford J. Maloney, Badrig Kurkjian, Marvin Schneiderman. Francis Dressel, as always, was helpful in many ways in making sure the program was a success. Thus many hands helped in guiding this Conference to a successful conclusion, and this is very much appreciated.

Herbert Solomon
Conference Chairman

TABLE OF CONTENTS

AGENDA

THE TWENTY-FIRST CONFERENCE ON THE DESIGN OF EXPERIMENTS

IN ARMY RESEARCH, DEVELOPMENT AND TESTING

22-24 October 1975

The Armed Forces Institute of Pathology

***** Wednesday, 22 October *****

0830-0930 REGISTRATION -- Lobby of Sternberg Auditorium (WRAIR)

0930-1220 GENERAL SESSION I -- Sternberg Auditorium

CALLING OF CONFERENCE TO ORDER

Dr. Walter D. Foster, Chairman on Local Arrangements, Armed Forces Institute of Pathology, Washington, D. C.

WELCOMING REMARKS

Major General Robert Berstein. Commander, WRAMC

CHAIRMAN OF SESSION I

Dr. Frank E. Grubbs, Program Committee Chairman, Aberdeen Proving Ground, Maryland

SUCCESS IN SOCIAL AND MEDICAL EXPERIMENTATION

Professor Frederick Mosteller, Department of Statistics, Harvard University, Cambridge, Massachusetts

1050-1120 BREAK

1120-1220 GENERAL SESSION I (CONTINUED)

SECOND PART OF THE ADDRESS BY PROFESSOR MOSTELLER

1220-1320 LUNCH -- Officers' Open Mess, WRAMC

***** Wednesday *****

1320-1430 CLINICAL SESSION A -- Dart Auditorium (AFIT)

CHAIRMAN

Boyd Harshbarger, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

PANELISTS

Robert Bechhofer, Department of Operations Research, Cornell University, Ithaca, New York

Joseph M. Cameron, Statistical Engineering Laboratory, National Bureau of Standards, Wahsington, D. C.

A. Clifford Cohen, Institute of Statistics, University of Georgia, Athens, Georgia

J. Richard Moore, U.S. Army Ballistics Research Laboratories Aberdeen Proving Ground, Maryland

INVESTIGATIONS OF INTERFACE BETWEEN 5.56MM BULLETS AND RIFLING CONFIGURATIONS

Dennis Conway, Munitions Development and Engineering Directorate, Frankford Arsenal, Philadelphia, Pennsylvania

A STEP TOWARD THE RATIONAL DESIGN OF EXPERIMENTS IN METAL-FORMING TECHNOLOGY

Paul Gordon, Materials Engineering Division, Pitman-Dunn Laboratory, Frankford Arsenal, Philadelphia, Pennsylvania

1320-1430 TECHNICAL SESSION 1 -- Owen Conference Room (AFIP)

CHAIRMAN

Lang Withers, U.S. Army Operational Test and Evaluation Agency, Fort Belvoir, Virginia

DESIGN OF EXPERIMENTS DEALING WITH MAN-MACHINE INTERFACE IN CURRENT COMMUNICATIONS SYSTEMS

R. J. D'Accardi, H. S. Bennett, U. S. Army Electronics Command, Fort Monmouth, New Jersey

J. R. Hennessy, U.S. ARMY MERDC, Fort Belvoir, Virginia

PLANNING FOR THE MEASUREMENT OF FLIGHT TRAJECTORY

J. B. Gose and J. V. Carrillo, Quality Assurance Office U.S. Army White Sands Missile Range, White Sands, New Mexico

***** Wednesday *****

1430-1500 BREAK

1500-1710 GENERAL SESSION II -- Sternberg Auditorium (WRAIR)

CHAIRMAN

Dr. Marvin A. Schneiderman, National Cancer Institute, Bethesda, Maryland

NONRANDOMIZED CLINICAL TRIALS

Professor Edmund A. Gehan, Department of Biomathematics, University of Texas System Cancer Center, Houston, Texas

RANDOMIZED CLINICAL TRIALS

Professor Paul Meier, Department of Statistics, The University of Chicago, Chicago, Illinois

1830-1915 SOCIAL GATHERING -- Officers' Open Mess, WRAMC

1915- BANQUET

PRESENTATION OF THE SAMUEL S. WILKS MEMORIAL AWARD

Dr. Frank E. Grubbs, Master of Ceremonies

***** Thursday, 23 October *****

0830-1010 CLINICAL SESSION B -- Dart Auditorium (AFIP)

CHAIRMAN

Badrig Kurkjian, U.S. Army Materiel Command, Alexandria, Virginia

PANELISTS

Robert Bechhofer, Department of Operations Research, Cornell University, Ithaca, New York

Seymour Geisser, School of Statistics, University of Minnesota, Minneapolis, Minnesota

J. Richard Moore, U.S. Army Ballistics Research Laboratories, Aberdeen Proving Ground, Maryland

Richard L. Moore, U.S. Army Armament Command, Rock Island, Illinois

***** Thursday *****

CLINICAL SESSION B (CONTINUED)

EMPIRICAL COMPARISON OF CRITERION-REFERENCED MEASUREMENT MODELS

Frederick H. Steinheiser, Jr. and Kenneth I. Epstein, U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, Virginia

PRESSURE IMPULSE METHODOLOGY

Barry H. Rodin, Concepts Analysis Laboratory, Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland

0830-1010 TECHNICAL SESSION 2 -- Owen Conference Room (AFIP)

CHAIRMAN

Douglas B. Tang, Department of Biostatistics/Applied Mathematics, Division of Biometrics and Medical Information Processing, Walter Reed Army Institute of Research, Washington, D. C.

NONRANDOMIZED FACTORIAL DESIGNS CHARACTERIZED BY TREND ELIMINATION AND A MINIMUM NUMBER OF FACTOR LEVEL CHANGES

Les Lancaster and Steve Reynolds, U.S. Army Operational Test and Evaluation Agency, Fort Belvoir, Virginia

A METHOD OF ESTIMATING ERROR VARIANCE IN A NON-REPLICATED EXPERIMENT BY PARTITIONING AN INTERACTION TERM INTO NON-ADDITIVITY AND ERROR

Lieutenant L. Douglas Peirce, Army Logistics Management Center, School of Logistics Science, Systems and Cost Analysis Department, Fort Lee, Virginia

PLANNING QUANTAL RESPONSE TESTS FOR ORDNANCE DEVICES: THE TWO-POINT STRATEGY AND ANALYSIS

R. E. Little, The University of Michigan-Dearborn, School of Engineering, Dearborn, Michigan

***** Thursday *****

0830-1010 TECHNICAL SESSION 3 -- Carroll Auditorium

CHAIRMAN

Eugene F. Dutoit, U.S. Army Infantry School, Directorate of Combat Developments, Fort Benning, Georgia

APPLICATIONS OF THE MONTE CARLO TECHNIQUE TO DETERMINE STATISTICAL STRESS AND STRAIN RESPONSE AROUND CUT-OUTS IN COMPOSITES

Donald M. Neal, Army Materials and Mechanics Research Center, Watertown, Massachusetts

TECHNIQUE FOR STATISTICALLY DETERMINING FLIGHT SUITABILITY OF AN ARTILLARY PROJECTILE

Gertrude Weintraub and Ronald Corn, Picatinny Arsenal, Dover, New Jersey

BAYESIAN SYSTEM RELIABILITY GROWTH ANALYSIS USING SUBSYSTEM DATA

John G. Mardo, Picatinny Arsenal, Dover, New Jersey

1010-1040 BREAK

1040-1220 CLINICAL SESSION C -- Dart Auditorium (AFIP)

CHAIRMAN

R. J. D'Accardi, U.S. Army Electronics Command, Fort Monmouth New Jersey

PANELISTS

A. Clifford Cohen, Institute of Statistics, University of Georgia, Athens, Georgia

Larry H. Crow, U.S. Army Materiel Systems Analysis Agency, Aberdeen Proving Ground, Maryland

Bernard Harris, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin

Herbert Solomon, Department of Statistics, Stanford University Stanford, California

***** Thursday *****

CLINICAL SESSION C (CONTINUED)

APPLICATION OF LIFE TESTING TECHNIQUES TO DETECTION DATA

Carl B. Bates, U.S. Army Concepts Analysis Agency, Bethesda, Maryland

TEST DESIGN CONSIDERATIONS IN COMOUGLAGE OF THE M60A1 TANK

MAJ William K. Emerson, USAMERDC, R&D Coordinator, Fort Belvoir, Virginia

1040-1220 TECHNICAL SESSION 4 -- Owen Conference Room (AFIP)

CHAIRMAN

Robert Burge, Department of Biostatistics/Applied Mathematics Division of Biometrics and Medical Information Processing, Walter Reed Army Institute of Research, Washington, D. C.

ON THE ROBUSTNESS OF THE EXPOTENTIAL DISTRIBUTION

George C. Canavos, School of Business, Virginia Commonwealth University, Richmond, Virginia

RANDOM INTERVAL RELIABILITY

Gerald R. Andersen, Office AMC Chief Mathematician, HQ, U.S. Army Materiel Command, Alexandria, Virginia

CONFIDENCE INTERVALS FOR A SUM OF RENEWAL PROCESSES WITH APPLICATION IN RELIABILITY

Ronald L. Racicot, Applied Math & Mechanics Division, Research Directorate, Benet Weapons Laboratory, Watervliet Arsenal, Watervliet, New York

STRUCTURAL VARIANCE ESTIMATION

Clifford J. Maloney and Lucille Carver, Bureau of Biologics, FDA, Rockville, Maryland

1220-1320 LUNCH -- Officers' Open Mess, WRAMC

***** Thursday *****

1320-1520 CLINICAL SESSION D- Dart Auditorium (AFIP)

CHAIRMAN

Clifford J. Maloney, Bureau of Biologics, FDA, Bethesda, Maryland

PANELISTS

Frank E. Grubbs, Aberdeen Proving Ground, Maryland

Bernard Harris, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin

Richard L. Moore, U.S. Army Armament Command, Rock Island, IL

Herbert Solomon, Department of Statistics, Stanford University Stanford, California

UNKNOWN SIGNAL DETECTOR IN A MULTIPLE OBJECT SITUATION

John Bart Wilburn, Jr., I&M Branch, U.S. Army Electronic Proving Ground, Fort Huachuca, Arizona

OUTLIER DETECTION PROCEDURES IN TRAJECTORY DATA REDUCTION

William S. Agee and Robert H. Turner, Analysis and Computation Division, National Range Operations Directorate, U.S. Army White Sands Missile Range, White Sands, New Mexico

1320-1520 TECHNICAL SESSION 5 - Owen Conference Room (AFIP)

CHAIRMAN

Ian McLean, Armed Forces Institute of Pathology, Washington, D. C.

APPLYING SIMULATION OF PHYSIOLOGICAL SYSTEMS TO THE DESIGN OF EXPERIMENTS: EXAMPLES OF ENDOCRINE AND RESPIRATORY FUNCTIONS

Stanley M. Finkelstein, Division of Biological Engineering and Department of Operations Research and System Analysis, Polytechnical Institute of New York, Brooklyn, New York

Stanley S. Reisman, Department of Electrical Engineering, New Jersey Institute of Technology, Newark, New Jersey

A DESIGN FOR THE DETECTION OF SYNERGY IN DRUG MIXTURES

P. V. Piserchia and B. V. Shah, Statistics Research Division, Research Triangle Institute, Research Triangle Park, North Carolina

***** Thursday *****

TECHNICAL SESSION 5 (CONTINUED)

A NEW SAMPLING RULE FOR SEQUENTIAL BINOMIAL CLINICAL TRIALS

R. Srinivasan, Department of Mathematics, Temple University, Philadelphia, Pennsylvania

1520-1550 BREAK

1550-1700 GENERAL SESSION III -- Dart Auditorium (AFIP)

CHAIRMAN

Professor J. Stuart Hunter, Committee on National Statistics, National Academy of Sciences, Washington, D. C.

PREDICTIVE SAMPLE REUSE

Professor Seymour Geisser, School of Statistics, University of Minnesota, Minneapolis, Minnesota

***** Friday, 24 October *****

0830-1020 TECHNICAL SESSION 6 -- Dart Auditorium (AFIP)

CHAIRMAN

Joseph Rothberg, Division of Neuropsychiatry, Walter Reed Army Institute of Research, Washington, D. C.

VAIROUS METHODOLOGICAL APPROACHES TO PEER EVALUATIONS

Ronald G. Downey and Paul J. Duffy, U.S. Army Research Institute for Behavioral and Social Sciences, Arlington, Virginia

OBJECTIVE ANALYSIS OF CAMOUFLAGE VIA IMAGE INTERPRETERS

Ronald Johnson, USAMERDC, Countersurveillance and Topographic Division, Fort Belvoir, Virginia

NATO JOINT FIELD TRIAL ON AIR DEFENSE SITE CAMOUFLAGE

Allan T. Sylvester II, USAMERDC, Countersurveillance and Topographic Division, Fort Belvoir, Virginia

***** Friday *****

0830-1020 TECHNICAL SESSION 7 -- Owen Conference Room (AFIP)

CHAIRMAN

Beatrice S. Orleans, Naval Ships System Command, Washington, D. C.

A SIMPLE METHOD FOR DETERMINING THE UNRESTRICTED AVERAGE OUTGOING QUALITY LIMIT (UAOQL) OF A CONTINUOUS SAMPLING PLAN

Richard M. Brugger, RAM Assessment Division, Product Assurance Directorate, U.S. Army Armament Command, Rock Island, Illinois

SEMI MARKOV CHAINS APPLIED TO MARKOV CHAIN FUNCTIONALS PARTIALLY DEPENDENT ON RANDOM RETROGRADE TIME SHIFTS

David L. Arp, Naval Weapons Center, China Lake, California

PROGRESSIVELY CENSORED SAMPLING IN THE LOG-NORMAL DISTRIBUTION

A. Clifford Cohen, University of Georgia, Department of Statistics and Computer Science, Athens, Georgia

1020-1050 BREAK

1050-1220 GENERAL SESSION IV -- Dart Auditorium (AFIP)

CHAIRMAN

Professor Herbert Solomon, Chairman of the Conference, Department of Statistics, Stanford University, Stanford, California

OPEN MEETING OF THE AMSC SUB-COMMITTEE ON PROBABILITY AND STATISTICS

Dr. Douglas B. Tang, Department of Biostatistics and Applied Mathematics Division, Biometrics and Med. Info. Proc., Walter Reed Army Institute of Research, Washington, D. C.

NORMALITY AND DISEASE

Professor Edmond A. Murphy, Professor of Medicine, The Johns Hopkins Hospital, Baltimore, Maryland

1220-1320 LUNCH -- Officers' Open Mess

INVESTIGATIONS OF INTERFACE BETWEEN 5.56MM BULLETS AND RIFLING CONFIGURATIONS

DENNIS J. CONWAY
MUNITIONS DEVELOPMENT & ENGINEERING DIRECTORATE
U.S. ARMY FRANKFORD ARSENAL
PHILADELPHIA, PA

Abstract. The interface between 5.56mm ball and tracer bullet designs and various rifling configurations are examined to determine the effects on ballistic performance and mechanical integrity as would be experienced under general purpose machine gun operational modes.

Two modes of projectile failure are examined against light machine-gun system design criteria. Based on these results, optimum rifling configurations are identified for use in a machine-gun system.

Verification of these optimized rifling designs through experimentation are discussed.

1. Introduction. Initial interest in the study of those parameters effecting barrel/bullet interface was generated at Frankford Arsenal under the 6mm tracer program. At that time, the 6mm ball and tracer cartridges were the prime ammunition candidates for the Squad Automatic Weapon (SAW), and consequently great concern was expressed at a high incidence of tracer projectile failures (break-up) then being observed during both test barrel and weapon barrel performance tests.

Table 1 categorizes various tracer projectile malfunctions from four and six-groove, plated and unplated weapon and test barrels. This chart shows the frequency of projectile failures from four-groove plated weapon barrels and to a lesser degree in four-groove plated test barrels.

As a result of this high incidence of projectile failure, an analytic stress study was undertaken to examine certain modes of failure which could explain the type of projectile break-up being exhibited.

TABLE 1

COMPARATIVE RESULTS SHOWING 6MM TRACER MALFUNCTIONS IN FOUR- AND SIX-GROOVE BARRELS

TRACER DESIGN	BARREL TYPE	QUANTITY	BREAK-UPS	% BREAK-UPS
1	6-GROOVE, UNPLATED, TEST BRL	40	-	-
1	6-GROOVE, PLATED, WPN BRL	40	-	-
1	4-GROOVE, UNPLATED, TEST BRL	40	-	-
1	4-GROOVE, PLATED, TEST BRL	29	1	3
1	4-GROOVE, PLATED, WPN BRL	40	15	38
2	6-GROOVE, UNPLATED, TEST BRL	40	-	-
2	6-GROOVE, PLATED, WPN BRL	40	-	-
2	4-GROOVE, UNPLATED, TEST BRL	40	-	-
2	4-GROOVE, PLATED, TEST BRL	40	1	2.5
2	4-GROOVE, PLATED, WPN BRL	40	7	18

NOTES:

1. DESIGN #1: 15-80-5 GMCS JACKET, .024 WALL, .2428 DIAMETER
2. DESIGN #2: 25-70-5 GMCS JACKET, .024 WALL, .2428 DIAMETER
3. ALL FIRING SINGLE SHOT

2. Stress Evaluation. Th typical 6mm tracer failure as observed in recovered projectiles was evidenced by a radial flaring of the projectile base and longitudinal separation of the projectile jacket, as if the pyrotechnic column exploded after muzzle exit.

The modes of projectile failure examined in the initial stress study were:

a. The shear deformation or out-of-roundness occurring in the projectile jacket.

b. The stress field encountered by the projectile jacket after engraving and during acceleration of the projectile.

Shortly after the initiation of the stress study, DA guidance was received eliminating the 6mm concept from inclusion as a SAW contender. Developmental efforts were redirected towards the consideration of a 5.56mm SAW ammunition contender, which was easily included in the analytic study. Shown in Table 2 are the pertinent projectile characteristics for the 5.56mm concepts under development. In selecting an ammunition design as a SAW contender, several design criteria were applied to the analysis in order to define the use of the projectile and weapon barrel in a light machine-gun role. These design criteria are outlined in Table 3. In addition to these design parameters addressing projectile integrity, any interior bore configuration must satisfy other basic performance requirements such as projectile accuracy, barrel life under machine-gun firing schedules, interior ballistics, terminal effectiveness and high rate manufacture by current methods.

The effect of shear deformation on the projectile integrity was considered by applying thin-ring theory to the projectile jacket with "n" distributed forces being applied corresponding to the number of lands. The results of the analysis indicated that during the engraving process it is desirous that the pressure under the land be as large as possible for any given deflection. The reason for this is that the engraving is caused by the jacket material becoming plastic, and the smaller the deflection that is encountered when the material goes plastic, then the less out-of-roundness that will be incurred by the jacket. When considering this result relative to the

TABLE 2

5.56MM PROJECTILE CHARACTERISTICS

PROJECTILE TYPE	BULLET DIAMETER	WALL THICKNESS	JACKET COMPOSITION
BALL BULLET AP-1	.2245 - .0006	.021 - .002	GILDING METAL
IMPROVED TRACER BULLET (UNDERSIZED)	.2238 - .0005	.027 - .002	GILDING METAL CLAD STEEL (15-80-5%)
IMPROVED TRACER BULLET (BALL SIZE)	.2245 - .0006	.027 - .002	GILDING METAL CLAD STEEL (15-80-5%)

TABLE 3

5.56MM (SAW) AMMUNITION/WEAPON INTERFACE

DESIGN CRITERIA CONSIDERED IN STRESS ANALYSIS

1. DESIGN BORE DIMENSIONS SUCH THAT BALL AND TRACER ROUNDS OF DIFFERING DIAMETERS (TRACER BEING UNDER BALL DIAMETER) CAN FUNCTION IN THE SAW WEAPON.

2. ALLOW MAXIMUM LAND AND GROOVE DIAMETRICAL TOLERANCE OF $\pm$ .0005 IN.

3. DESIGN FOR DIAMETRICAL BORE GROWTH AT A TEMPERATURE OF $1250^{\circ}F$.

4. DESIGN TO ACCOMMODATE 0.0005 IN. MAXIMUM DIAMETRICAL LAND WEAR.

5. DESIGN OF 0.0004 IN. AVERAGE MANUFACTURING ECCENTRICITY BETWEEN LAND AND GROOVE DIAMETERS.

pressures and deflections induced by four and six-groove barrels, the results clearly indicate that the six-groove configuration is clearly superior to the four-groove even when comparing a six-groove barrel with minimum land height to a four-groove with a maximum land height.

The stress field developed on the jacket after engraving and during acceleration was addressed by considering a pressure gradient acting from the bottom to the top of the engraved surface. By relating this pressure distribution to the depth of engraving, minimum values of engraving depth were calculated such that the probability of jacket shearing is reduced. This minimum depth of engraving was shown to be .0017 in. for the four-groove barrel and .0011 in.for the six-groove. These minimum engraving depths were applied to the analysis in determing optimum bore configurations.

Optimum Bore Dimensions and Projectile Compatibility. When considering the minimum engraving depths required together with the pertinent design criteria and projectile dimensions, it is possible to compute optimum rifling dimensions such that the types of system failures considered will be minimized. This was done for the projectiles being developed by relating the minimum engraving depths required such that jacket shear does not take place as a function of projectile diameter, bore diameter, barrel temperature, jacket deformation due to engraving and land wear. This relationship is shown in equation 1-1.

$$(1\text{-}1) \qquad le = Rp - Rbo\ (1 + \alpha\Delta T) - W_b - u_{Ly}$$

where,

le	=	minimum engraving depth required
Rbo	=	bore radius or land radius
Rp	=	projectile radius
α	=	coefficient of thermal expansion
ΔT	=	barrel temperature gradient under hot condition
W_b	=	barrel wear
u_{Ly}	=	jacket displacement before yielding

By solving equation 1-1 for Rbo, the land diameter suited to each projectile design can be found. The optimum groove size was derived such that the smallest projectile diameter used in the bore will have the same diameter as the groove at its highest temperature as shown in equation 1-2. This would correspond to the barrel temperature reached under sustained firing schedules.

(1-2) $$D_G = \frac{Dp\ min}{1 + \alpha\Delta T}$$, where D_G = groove diameter
Dp = minimum projectile diameter
α = coefficient of thermal expansion
ΔT = barrel temperature gradient

The optimum barrel dimensions calculated using equations 1-1 and 1-2 are shown in Table 4. Note that configurations 1 and 2 are optimum based on tracer projectiles of differing diameters while configuration 3 considers an increased land height for larger barrel wear over configurations 1 and 2. Standard 5.56mm barrel dimensions are shown as reference.

A numerical exercise was performed utilizing the optimum rifling dimensions and projectile dimensions to demonstrate the range of in-bore interferences and clearances possible under "best" and "worst" design conditions. Table 5 summarizes the results of this exercise giving a range of interference/clearance values for both standard 5.56mm bore configuration and optimized configurations. To properly compute these interference/clearance values, the following parameters were considered:

a. minimum and maximum bullet diameters (ball and tracer)
b. minimum and maximum land and groove diameters
c. .0005 in. diametrical land wear
d. diametrical bore expansion at 1250^{o}F

Table 6 lists the equations used to compute the ranges of interference/clearance and minimum land height values. In comparing the standard barrel designs with the optimized cases, it is important to view these results in a strictly statistical sense in that projectile deformation into the barrel grooves was not considered. However, despite the rather static condition under which these numbers were generated, a major difference among designs can be noted. In all cases, the optimized designs exhibit a greater projectile/barrel interference, or lesser projectile/barrel clearance than the standard barrel dimensions. This important difference is the direct result of attempting to accommodate differing ball and tracer projectile diameters while insuring satisfactory system performance over a temperature range from

TABLE 4

5.56MM (SAW) AMMUNITION/WEAPON INTERFACE

RIFLING CONFIGURATIONS FOR BULLET/BARREL INTERFACE STUDY

BARREL TYPE	GROOVE WIDTH	GROOVE DIAMETER	LAND DIAMETER
STANDARD 5.56MM (REF)	.074 - .076	.2235 - .2245	.2190 - .2200
CONFIGURATION 1 6-GROOVE FOR UNDERSIZED TRACER	.074 - .076	.2206 - .2216	.2169 - .2176
CONFIGURATION 2 6-GROOVE FOR BALL SIZE TRACER	.074 - .076	.2212 - .2222	.2174 - .2181
CONFIGURATION 3 6-GROOVE WITH INCREASED LAND HEIGHT	.074 - .076	.2206 - .2216	.2156 - .2166

NOTE: SIX LANDS AND GROOVES EQUALLY SPACED

TABLE 5
5.56MM (SAW AMMUNITION/WEAPON INTERFACE)

TABLE OF INTERFERENCES UNDER MINIMUM AND MAXIMUM DESIGN CONDITIONS

	PROJECTILE CONFIGURATION	RANGE OF INTERFERENCE FOR MAX/MIN LAND CONDITIONS	RANGE OF INTERFERENCE FOR MAX/MIN GROOVE CONDITIONS	MINIMUM LAND HEIGHT POSSIBLE
STANDARD 5.56MM BORE	5.56MM (BALL)	+.0008 to +.0028	-.0013 to +.0005	.0015
	5.56MM (TRACER) (UNDERSIZED)	+.0005 to +.0024	-.0016 to +.0002	.0015
	5.56MM (TRACER) (BALL SIZE)	+.0007 to +.0028	-.0013 to +.0005	.0015
CONFIGURATION 1	5.56MM (BALL)	+.0019 to +.0038	+.00015 to +.00195	.00125
	5.56MM (TRACER (UNDERSIZED)	+.0017 to +.0035	-.0001 to +.0016	.00125
CONFIGURATION 2	5.56MM (BALL)	+.0017 to +.0035	-.0015 to +.0017	.0013
	5.56MM (TRACER) (BALL SIZE)	+.0017 to +.0035	-.00015 to +.0017	.0013
CONFIGURATION 3	5.56MM (BALL)	+.0024 to +.0045	+.00015 to +.002	.0018
	5.56MM (TRACER) (UNDERSIZED)	+.0022 to +.0041	-.00015 to +.0016	.0018
	5.56MM (TRACER) (BALL SIZE)	+.0024 to .0045	+.00015 to +.002	.0018

NOTES:

1. MINIMUM ENGRAVING REQUIRED FOR SIX-GROOVE CONFIGURATION IS .0011.
2. MAXIMUM INTERFERENCE CONSIDERS NO WEAR AND NO THERMAL EXPANSION.
3. BORE DIMENSIONS INCREASE BY .0019 IN. ON DIAMETER DUE TO THERMAL EXPANSION.
4. LAND DIMENSIONS INCREASE BY 0.0005 IN. ON DIAMETER DUE TO WEAR.

TABLE 6

5.56MM (SAW) AMMUNITION/WEAPON INTERFACE

METHODS OF COMPUTATION FOR INTERFERENCES AND CLEARANCES

$$\text{MAXIMUM LAND INTERFERENCE} = \frac{\text{DPMAX} - \text{DLMIN}}{2}$$

$$\text{MINIMUM LAND INTERFERENCE} = \frac{\text{DPMIN} - (\text{DLMAX} + .0005 + .0019)}{2}$$

$$\text{MAXIMUM GROOVE INTERFERENCE} = \frac{\text{DPMAX} - \text{DGMIN}}{2}$$

$$\text{MINIMUM GROOVE INTERFERENCE} = \frac{\text{DPMIN} - (\text{DGMAX} + .0019)}{2}$$

$$\text{MINIMUM LAND HEIGHT POSSIBLE} = \frac{(\text{DGMIN} + .0019) - (\text{DLMAX} + .0005 + .0019)}{2}$$

NOTE:

DPMIN = MINIMUM PROJECTILE DIAMETER

DPMAX = MAXIMUM PROJECTILE DIAMETER

DLMIN = MINIMUM LAND DIAMETER

DLMAX = MAXIMUM LAND DIAMETER

DGMIN = MINIMUM GROOVE DIAMETER

DGMAX = MAXIMUM GROOVE DIAMETER

NEGATIVE INTERFERENCES DENOTE CLEARANCES

ambient to 1250°F. These design parameters are further aggravated by considering land wear.

Comparing the interferences and clearances shown in Table 5 with the minimum required land engagement of .0011 in. for six-groove configurations shows possible problem areas. Despite the fact that the minimum land heights under worst conditions exceed this .0011 in. requirement, it is not necessarily true that proper engraving will occur. This situation occurs in the 5.56mm standard six-groove design, for both ball and tracer comparisons. Although the minimum land height at 1250°F is adequate for the required .0011 in. engraving, this engraving cannot occur if the projectile/land interferences run as low as .0005 in., as it does for the tracer. This minimal interference could lead to a serious skidding problem.

Experimental Evaluation. The accuracy of the analysis, as well as the suitability of any barrel design to field use, can only be verified through extensive testing. Toward this end, a quantity of barrels of various configurations has been procured for evaluation of system performance levels. Table 7 is a matrix showing the quantity and types of barrels which will be the core of an exhaustive barrel performance program. These barrels will be tested along with approximately 45,000 rounds of 5.56mm ball and tracer ammunition against current SAW performance requirements so that sufficient statistical significance is obtained, pointing to a singular rifling configuration.

Plans for testing currently envision adhering to current acceptance standards for 5.56mm and 7.62mm ammunition and will mirror sample sizes of barrels and ammunition contained therein.

TABLE 7

5.56MM (SAW) AMMUNITION/WEAPON INTERFACE

BARREL MATRIX

BARREL TYPE / BORE CONFIGURATION	ACCURACY	PRESSURE	WEAPON* (CHROMED)	WEAPON* (UNCHROMI
		QUANTITY		
STANDARD 5.56MM RIFLING	2	2	3	2
6-GROOVE BORE 1 IN 12 TWIST UNDERSIZED TRACER (CONFIG. 1)	2	2	3	2
6-GROOVE BORE 1 IN 11 TWIST UNDERSIZED TRACER (CONFIG. 1)	2	2	3	2
6-GROOVE BORE 1 IN 12 TWIST BALL SIZE TRACER (CONFIG. 2)	2	2	3	2
6-GROOVE BORE 1 IN 11 TWIST BALL SIZE TRACER (CONFIG. 2)	2	2	3	2
6-GROOVE BORE 1 IN 11 TWIST INCREASED LAND HEIGHT FOR ECCENTRICITY (CONFIG. 3)	2	2	3	2

DESIGN OF EXPERIMENTS DEALING WITH MAN-MACHINE INTERFACE IN CURRENT COMMUNICATIONS SYSTEMS

R. J. D'Accardi and H. S. Bennett, U.S. Electronics Command, Fort Monmouth, New Jersey

J. R. Hennessy, U.S. Army MERDC, Fort Belvoir, Virginia

ABSTRACT. Recently, the US Army Electronics Command has supported experiments dealing with man-machine interface problems occurring in Tactical Communications Systems. The aim was to characterize communications system operators' performance under various environmental conditions related to tactical operations. The study was directed towards system equipment such as the standard teletype and optical-read-only terminal equipments. Using these devices, the significance of acoustic noise and ambient light on operator performance was studied under sixteen combinations of environmental conditions.

The object of this presentation is threefold. First, we discuss the methods of evaluating message transfer over man-machine interfaces to include audio and visual. Second, we discuss the design of the experiment and modeling to determine the operator characteristics under different environmental conditions, and third, we present statistical estimates of: (a) the effects of the controlled variables (ambient light and acoustic noise) upon the transcription accuracy of several operators, (b) measures of experimental error to define a range of values, for a prescribed level of confidence, within which the true value of the estimates may be found, and (c) the most significant combinations of environmental effects on operator performance. Several multivariate regression models which characterize operator performance are presented and the criteria for choosing the best model are discussed.

INTRODUCTION. Information gained in evaluating and solving man-machine interface problems that occur in complex communications systems is extremely important to systems engineers committed to the mission of the design and fabrication of future generations of equipment. Sophisticated systems of Command and Control, computer-aided man-in-the-loop systems (e.g., manned space craft), human response to audio and visual displays, management functions, pattern recognition, man-computer languages, cutaneous communication and many other facets are of concern where an operator must perform a control task, or decision task. At present there is a large volume of on-going work oriented towards man-machine interfaces which span the projected needs of the Armed Forces. For example, work in progress by the Naval Electronics Systems Command, 6570th Aerospace Medical Research Laboratory, DA ARI for the Behavioral Sciences, ECOM and HEL (to name a few) generally deal with evaluation of complex system interfaces, assessment of operator performance capabilities for a wide variety of tasks, analysis of manual functions into tasks, analysis of human control functions, and the physical and psychological characteristics which affect the assessment of operator performance capabilities. Much of the on-going work concerns the psychological and physiological aspects of command and control in tactical operations, weapons systems, vehicles management, logistics, and communications. Some of the more

specific areas of investigation are:

1. Work/rest schedules and effects on man-machine performance.

2. Utilization of Bio-electric phenomena to automatically control complex systems.

3. Measures of operator performance under different mixes of equipment, personnel and procedures.

4. Physiological aspects (fatigue, alertness, metabolism, endocrine gland functions, and central nervous system) of operator efficiency and man-machine interface.

5. System simulation to study the impact of operator performance on complex systems as a function of environmental threat, mission, and work load stress.

6. Army Tactical Flight operations under adverse visibility conditions.

7. Influence of USAF operational environments on air crew utilization.

Examination of ongoing research in these areas indicate that there is no clear cut procedure to evaluate the human subsystem in a sophisticated communications system or the effects of environmental stress on operator performance. Army communications requirements in a tactical situation often require 24 hour operations and personnel are required to work either on standard or unpatterned and frequently extended duty schedules, in a variety of environments, each characterized by multiple stresses occurring in a random manner. For example, the accuracy in reading an optical display is dependent on many variables such as number of lines, characters, ambient lighting, environmental noise, speed of display, correction time, back-log, operator physiology (e.g., mood, fatigue, attention, and training), display brightness and size, and effective signal-to-noise ratio (legibility) to Name a few. Since future Army requirements include optical display terminals, it is essential to provide insight into those variables that affect accuracy through the man-machine interface and the effects caused by physiological factors. To answer the Army's need for measures of man-machine interfaces which occur in communications systems and to enhance the design of future families of equipment, this report will address teletype operator performance as the environmental factors of ambient light and acoustic noise are varied. The design of the experiment performed at Ft. Monmouth, New Jersey during April and May 1975 and results are discussed. Experimental results and several models are presented which show the significance of these variables on experienced teletype operators.

DESIGN OF THE EXPERIMENT. The significance of acoustic noise and ambient light on operator performance was investigated using a visual display transmission device, see figure 1. This is a visual terminal designed to interface with computers or store-and-forward devices. Primarily, it is a developmental equipment intended to visually present messages on a CRT display where an operator can see and correct his message prior to transmission. The advantages of this equipment over the standard military teletypewriter were not addressed in this experiment.

The experiment consisted of testing the transcription accuracy of six experienced communications-center operators under 16 combinations of environmental conditions. Ambient light was varied at four levels, ranging from 24 ft-candles to 3 ft-candles, and acoustic noise was concurrently varied at four levels ranging from 55 dBa to 95 dBa. Sound pressure level (SPL) measured in dBa is in reference to .0002 dynes/cm^2. This is considered the threshhold of hearing and is roughly equivalent to a leaf "falling" on a quiet day. The 55dBa level was considered the quiet condition where only the inherent noise from the terminal equipment, sound room noise, and thermal noise were recorded. The 95dBa level represented an extremely annoying and distracting "pink" noise. The noise-power per unit frequency for this type of noise is inversely proportioned to frequency over a specified range and slopes down at 3dB per octave from 20Hz to 20KHz. These characteristics are more common to conference type noise where the higher and lower frequency components characterize motor and equipment noises. Pink noise was also used because it has relatively constant energy per octave-bandwidth. The 24 ft-candle light level compared favorably to the Army Corps of Engineers standard for office lighting. The other chosen levels of 12, 6 and 3 ft-candles, respectively, represented successively deteriorating ambient light conditions. Throughout the testing, the brightness of the visual display was constant.

For each test the operator was required to type his name, treatment combination, and date as part of the message, see figure 2. The messages for the experiment consisted of forty random-letter word groups of five characters each. They were derived through a random number generator and an alphanumeric conversion. No message was a duplicate nor were they duplicated by any of the operators on either terminal equipment. The random letter format was used so that the operator could not identify or recognize message words and therefore would have to concentrate on the given formats to avoid making transcription errors. The aim of the experiment was to vary the environmental variables and to observe the accuracy and speed of transcribing the random letter formats as a function of these variables. The response variable, accuracy, was the measure of transcription errors that each operator committed per message format. The errors considered were the following:

1. transposition
2. missing letter
3. extra letter

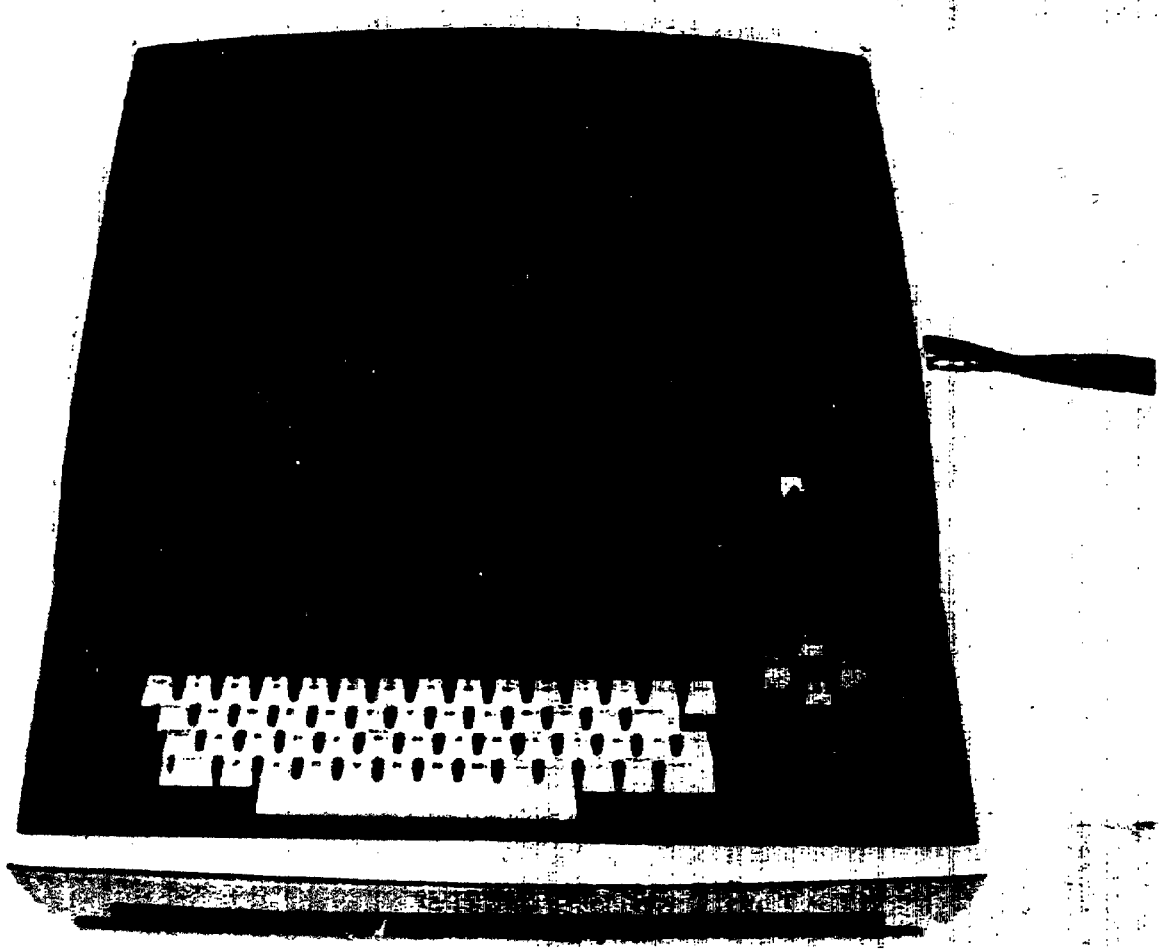


Figure 1 - Visual Display Transmission Terminal

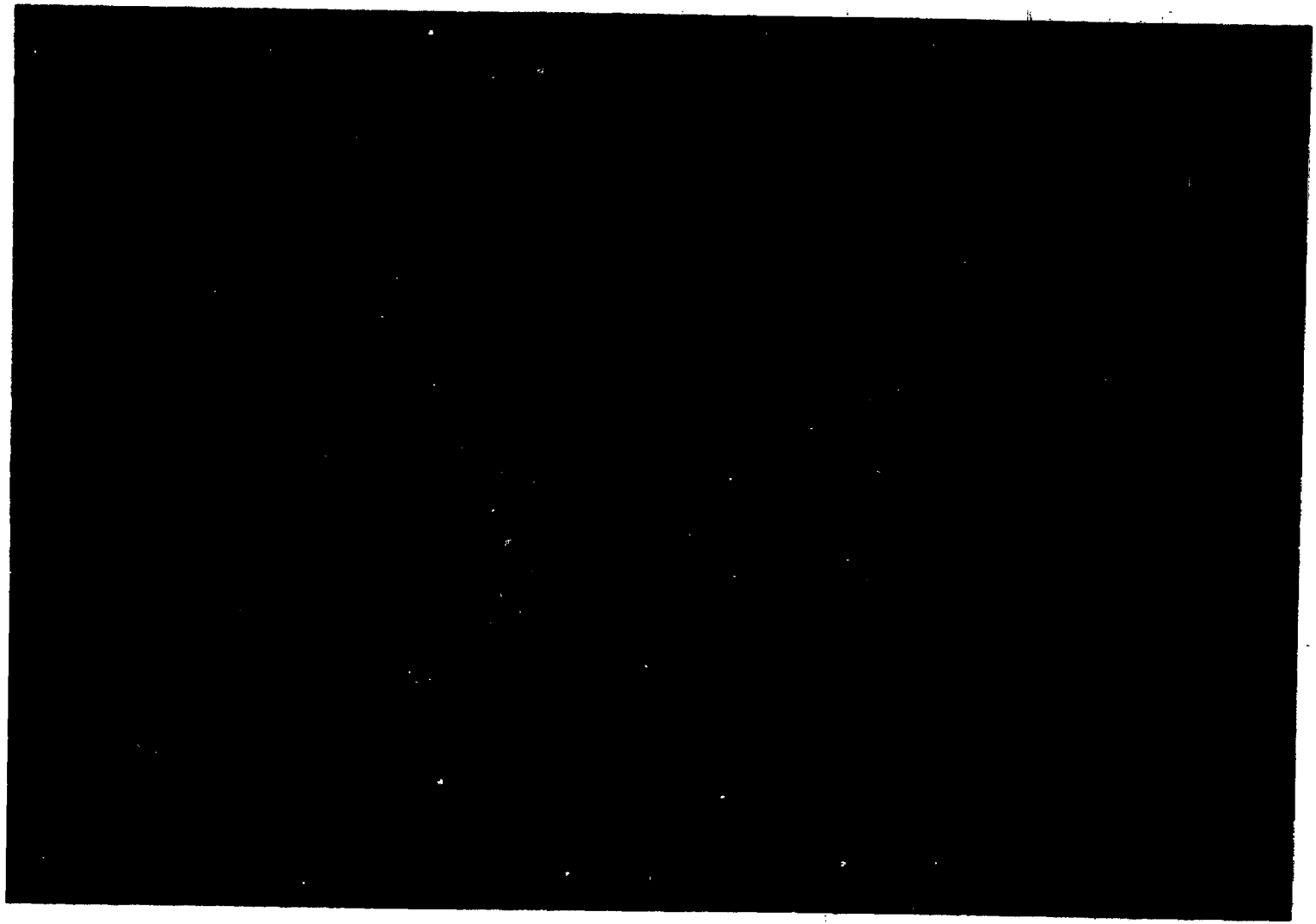


Figure 2 - Message Format

4. incorrect space

5. extra line feed

6. missing word groups

7. wrong letter

8. line out of sequence (skipped line inserted after detection)

9. word group out of sequence

The results were compared to an acceptable operator norm, i.e., typing a message format on a standard teletype terminal (see figure 3) under the same conditions. Each operator was tested in four sessions, each session programmed for eight random environmental combinations, four for each terminal equipment, where tests were alternated between the optical display and the standard teletypewriter. This was done to reduce the effects of learning. A thirty minute familiarization period was given each operator prior to the tests, and a standard instruction sheet was distributed during this period to insure uniform orientation with the equipment and with the purpose and procedure of the experiment.

The effect of any environmental combination is considered to be the sum of three effects, namely, those of sound, light, and the interaction of light and sound. To adequately analyze these effects, a two-level factorial experiment was formulated with six replications. The four levels of acoustic noise are combined with the four levels of ambient light giving 4 x 4 or sixteen treatment combinations. For a two-factor factorial experiment with n observations per cell, run as a completely randomized design, [1] , [2] , a general model is:

$$Y_{ijk} = \mu + A_i + B_j + A_iB_j + \varepsilon_{k(ij)}$$

where Y is the response variable, i.e., the number of transcribed errors, and A and B are the main effects of light and sound, AB is their interaction, ε is the experimental error, (i.e., the extent to which the observed data and the general model disagree) and their respective levels are i = 1,2,3,4; j = 1.2.3,4, with k = 1,2----6 observations per cell. The interaction term adjusts for the failure of either one of the main effects to remain constant for each level of the other. The test runs were randomized as shown in table I. This was done to minimize the effects of training.

Figure 3 - Teletypewriter Terminal

TABLE 1

TREATMENT SCHEDULE PER OPERATOR

Session	Run	Environmental Treatment Combinations: Optical Display Terminal	Environmental Treatment Combinations: Teletype Terminal
I	1	1,4	3,1
	2	4,3	4,4
	3	3,2	2,2
	4	2,1	1,3
II	5	3,1	4,1
	6	4,4	1,2
	7	2,2	3,4
	8	1,3	2,3
III	9	4,1	2,4
	10	1,2	3,3
	11	3,4	1,1
	12	2,3	4,2
IV	13	2,4	1,4
	14	3,3	4,3
	15	1,1	3,2
	16	4,2	2,1

(Treatment = (Ambient Light Level, Acoustic Noise Level)

Ambient Light Level	Value
1	24 ft-candles
2	12 ft-candles
3	6 ft-candles
4	3 ft-candles

Acoustic Noise Level	Value
1	55 dBa
2	70 dBa
3	80 dBa
4	95 dBa

ANALYSIS: The following ANOVA tables and statistical estimates were formulated to analyze the transcribed errors for the standard teletype terminal and for the optical display terminal (tables II, III, IV and V):

TABLE II

ANOVA FOR STANDARD TELETYPE TERMINAL

Source	Sum or Squares	Degrees of Freedom	Mean Square Error	"F" ratio
Ambient Light, A_i	55.94	3	18.65	0.33
Acoustic Noise, B_j	99.70	3	33.23	0.59
Interaction, A_iB_j	109.93	9	12.21	0.22
Error, $E_{k(ij)}$	4494.67	80	56.18	-----
TOTAL	4760.24	95		

TABLE III

ANOVA FOR THE OPTICAL DISPLAY TERMINAL

Source	Sum of Squares	Degrees of Freedom	Mean Square Error	"F" ratio
Ambient Light	65.28	3	21.76	0.32
Acoustic Noise	276.03	3	92.01	1.35
Interaction	55.18	9	6.13	.10
Error	5437.50	80	67.97	
TOTAL	5840.12	95		

TABLE IV

STATISTICAL ESTIMATES OF TRANSCRIBED ERRORS
FOR THE TELETYPE TERMINAL

Ambient Light Level	Statistic	Acoustic Noise Level 50dBa	70 dBa	80 dBa	95 dBa	For All Sound Levels
24 ft-candles	$\bar{Y}$	3.0	5.8	5.8	6.2	5.7
	S_Y	1.87	3.96	3.7	6.42	4.23
	$S_{\bar{Y}}$	0.84	1.77	1.66	2.87	0.95
12 ft-candles	$\bar{Y}$	2.2	6.8	6.8	9.8	6.4
	S_Y	2.17	2.59	5.54	8.47	5.63
	$S_{\bar{Y}}$	0.97	1.16	2.48	3.79	2.6
6 ft-candles	$\bar{Y}$	5.0	3.8	5.0	7.2	5.25
	S_Y	3.94	2.59	6.2	4.6	4.34
	$S_{\bar{Y}}$	1.76	1.16	2.77	2.06	0.97
3 ft-candles	$\bar{Y}$	4.4	4.0	3.8	4.2	4.10
	S_Y	3.36	4.95	3.03	1.79	3.19
	$S_{\bar{Y}}$	1.50	2.21	1.36	0.80	0.71
						Overall
For All Light Levels	$\bar{Y}$	4.15	5.10	5.35	6.85	5.36
	S_Y	3.30	3.60	4.55	5.76	4.43
	$S_{\bar{Y}}$	0.74	0.80	1.02	1.29	0.50

TABLE V

STATISTICAL ESTIMATES OF TRANSCRIBED ERRORS FOR THE VISUAL DISPLAY TERMINAL

Ambient Light Level	Statistic	Acoustic Noise Level 55 dBa	70 dBa	80 dBa	95 dBa	For All Sound Levels
24 ft-candles	$\bar{Y}$	3.4	5.80	6.20	9.2	6.15
	S_Y	2.7	4.76	5.17	4.82	4.61
	$S_{\bar{Y}}$	1.21	2.13	2.31	2.15	1.03
12 ft-candles	$\bar{Y}$	6.8	5.0	7.0	8.60	6.9
	S_Y	3.77	2.77	2.45	6.35	3.99
	$S_{\bar{Y}}$	1.69	1.24	1.10	2.84	0.89
6 ft-candles	$\bar{Y}$	5.0	5.2	6.2	5.5	5.46
	S_Y	3.16	2.39	3.96	4.16	3.28
	$S_{\bar{Y}}$	1.41	1.07	1.77	1.86	0.73
3 ft-candles	$\bar{Y}$	6.0	5.2	5.4	8.2	6.2
	S_Y	3.67	3.42	5.5	4.71	4.23
	$S_{\bar{Y}}$	1.64	1.53	2.46	2.11	0.94
						Overall
For All Light Levels	$\bar{Y}$	5.3	5.35	6.20	7.90	6.19
	S_Y	3.34	3.18	4.11	4.87	4.00
	$S_{\bar{Y}}$	0.75	0.71	0.92	1.09	0.45

Although one might expect that acoustic noise and ambient light would strongly affect the production of transcription errors, no conclusive statistical significance as to environmental effects can be adjudged from the data. Examination of the MSE, however, shows that acoustic noise has a stronger effect on error production than either the Ambient Light or the interaction of the two (see tables II and III). Table IV and V show, for all light levels, the average transcription error production increased by about 60%. For all sound levels, the transcription error did not vary significantly.

The operators chosen were all of the same minimum proficiency, each able to transcribe messages at 60 w.p.m., with the exception of one trainee. Thus, examining the variation of transcription errors for the visual display terminal at 70 dBa (see table V) for light levels below 24-ft candles, the mean $\bar{Y}$ and standard deviation, S_y, decrease from the 55 dBa values, then increase as noise is increased to 95 dBa.

Interviews with the subjects seem to indicate that 70 dBa is the approximate level of noise to which they are accustomed, and therefore they were less distracted by environmental changes in ambient light at this sound level. The findings indicate that for the visual display terminal under quiet conditions (i.e., at 55 dBa, the noise below standard comcenter operational levels) at lower levels of Ambient Light, more errors were made than at normal operating (70dBa) level. The effect of noise at the higher levels (80 and 95 dBa) indicates the variability and adaptability of the operators to acoustic and photic noise. It was also noted (as was expected with the visual display terminal) that changing light levels had the least effect on operator performance.

Six multiple linear and non-linear regression models were fitted to the data, by the least squares method, to characterize operator performance. The models were of the form:

$$(1)\quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_{12}$$

$$(2)\quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon_{12}$$

$$(3)\quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 + \varepsilon_{12}$$

$$(4)\quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1^3 + \beta_6 X_2^3 + \beta_7 X_1 X_2 + \beta_8 X_1^2 X_2 + \beta_9 X_1 X_2^2 + \varepsilon_{12}$$

$$(5)\quad Y = \sum \beta_i X_1^j X_2^k + \varepsilon_i \qquad 0 \leq j+k \leq 3$$

$$(6)\quad Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 X_2 + \beta_3 \ln^2 X_1 + \beta_4 X_2^2 + \beta_5 X_2 \ln X_1 + \varepsilon_{12}$$

Where Y is the observed operator response, X_1 and X_2 are independent variables corresponding to ambient light and acoustic noise respectively. The estimated values of the coefficients, standard errors of the estimates, and coefficients of determination are summarized in the following table:

TABLE VI

Least Squares Estimates Using Coded and Uncoded Data for the Optical Display Terminal

Estimate	Model 1	2	3	4	5 Uncoded	6 Uncoded
$\hat{\beta}_0$	7.078	7.078	6.785	6.684	21.049	13.715
$\hat{\beta}_1$	0.100	.100	0.190	1.752	-3.495	-8.045
$\hat{\beta}_2$	0.680	.680	0.680	0.655	-0.362	-0.260
$\hat{\beta}_3$		.321	-0.099	0.449	0.084	24.986
$\hat{\beta}_4$			0.225	0.110	0.002	0.002
$\hat{\beta}_5$			0.320	0.225	0.004	-0.636
$\hat{\beta}_6$				-0.543		
$\hat{\beta}_7$				0.232		
$\hat{\beta}_8$				-0.076		
$\hat{\beta}_9$				-0.108		
$S_{\hat{\beta}_0}$	.307	.260	0.537	0.509	9.353	9.024
$S_{\hat{\beta}_1}$	.224	.190	0.260	1.279	3.246	14.836
$S_{\hat{\beta}_2}$	.210	.179	0.171	0.743	0.229	0.245
$S_{\hat{\beta}_3}$		.130	0.204	0.521	0.627	24.845
$S_{\hat{\beta}_4}$			0.133	0.166	0.001	0.001
$S_{\hat{\beta}_5}$			0.125	0.124	0.023	1.000
$S_{\hat{\beta}_6}$				0.481		
$S_{\hat{\beta}_7}$				0.131		
$S_{\hat{\beta}_8}$				0.091		
$S_{\hat{\beta}_9}$				0.181		
$S^2_{(Y-\hat{Y})}$	1.227	1.041	0.996	0.930	1.114	1.221
$R^2_{Y\hat{Y}}$	0.450	0.634	0.721	0.854	0.651	0.481

Clearly, the higher order model (4) fits the data best on the basis of minimum residual variance, $S^2_{(Y-\hat{Y})}$, and maximum coefficient of determination, $R^2_{Y\hat{Y}}$.

This provides the model:

$$\hat{Y}_1 = 6.785 + 1.752X_1 + 0.655X_2 + 0.449X_1^2.$$

$$+ 0.110X_2^2 + 0.225X_1^3 - 0.543X_2^3$$

$$+ 0.232X_1X_2 - 0.076X_1^2X_2 - 0.108X_1X_2^2$$

Testing for fit, the sum squared error due to regression and the respective degrees of freedom for the variation of Y_1 from the curve are 3.378 and {9,6} respectively. If the model is correct, the residual mean square has the expected value of σ^2. Using $S^2_{(y-\hat{y})} = \sigma^2_y = 0.5187 = MS_\varepsilon$, the "F" ratio is:

$$F = \frac{MSc}{MS_\varepsilon} = \frac{3.378}{0.518} = 3.907$$

and is not significant since $3.907 < 5.520$. Thus, on the basis of minimum $S^2_{(y-\hat{y})}$, maximum $R^2_{y\hat{y}}$ and this test, we have no reason to doubt the adequacy of this particular model. This technique is presented to show the feasibility of using multiple least squares regression for this type of man-machine interface problem. A more sophisticated approach is planned at a later time when more data is obtained.

Conclusions: Several adverse aspects of the terminal equipment were discovered which may affect error production. The angle of the keyboard (see figures 4 and 5) of the visual display terminal was apparently not conducive to optimum performance. The teletypewriter keyboard was unanimously considered more comfortable. Also, the detent pressure of the individual keys and the absence of feedback "thump" seemed to increase the probability of transcription error with the visual display terminal.

While the results do not show statistical significance of the environmental effects, the trends in the statistics (particularly the MSE and overall means, see tables II, III, IV and V) indicate the possibility that with a larger population of more homogeneous (as to expertise) subjects, statistical significance will emerge. That is, the variations in human performance will be greater under abnormal environmental conditions. If such abnormal conditions are to be expected under battlefield conditions, then significant training information could be extracted from such a follow-on experiment.

Figure 4 - Visual Display Terminal,
Keyboard Sideview

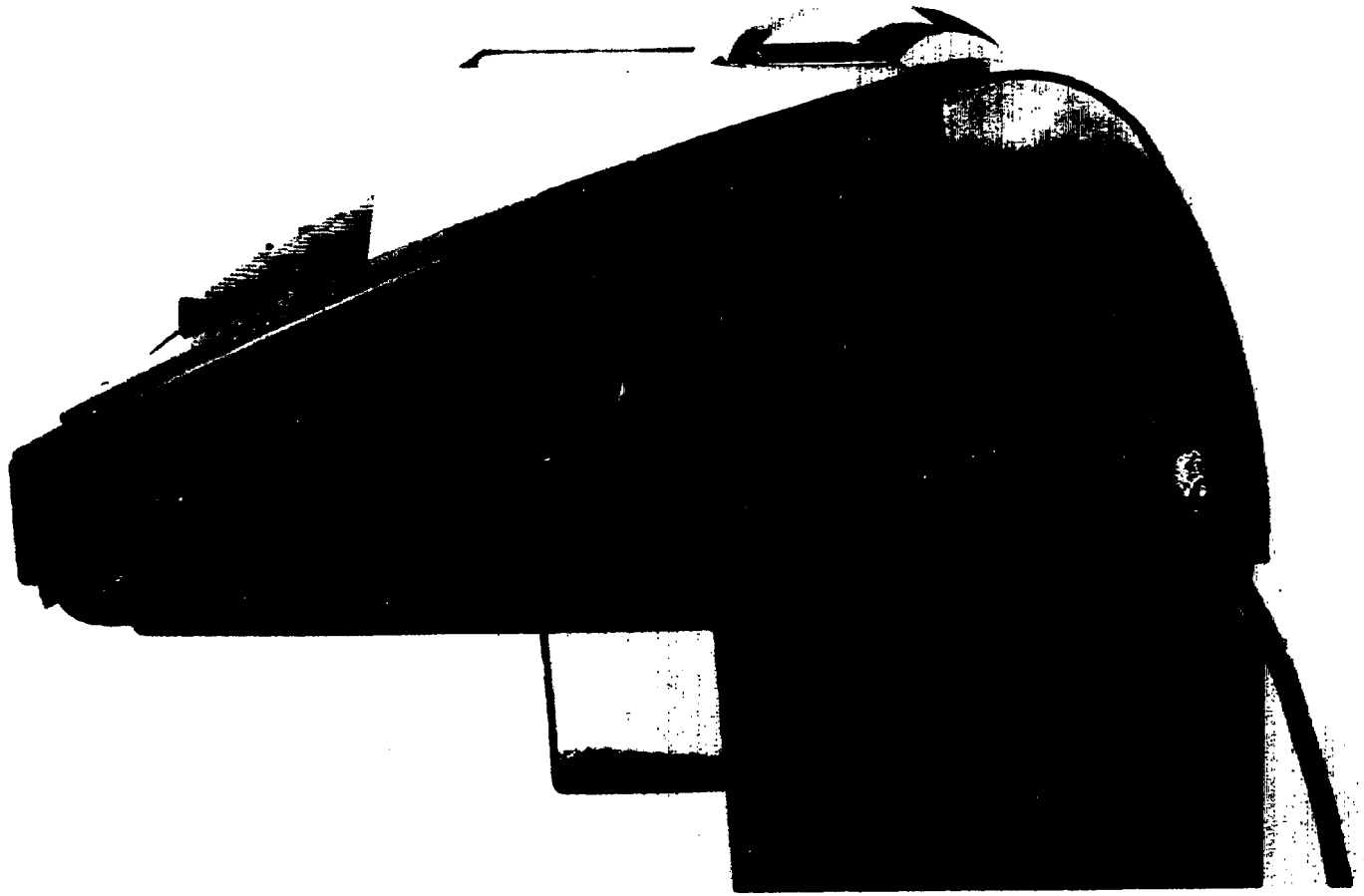


Figure 5 - Teletypewriter Terminal,
Keyboard Sideview

Another measure that could attain statistical significance is the mean transcription error production for the group. Such statistics would indicate the outer bounds of expectation under battlefield conditions.

REFERENCES

1. Hicks, C. R., "Fundamental Concepts in the Design of Experiments", Holt, Rinehart and Winston, New York, 1973.

2. Draper, N. R. and H. Smith, "Applied Regression Analysis", John Wiley & Sons, New York, 1966.

PLANNING FOR THE MEASUREMENT OF FLIGHT TRAJECTORY

J. B. GOSE
J. V. CARRILLO
Quality Assurance Office
US Army White Sands Missile Range
New Mexico

ABSTRACT. This paper describes a procedure used at White Sands Missile Range, New Mexico for selecting instruments to measure a test object's location and body angles. Criteria for selection include number and location of instruments, types and quality of measurements, probability of operation, and data reduction procedures. Optimizations are made in terms of cost-to-support, probability of success, expected error in data and instrumentation system used. Constraints include expected trajectory and object dimensions, optical image size and aspect angle, tracking rate, atmospheric distortion, and for some applications, locations of existing facilities.

The procedure employs both theoretically and pragmatically derived models and utilizes observed error distribution and reliability data. It has been automated for computation on a UNIVAC 1108 computer.

1. INTRODUCTION. The purpose of this report is to outline the mathematical and statistical scheme used for the Resource Conservation Planning (RCP) Model. The RCP is used as a tool for evaluating and formulating test support plans.[1] The model developed is formulated from the multi-station solution now in use at WSMR, better known as the Davis Solution.[2] This is a least-squares solution which is identical to the maximum likelihood estimates of missile position in the particular case in which the instrumentation measurements are normally distributed. In 1965, 1LT Charles A. Hall, PhD, expanded the least-squares formulation to provide an improved estimate and to minimize the number of observations required. This concept became known as Minimal Station Participation (MSPAR).[3] The RCP is an extension of this concept. The scheme has been

---

[1]J. V. Carrillo and R. L. Garcia, A Technique for Computing The Probability of Meeting a User's Trajectory Requirement, QA Technical Report No. 121, (WSMR, NM, 1975).

[2]R. C. Davis, Techniques for the Statistical Analysis of Cinetheodolite Data, (China Lake, California, 1951), page 1.

[3]C. A. Hall, Deleting Observations From a Least-Squares Solution, Proceeding of the Eleventh Conference on the Design Experiments in Army Research Development and Testing, ARD-D Rpt 66-2, (Durham, NC, 1966).

adapted to cinetheodolites, Telescopes, Radar, and DOVAP for position and attitude applications. The RCP model uses for input empirically developed measurement error probability tables from each measurement system, a proposed flight test trajectory of a specified test object, and the uncertainty (flight test requirements) in the flight test data that a Range User can tolerate in his experiment. The probability tables are used to compute the probability of a particular data error for a selected or given geometry configuration. The final output is in terms of the probability of meeting a particular Range User requirement. Hence, cost-to-support trade-offs can be developed based on the risk a user may want to take in completing his experiment. The less risk the user can accept, the higher the support cost.

Restating the problem as: "Determine the probability of satisfying a Range User's requirement for a test object's position and/or attitude over a given interval, such that the results will allow cost trade-off analyses."

The problem statement gives rise to the specific questions of how to identify the minimum set? How to find the probability of success? and How to solve the problem with a computer? The approach taken obviates the need to answer the first question (as we shall see). The latter two are the substance of this paper.

2. ESTIMATION OF THE PROBABILITY OF SUCCESS. Error estimates can be described probabilistically and, of course, reliabilities are probabilities. Thus, they can be combined in a probabilistic formulation. The probabilities involved in the estimation of meeting a requirement for one point of a trajectory can be expressed in equation form as:

$$P(\text{Rqmt})_i = \sum^{M} [P(\sigma_\theta^2 \leq S_x) \times P(\text{Sta Opr})]_i \qquad \text{(Eq 1)}$$

where,

$P(\text{Rqmt})_i$ = Probability of meeting the requirement at the ith point

σ_θ^2 = Error in observed data

S_x^2 = Maximum allowable error from the requirement

P(Sta Opr) = The probability of successful station operation

$$M = \Sigma\binom{x}{c}$$

where c = 2, 3, 4, ..., x
x = total number of sites available

The probability for the entire trajectory is the distribution of the chances for success at all points from the population of occurrences and is found by simply averaging the risk over all points:

$$\hat{P}(Rqmt) = \frac{\sum_{i=1}^{R} P(Rqmt)_i}{R} \qquad \text{(Eq 2)}$$

where

R = the number of trajectory points.

The only unknown parameter in Equation 1 is σ_θ^2. σ_θ^2 is found in the following manner.

The basic regression relationship is

$$\phi = B\theta$$

where,

ϕ = Matrix of Observations

B = Jacobian Matrix

θ = Matrix of Derived Trajectory Data

Solving for θ

$$\theta = (B^t W B)^{-1} B^t W \phi \qquad (W = \text{Weight Matrix})$$

$$\sigma_\theta^2 = \sigma_\phi^2 (B^t W B)^{-1}$$

or

$$\sigma_\theta^2 = (B^t W B)^{-1} \qquad \text{for} \qquad W = [\sigma_\phi^2]^{-1}$$

$$\sigma_\theta^2 = \sigma_\phi^2 (B^t B)^{-1} \qquad \text{(Eq 3)}$$

This last equation (Eq 3) defines the data error in terms of Geometric Dilution of Precision (GDOP) and measurement error; both of which are known or knowable. For a given geometry, $(B^t B)^{-1}$ is deterministic while σ_ϕ^2 is probabilistic. Thus, the probabilistic nature of σ_θ^2 is dependent on the probabilistic nature of σ_ϕ^2.

In actual practice, a requirement, S_x^2, is defined as the trace of a variance-covariance matrix. We may, therefore, attack the heuristic nature of σ_ϕ^2 simply by introducing a scalar "s^2".

$$s^2 = (S_x^2 / \sigma_\theta^2)_{TR}$$

into Eq 3, which becomes

$$s^2\sigma_\theta^2 = s^2\sigma_\phi^2(B^tB)^{-1}$$

The probability of measurement error (S_ϕ) $\leq s\sigma_\phi$ is the probability that $\sigma_\theta^2 \leq S_x^2$ (see example Figure 1). These data are available from histories of performance.

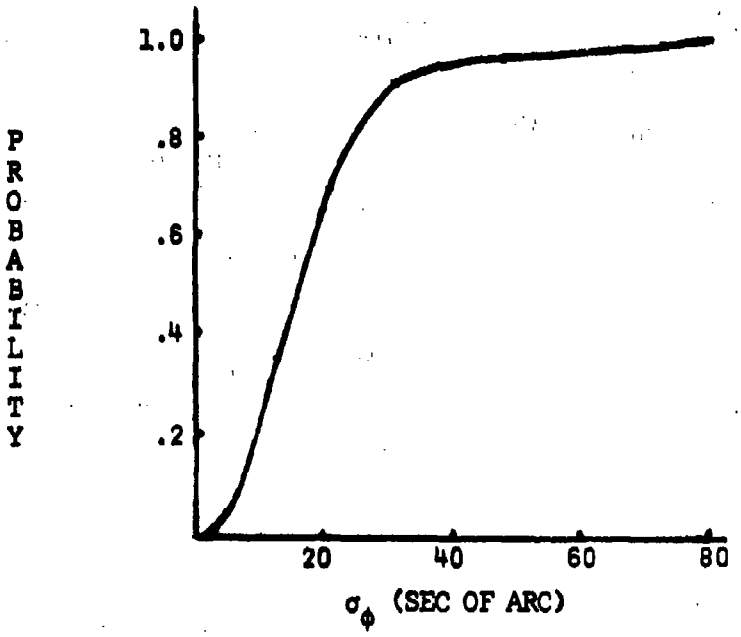


FIGURE 1

Equation 1 becomes

$$P(\text{Rqmt})_i = \sum^{M} [P(S_\phi \leq s\sigma_\phi) \times P(\text{Sta Opr})]_i \qquad \text{(Eq 4)}$$

The formula for computing the probability that exactly M of N scheduled instruments operate successfully is:

$$P(\text{Sta Opr}) = (R_1 \cdot R_2 \cdots R_M)(Q_{M+1} \cdot Q_{M+2} \cdots Q_N) + \cdots + (Q_1 \cdot Q_2 \cdots Q_{N-M})$$

$$(R_{N-M+1} \cdot R_{N-M+2} \cdots R_N) \qquad \text{(Eq 5)}$$

where: $R_1, R_2, \ldots, R_M$ are the reliability values for instrument 1, 2, 3, ..., M. $Q_1, Q_2, \ldots, Q_N$ are the $(1-R_1), (1-R_2), \ldots, (1-R_N)$ values for each of the instruments, respectively. Note that there are

$$\frac{N!}{M!(N-M)!}$$

terms to be added in Eq 5.

An example of the computational procedure for a point is shown in Appendix 1.

3. FITTING THE MODEL ON THE COMPUTER. A little thought on the computational times for Equation 5 will lead one to the realization that the time will approximately double for each additional site added. This was verified for the program prepared for the UNIVAC 1108 computer: A 5

station solution taking 2 seconds, 11 stations taking 1 minute, 15 stations taking 14 minutes, etc. Alternatives to minimize this problem were (1) to improve the speed of each computation or (2) to reduce the number of candidate sites. The latter course was pursued.

An initial screening was derived based on instruments operating limitations.

OPTICS - Elevation Angle - Between 3° and 80°

Image Size - >35 Microns (μ) for Position

>100 Microns (μ) for Attitude

RADAR & DOVAP - Elevation Angle - Between 10° and 80°

Next, each surviving site is ordered in accordance with its contribution to the error. For each point, an error constant[4] D_j is calculated from:

$$D_j = \sum_{KL} H^2_{jKL} \qquad \text{for the jth site}$$

K is an index of observation (ϕ)

L is an index of computed values (θ)
$L = 1, 2, 3, \ldots, \theta$

and

$$H_j = (B^t W B)^{-1} B_j{}^t W_j{}^{\frac{1}{2}}$$

W_j is a weight matrix from $\sigma_\phi{}^2 W = I$

from $\sigma_\theta{}^2 = \sigma_\phi{}^2 (B^t B)^{-1}$

$$\sigma_\theta{}^2 = (B^t W B)^{-1}$$

---

[4] c.f., Ref 1

D_j's then relate to σ_θ^2 from

$$\sigma_\theta^2 TR = \frac{\sum_{j=1}^{A} D_j}{L}$$

where,

A = The set of sites used

L = 3 for Position data
    2 for Attitude data

The D_j's vary with GDOP, therefore, the largest value at one point may be smaller than the smallest value at another point. Since all points are assumedly of equal importance to a customer, the GDOP effect (D_j's) must be normalized. This is accomplished by the following scheme. First, an average D_j is computed. This average value is divided into each D_j value for all points. Then, each site's normalized point value is summed over all points. The sites are then ordered (largest to smallest) based on the magnitude of the sum. The first three sites (with the largest values) are then selected for the first estimate of meeting a user's trajectory requirement. If the probability of meeting the requirement is sufficient, the computation is terminated. If the probability is insufficient, the site with the next largest value is added to the computation. This procedure is continued until the desired probability is obtained or all the sites in the group are used. This procedure has resulted in minimizing the number of sites required.

In evaluating the procedure, it was found that the sites selected produce the maximum $\hat{P}$(Rqmt) 95% of the time; and for the remaining 5%, the $\hat{P}$(Rqmt) was within 3% of the maximum.

6. CONCLUSIONS. The models discussed in this paper can be used for analyzing cost-to-support trade-offs. Cost-to-support is related directly to the type and amount of instrumentation necessary to meet a particular user requirement. Thus, the output of the RCP Model provides the information necessary for risk analysis from a measurement aspect. It is readily apparent that the more stringent the error requirement or the less risk of data loss a user can accept, the higher the cost-to-support.

There are limiations to the model. First, since the error and reliability values used are based on history, changing performance will result in erroneous answers; further, since the present reduction process is modeled in the equations, a change in the procedure will necessitate revision of the model.

APPENDIX 1

EXAMPLE OF COMPUTATIONAL SEQUENCE, $S_x = 5$

SITE COMBINATION	COMPUTED ERROR (S_θ)	(S_x/S_θ)	PROBABILITY OF OBSERVATION ERROR $S_\phi \leq s\sigma_\phi$	PROBABILITY THAT ONLY THESE SITES OUT OF 4 OPERATED	PROBABILITY OF MEETING REQUIREMENT
1, 2	22.8	0.22	0.059	0.0388	0.0023
2, 3	54.2	0.09	0.020	0.0388	0.0008
2, 4	10.2	0.49	0.398	0.0388	0.0154
1, 3	7.7	0.65	0.586	0.0388	0.0228
1, 4	4.5	1.11	0.919	0.0388	0.0357
3, 4	3.9	1.30	0.967	0.0388	0.0375
1, 2, 3	5.0	1.00	0.880	0.1050	0.0924
1, 2, 4	2.8	1.78	1.000^-	0.1050	0.1050
2, 3, 4	2.5	2.07	1.000^-	0.1050	0.1050
1, 3, 4	2.2	2.30	1.000^-	0.1050	0.1050
1, 2, 3, 4	2.0	2.50	1.000^-	0.2840	0.2840

$P(Rqmt)_i = 0.8060$

NON-RANDOMIZED CLINICAL TRIALS

E. A. Gehan and E. J. Freireich

The University of Texas System Cancer Center

Houston, Texas

ABSTRACT

This paper gives a general discussion of some principles involved in planning comparative studies, namely, the objectives, comparability of patients, feasibility, and ethics. For each principle, circumstances are given for which a non-randomized study is to be preferred to a randomized one. Examples of non-randomized, controlled studies are presented utilizing literature controls, an acute leukemia late intensification study involving matched controls, and an acute leukemia sequence of three studies. In the latter example, adjustment for prognostic factors was carried out to enable the studies to be compared with respect to response rate and survival.

NON-RANDOMIZED CLINICAL TRIALS

E.A. Gehan

and

E.J Freireich

The University of Texas System Cancer Center

1. Introduction

Consider the design of the following Army experiment (hypothetical). Because of the need for saving money, an officer in the Quartermaster Corps does a study of shoe sizes for Army recruits. He finds that the distribution of shoe sizes has several peaks and that it would be possible to save money in buying shoes by ordering only a small number of sizes. He decides that the best way to determine which sizes to buy is from a randomized comparative study. His idea is to issue three sizes of shoes: 8½, 9½ and 10½ randomly to incoming recruits and their "response" to a particular shoe will be measured following a ten mile hike by interviewing and a physician's examination. The ultimate objective is to choose a single size of shoe for all recruits. What is wrong with this experiment? The objective is stated clearly, the designed experiment could be carried out, treatments would be assigned at random and there wouldn't be much difficulty in measuring reaction of the recruits to the assigned shoes. It is obvious that the whole experiment is ridiculous because each individual has his own shoe size and a choice of shoes should be made accordingly. Randomization, in this case, added only a pseudo-scientific aspect to the experiment.

The outcome could be predicted well and a great deal of suffering would be caused among the Army recruits selected for the study - either by randomization or otherwise. In clinical research, treatment must often be tailored to the individual patient either in terms of dosage or schedule and a randomized comparative study is difficult to accomplish when treatment is individualized. Too often, randomized comparative clinical trials are analogous to the hypothetical Quartermaster who proposed a randomized comparison of shoes of different sizes.

In cancer clinical trials and in other disease entities, the patient is in a life or death struggle against his disease. His objective is to win the battle and he clearly would like to be in the hands of a physician who would give him the best chance of winning. Would the best chance be as a patient in a randomized comparative study or as an individual receiving care from an outstanding physician who used his best knowledge of patient, disease and treatment to choose a treatment plan? An analogy might be the selection of a designer for a car to win the Indianapolis 500 mile race. Would a designer be chosen who did a randomized comparative study of every design feature to be added to the car or would one choose an experienced designer with a good record and ask him to use his best judgment to design a car to win the race. Not many individuals would do randomized comparative studies in an attempt to win the Indianapolis 500; why then the emphasis on randomized comparative studies to win the battle against cancer or heart disease?

In this paper, a discussion will be given to the general considerations involved in planning a randomized vs. non-randomized comparative study and some specific examples of successful non-randomized studies will be given. These studies involve selection of control patients from the literature, from matched patients and from the previous study in a sequence of clinical studies. Recent papers stressing the value of non-randomized studies are by Gehan and Freireich (1974) and Freireich and Gehan (1974).

2. General Considerations

Four aspects of the comparative clinical trial will be considered. These are: (a) objectives; (b) comparability of patients; (c) feasibility; and (d) ethics.

(a) Objectives

Chalmers, Block and Lee (1972) have published a paper on controlled clinical trials in which the main theme is illustrated by a humorous conversation between two biostatisticians. First biostatistician, "How's your wife?". Second biostatistician, "Compared to whom?". The humor of this parable emphasizes two important and distinctive facts about the man's wife: the first being how does his wife differ from other wives, a comparative fact; the second, how is his wife in his own judgment, that is, what is his estimation of his wife's capabilities. This fundamental difference is frequently overlooked in the design and conduct of a clinical study. It should be emphasized that an important result of a therapeutic investigation is the measurement in a quantitative sense of the effectiveness of a given treatment. There are situations in which the important question is not how effective is this treatment, but is this treatment more or less effective than a standard or some other form of treatment. In general, the latter question is not as significant as the former - for both treatments and wives.

An essential ingredient of clinical research is a significant objective. Too often the concept of randomization is equated with the concept of research while non-randomization is equated with "non-scientific" or "uncontrolled". One cannot replace the intelligent, imaginative, creative work of a clinical scientist with the routine application of a clinical trial technique. In cancer research, there are many examples of non-randomized studies that have led to important alterations in methods of treating patients. Examples are the discovery of mechlorethamine in the treatment of Hodgkin's disease, the first antimetabolite methotrexa

in the treatment of patients with acute leukemia, vincristine in acute leukemia, and combination chemotherapy in lymphoma and Hodgkin's disease. These were all dramatic advances in the treatment of patients with malignant disease and this knowledge was derived from non-randomized clinical studies. What new and effective treatments have been discovered utilizing randomized clinical studies?

(b) Comparability of patients

As A.B. Hill (1962) has put it, a sine qua non in the proper conduct of a controlled clinical trial is having comparable groups of patients. A clinical trial designed to evaluate the relative effectiveness of two or more treatments should be planned so that the only differences among treatment groups are in the actual treatment received. This requires comparability of patients as they are entered into study, managed when on study, and analyzed when the study is completed.

The entry of patients will be discussed here and one technique for achieving comparability of patients is randomization, possibly stratified so that there are separate randomizations of patients in prognostic categories. Even the proponents of randomization agree that randomization guarantees comparability of patients <u>on the average</u> and this needs to be checked in every clinical trial. It may even be argued that randomization is a guarantee of non-comparability of treatment groups with respect to some patient characteristics, if enough patient characteristics are examined. For example, if there were a 5% chance that the random assignment of patients would lead to a significant difference between treatment groups with respect to a given patient characteristic and the distribution of 20 characteristics were considered, it would be expected that there would be a significant imbalance between groups with respect to at least one characteristic. As Daniel (1970) has pointed out, "Randomization is a confession of ignorance. Full randomization is a confession of full ignorance." In other words, a full randomization should be accomplished only when a clinical investigator is not cognizant of any patient characteristics that influence prognosis.

Another technique for achieving comparability of patients at time of entry into study is to select patients for a control group according to certain characteristics, namely those which are known to influence prognosis. If treatment A is the treatment under study and treatment B is a standard or "control" treatment which is to be compared with A, the control group of B patients could be selected from the literature, chosen on a matched basis from previously or concurrently conducted clinical studies, or selected from the previous study in a sequence. The primary assumption needed for selecting a control group is that the important patient characteristics related to prognosis are known, so that there is a firm basis for selecting a comparable group of patients. Further, it must be assumed that differences which do exist between the groups selected (such as time, institution, physician, or the availability of supportive care) have little or no relation to the outcome of the treatment. In a disease which has been studied extensively, techniques of regression analysis can be used to determine patient characteristics related to prognosis. See Armitage and Gehan (1974) for a review of available methods. Some examples will be discussed in section 3.

(c) Feasibility

In general, the feasibility of a particular study relates to the number of patients required and its duration. For a particular investigator or group of clinical investigators, one can compare the strategy of proceeding from one fairly large study to the next, each based on a single treatment vs. the strategy of randomizing between two treatments in each study. Suppose the investigators in both circumstances has exactly the same requirements concerning the number of patients to be studied on each treatment. Suppose the number required for each treatment is N and the group of investigators accrues this number of patients in one year. Assuming that no follow-up period is required for observing the effect

of treatment, the strategy of proceeding sequentially from one study to the next means that one year will be required for each study. The investigator who always randomizes between two treatments requires two years to complete each study. It is true that at the end of two years, an investigator following either strategy will have evaluated two treatments, however the investigator who does sequential studies will have an opportunity to choose a second treatment based upon the results of the first. Further, some investigators adopt the practice of always carrying along the best treatment from a previous study in the current study; this results in evaluating three treatments every four years compared with four treatments for the investigator who proceeds sequentially. The latter investigator will have had the opportunity to build upon knowledge gained from previous studies to choose three treatments, while the investigator preferring simultaneous comparisons will have chosen only one new treatment based upon the results of a previous study.

Suppose an investigator is doing a simultaneous comparison of treatments A and B in which a fixed number of patients is to receive each treatment so that the difference in response rates can be detected at a given significance level and power of test. These specifications lead to n patients being required on each treatment and tables of n are readily available in textbooks (Cochran and Cox, 1957) (Holland and Frei, 1973). An experimenter who does studies in sequence of one treatment might be prepared to assume that the response rate to the control treatment (B) is so well known that it may be taken as a fixed quantity, say p, and no patients need receive B in the trial. To carry out a statistical test of the difference between the proportion of patients responding to A and B at the same significance level and power assumed above, only n/2 patients are needed on treatment A, which is only 1/4 the total number of patients required for the randomized comparative trial. When the cost of supporting clinical studies is often

in excess of $1,000,000 per year, a savings of patients and duration of study has a substantial dollar equivalent. Even when the response rate to the control treatment is not known precisely, it may still be reasonable to proceed as if it is known. For example, in the treatment of patients with advanced lung cancer, the expected percentage of patients responding to standard treatment is very low (less than 20%) and survival is poor. In this circumstance, it would be sensible to test a proposed therapy against a specified percentage, say 20%. The objective would be to find a new treatment that has a response percentage significantly higher than 20%.

(d) Ethics

All clinical investigators seek results which demonstrate that the overall prognosis for patients is getting better. Clinical trials in which patients do less well than they have in the past are to be avoided at all costs and to be concluded as early as possible. A comparative clinical trial should not be started unless there is some preliminary evidence suggesting that the new therapy is at least as good and possibly better than the standard. If this is accepted, the question can be raised whether it is ethical to enter patients on the standard therapy when there is little or no chance that the standard could be better than the new therapy. That is, the objective should be to study the new therapy until it can be concluded whether the new therapy is significantly more effective than the standard or not. Study of the new therapy could be stopped when the probability of its being more effective than the standard becomes very low.

The clinical investigator conducting studies in sequence of treatments is always giving what he considers to be the best treatment to his patients. Recruitment of patients to a clinic to receive this treatment is much easier than for the investigator who proceeds by simultaneous comparisons. The former investigator can promise all patients, even those who come from long distances, that they will receive what the investigator thinks is the current best treatment. The

latter type of investigator can promise only that the choice of treatment will be determined essentially by flipping a coin and that the treatments in the clinical trial are reasonably good ones.

Meier (1975) has stated the ethical problem as follows: "The view is often expressed that each patient must be afforded the presumed benefit of any estimated advantage of one treatment over another, regardless of how slight or uncertain that advantage may be. I insist that this view does not reflect my attitude about myself as a patient, nor does it reflect the attitude of most of us. Make no mistake about it, this position is incompatible with any experimenting whatever, controlled or casual. It does not favor judicious experimenting with a new technique or drug on carefully selected patients. That, after all, can be done in a controlled study. Rather, it forbids any experimenting at all." The ethical dilemma disappears if one proceeds sequentially in evaluating treatments - the presumed best treatment is always being given. However, what Meier and many other statisticians do not accept is that conducting studies in sequence can resolve the scientific problem of properly evaluating the relative effectiveness of treatments. This will be demonstrated by some examples from cancer clinical trials.

3. Examples of Non-Randomized Clinical Trials

In this section, some examples of non-randomized clinical trials are given in which patients in the control group were selected to be comparable to those receiving a study treatment. Patients in the control group were selected based upon their prognostic characteristics and the assumption was made in all studies that the patient characteristics chosen accounted for the major proportion of the patient-to-patient variability in response. Literature controls, matched controls, and patients from a sequence of studies will be considered in relation to the evaluation of study treatments.

(a) Literature Controls

In all circumstances in which the same or similar treatments have been used by others in a clinical investigation, it is desirable to use these patients as controls, even when there is also an internal group of control patients in the trial. Unfortunately, it is usually true that authors do not provide sufficient data in their papers so that it can be checked whether the patients reported in the literature are comparable to those in a given clinical trial. It certainly would be helpful if authors and those engaged in large cooperative group studies could make available basic data on punch cards or computer tape so that others might use the data for literature controls.

An example of a literature control group is given in the study reported by Luce et al (1971) in which combined cyclophosphamide, vincristine (Oncovin), and prednisone therapy (COP) for malignant lymphoma was compared to single agent treatment with cyclophosphamide or a vinca alkaloid (vinblastine for Hodgkin's disease and vincristine for lymphosarcoma) as reported by Carbone, Spurr, et al (1968). All patients in both studies had stage III or IV disease. However, patients who had received major prior chemotherapy or those with moderately impaired bone marrow reserve were excluded from the Carbone study. Thus, in terms of prior treatment and bone marrow reserve - two important prognostic factors - patients who had received little or no prior treatment in the Luce study were comparable to those in the Carbone study. The age and sex distributions were similar in the studies. Hence, when adjustment was made for prior therapy, it could be concluded that patients in the Carbone study were comparable to those in the Luce study. The complete remission rate following COP treatment was 36-50% in malignant lymphoma compared with 6-20% for the single agent treatment reported by Carbone. In addition, other series of patients receiving either single agents or COP treatment by a

slightly different schedule had similar results. Because both single agents and COP had response rates that were consistent from one study to the next and the evidence that COP was significantly superior, it seemed safe to conclude that COP was superior to single agent treatment in the induction of complete remissions.

Another example is that given by Sutow et al (1970) in which the survival experience of patients with Wilm's tumor or neuroblastoma, first treated in 1962, was compared to that of patients first treated in 1956. A total of 35 institutions participated in the study and, for patients with Wilm's tumor, it was demonstrated that the age distribution, percentage of children with metastases, and intensity of surgical and radiation therapy were comparable between the two time periods. However, 94% of patients received drug therapy (mainly actinomycin-D, vincristine, and cyclophosphamide) in 1962 compared with 28% in 1956. A significant improval in survival was demonstrated for patients of all ages without meatstases and for patients two years or older with metastases. The authors concluded that the increased clinical use of chemotherapeutic agents resulted in the significant improvement in the survival curves. For patients with neuroblastoma, though there was a slight difference in the survival experience for both non-metastatic and metastatic patients favoring those first treated in 1962, the difference was not near statistical significance and it was concluded that the increased use of chemotherapeutic agents did not result in a significant improvement in survival time.

A literature control group is useful when patients can be checked for comparability and, in some circumstances, when it can be demonstrated that patients in the literature have more favorable prognostic indicators. Authors should be encouraged to have details of their data available to others for comparison purposes.

(b) Matched Controls

In a matched control study in which patients are to be selected from a group of patients treated in the past, all new patients would receive the treatment

to be evaluated, say treatment A. A pairmate for each patient receiving A would be chosen at random from among the possible pairmates in the group of historical control patients who received treatment B. The applicability of this approach depends upon having a sufficiently large group of patients for potential pairmates. Patients obtained by this process who receive treatment A would be as comparable as possible to those on treatment B with respect to the patient characteristics used as a basis for the pairing. If sufficient patients are available, it may be desirable to select two control patients for each treated patient, making a comparison between control patients to test the selection process.

An example of this type of study is given by Bodey et al (1976) who compared the length of complete remission for patients with acute leukemia between two groups: a study group receiving late intensification chemotherapy and immunotherapy a median of 89 weeks (range of 58 to 194 weeks) after achievement of complete remission vs. a matched control group of patients who received maintenance therapy at monthly intervals, generally the same therapy that induced the remission. The objective of the late intensification study was to cure the patient by administering an intense program of therapy with new agents when the leukemia cell population was at a minimum. Patients were matched by age group, cell type, and length of remission prior to the start of late intensification therapy. There were 17 patients in the matched control group and 19 in the group receiving late intensification therapy (matched controls could not be found for two patients). The median duration of complete remission subsequent to late intensification therapy has not yet been reached but will be in excess of 98 weeks, only 5 patients relapsing of 19. The median length of subsequent remission in the matched control group was 24 weeks and there is a highly significant statistical difference between the two remission curves ($P<.01$). Comparing survival times between groups,

16 of the 19 patients receiving late intensification treatment are still alive and their median follow-up time is 97 weeks. The median survival time for patients in the matched control group is 56 weeks and the difference between curves was highly statistically significant ($P<.01$). Thus, this study has demonstrated the importance of a new concept in the treatment of patients with acute leukemia that may have resulted in a cure of some patients.

Another study by Bodey et al (1971) in patients with acute leukemia demonstrated that patients in a protected environment (PE) receiving prophylactic antibiotics and chemotherapy had significantly better length of complete remission (median of 55 weeks for PE, 26 weeks for controls), length of survival (median of 34 weeks for PE, 23 weeks for controls), and percentage of days spent with infection as related to neutrophil count than a matched control group of patients treated outside a protected environment.

(c) Controls Selected from a Sequence of Studies

There are many cooperative groups engaged in cancer research in the USA who proceed from one study to the next. Generally, there is little change over short intervals of time in institution, type of patient, criteria for diagnosis and response, and availability of supportive therapy. In this circumstance, it is sensible to compare results from a previous study with those of a current one. Using patients from a previous study as controls might be misleading if a relatively long time interval had elapsed between studies (say greater than 3 years) or if it could be demonstrated that important changes had taken place with respect to clinical investigators, type of patient, criteria for evaluation, etc. There are about 25 cooperative groups in the United States supported by the National Cancer Institute that proceed directly from one study to the next, have a stable group of clinical investigators, see the same types of patients from year to year, have the same access to supportive therapy measures and generally use the same

criteria of response in successive studies. Using patients from a previous study as controls would often be feasible for such cooperative groups.

Examples from studies conducted by the Southwest Oncology Group demonstrate that the same treatment administered in successive studies may be expected to lead to the same general result. In consecutive studies of previously untreated pediatric patients with acute leukemia, the complete bone marrow remission rates for patients treated with vincristine plus prednisone were 83% (72/87) in the ALinC #6 study and 86% (237/276) in the ALinC #7 study (Lonsdale et al, 1975). In consecutive studies of patients with Hodgkin's disease, the complete remission rate following MOPP treatment has remained very close to 80% for previously untreated patients with stage III or IV disease.

When consecutive studies of different treatments have been conducted, regression models can be utilized to test whether there are significant treatment differences, adjusting for values of the prognostic characteristics in the successive studies. If response is the end point for analysis, stepwise logistic regression procedures can be carried out to interpret the data (Cox(1970), Lee (1974)) If survival or length of response is the end point, Cox's regression model (Cox (197 may be used. An example will be given from successive studies conducted in the Southwest Oncology Group.

Over the past several years, the Southwest Oncology Group (SWOG) has conducted the following clinical studies in patients with adult acute leukemia: COAP vs. OAP vs. DOAP (from 2/71 to 10/72); a 10-day OAP study (from 6/73 to 1/75); and a CIAL study (from 1/75 to present). The designations of the drugs are as follows: C=Cyclophosphamide, O=Vincristine (Oncovin), A=Cytosine Arabinoside, and P=Prednisone. The CIAL study in the remission induction phase consisted of givir vincristine plus prednisone to all patients with less than 30,000 blasts in the

peripheral blood. For patients with 30,000 or more blasts, patients were randomized between sequential vs. simultaneous adriamycin-OAP treatment. In the first study, OAP was given by continuous infusion over a period of five days.

The complete remission rate for 5-day OAP was 43% (39/90), that for 10-day OAP was 53% (92/173), and the current complete response rate for patients in the combined groups on CIAL is 60% (70/117). The question arises, do these data indicate significantly improved complete remission rates by study, or is there evidence that the types of patients on the three studies might explain the differences in complete remission rates?

From previous studies in adult acute leukemia, the following patient characteristics have been identified as being predictive of response: age (years), infection status at start of study (0=no, 1=yes), acute myelocytic leukemia (0=no, 1=yes), hemoglobin value (gms %), and logarithm (white blood count). These five patient characteristics and two variables representing the linear and quadratic effect of treatments were included in a logistic regression equation. The regression equation obtained is as follows:

$$\log\left\{\frac{p_i}{1-p_i}\right\} = + .1276 - .0417(\text{Age}-44.73) + .5027(\text{Treat.linear}-.101)$$
$$- .7000(\text{Infection status}-.388) - .3806(\text{AML}-.830)$$
$$+ .0501(\text{Hemoglobin}-9.21) - .0597(\log(\text{WBC})-4.144)$$
$$+ .0207(\text{Treat.quadratic}+.407)$$

where p_i is the predicted complete remission rate based on the 7 patient characteristics.

The coefficients in the equation were determined by stepwise logistic regression (Lee(1974)) so the significance level of each entering characteristic can be calculated. The statistical significance level of each entering variable was: age ($P<.01$), treatment linear ($P<.01$), infection status ($P<.01$), AML ($P=.18$), hemoglobin ($P=.33$), log WBC ($P=.76$) and treatment quadratic ($P=.80$). This analysis

demonstrates that there is statistically significant evidence of a linear increasing trend in response rate by study and that age and infection status are significantly related to response rate.

Evidence that the five patient characteristics do predict complete remission rate is given in Table 1. A logistic regression equation was fit to the five patient characteristics in the 5 and 10-day OAP studies (excluding treatment as a possible characteristic). This equation is as follows:

$$\log\left\{\frac{p_i}{1-p_i}\right\} = .02888 - .04238(\text{Age}-.44031) - .59297(\text{Infection status}-.37) - .35854(\text{AML}-.872) - .01431(\text{Hemoglobin}-9.155) - .0208(\log(\text{WBC})-4.127).$$

Table 1 gives the observed and predicted numbers of patients responding on the 10-day OAP and CIAL studies. As would be expected, the relationship between observed and predicted probability of response was excellent for the 10-day OAP, since the equation is being re-applied to the same data from which it was derived. Note that there is also a good relationship between observed and predicted probability of response for patients on the CIAL study. The observed percentages responding were higher than predicted in patients with predicted probabilities under .60 and were in accord with predictions for patients with predicted probabilities over .60. Hence, there is some evidence that patients on the CIAL study produced higher observed responses in patients with relatively low predicted probabilities of response. When the equation was applied to the patients from 5-day OAP, the predicted complete remission rate was 52.1%; it was 50.0% for patients on 10-day OAP, and 50.8% for CIAL. Hence, there was strong evidence that patients on all three studies were comparable with respect to the five patient characteristics.

Cox's regression model was fit to the survival data from the three studies using the same five patient characteristics and treatment variables as in the analysis of response. Cox's model may be written as follows:

$$\lambda(t) = \exp\left\{\beta_1(x_1-\bar{x}_1) + \beta_2(x_2-\bar{x}_2) + \ldots + \beta_p(x_p-\bar{x}_p)\right\}\lambda_o(t)$$

where $\lambda(t)$ is the hazard function at time t, the β's are regression coefficients, the x's are patient characteristics potentially related to survival, the $\bar{x}$'s are average values, and $\lambda_o(t)$ is an arbitrary hazard function when all the x's are at their mean values. The model fit to the survival data from the three studies is as follows:

$$\log_e\left\{\frac{\lambda(t)}{\lambda_o(t)}\right\} = + .0319(\text{Age}-44.74) - .4269(\text{Treat.linear}-.10)$$
$$+ .4978(\text{Infection status}-.39) + .1435(\log(\text{WBC})-4.14)$$
$$- .0429(\text{Treat.quadratic}+.41) - .0097(\text{Hemoglobin}-9.21)$$
$$+ .0006(\text{AML}-.83).$$

The model was fit in forward stepwise fashion and the statistical significance of adding variables at each step was as follows: age ($P<.01$), treatment linear ($P=.001$), infection status ($P=.001$), log (WBC) ($P=.30$), treatment quadratic ($P=.39$), hemoglobin value ($P=.77$) and AML ($P=.99$). Hence, as in the analysis of response, age and infection status are the two characteristics most significantly related to survival time and there is evidence of a linear trend which indicates increasing survival time by study. Figure 1 gives the survival curves for patients on the three studies. The median survival time for patients receiving 5-day OAP was 7 weeks, that for patients receiving 10-day OAP was 38 weeks, and the median has not yet been reached for patients on the CIAL study. There is evidence of a significant advantage in survival for 10-day vs. 5-day OAP patients ($P<.015$) and nearly significant evidence that CIAL has superior survival to 10-day OAP ($P=.059$).

These regression analyses have permitted comparison to be made among treatment programs, adjusting for patient characteristics related to prognosis. Based upon these analyses, one could more confidently assert that there were real differences in response rate and survival among the three studies because patient characteristics were adjusted for in both analyses, patients were comparable in the three studies with respect to predicted probability of complete remission, and the same patient characteristics (namely, age and infection status) were significantly related to response and survival.

4. Discussion

The point of view has been presented that rational, scientific, and controlled clinical studies can be accomplished without randomization. In some circumstances, patients that are comparable in prognosis can be identified in successive studies which allow comparison between a group of patients under investigation and other groups treated in the past. Recording data which differs significantly from that observed in the past forms the basis for new knowledge. Confirmation of data by the same investigator and by other investigators in other institutions provides a convincing mechanism for generating knowledge which predicts for the future.

The major reasons for preferring the non-randomized to the randomized study are: a clinical investigator in a non-randomized study is always administering what he believes to be the best treatment for the disease under investigation so there is no ethical dilemma, and non-randomized studies require fewer patients and proceed more quickly so that new knowledge is gained faster.

Randomized studies are useful if there is no basis for choosing comparable patients treated in the past since patient characteristics related to prognosis

are unknown. Also, such studies could be considered when there is no preliminary evidence that one treatment is substantially better than another so that the ethical dilemma does not really arise. Thirdly, previous data will sometimes suggest that the same treatment program be studied according to different dosages or schedules, etc., and it is convenient to have these treatments in the same study. Fourthly, when studies are to be conducted over a very long term (say, 3-5 years or more) then patients could be randomized because there was genuine doubt that the ancillary aspects of the successive studies would be comparable.

In planning any clinical trial, there is no substitute for imaginative, original, and creative thought. The best clinical trials are those that have the best treatments in them, whether randomized or not. Clinical knowledge will advance when there has been careful analysis of past results as a basis for the formulation of significant hypotheses to be tested in objective and scientifically valid studies.

Acknowledgement

Dr. Thomas Lincoln suggested the Army experiment and Indianapolis 500 example in the Introduction.

References

1. Armitage, P. and Gehan, E.A. (1974),"Statistical methods for the identification and use of prognostic factors", International Journal of Cancer,13: 16-36.
2. Bodey,G.P., Gehan, E.A., Freireich, E.J et al (1971), "Protected environment-prophylactic antibiotic program in the chemotherapy of acute leukemia", American Journal of Medical Sciences, 262: 138-151.
3. Bodey, G.P., Freireich, E.J, Gehan, E.A. et al (1976), "Late intensification chemotherapy and immunotherapy for patients with acute leukemia in remission", Journal of the American Medical Association, in press.
4. Carbone, P., Spurr, Schneiderman, M. et al (1968), "Management of patients with malignant lymphoma: a comparative study with cyclophosphamide and vinca alkaloids", Cancer Research, 28: 811-822.
5. Chalmers, T.C., Block, J.D., Lee,S. (1972), "Controlled studies in clinical cancer research", New England Journal of Medicine, 287: 75-78.
6. Cochran, W.G., Cox, G.M. (1957), Experimental Designs, 2nd edition, New York, John Wiley and Sons.
7. Cox, D.R. (1970), The Analysis of Binary Data, London, Methuen Company.
8. Cox, D.R. (1972), "Regression models and life tables", Journal of the Royal Statistical Society (B), 34: 187-226.
9. Daniel, C. (1970), "Randomization and replication", Proceedings of the"Symposium on Statistical Aspects of Protocol Design", National Cancer Institute, San Juan, Puerto Rico, pp. 37-46.

10. Freireich, E.J and Gehan, E.A. (1974), "Historical controls reincarnated: an examination of techniques for clinical therapeutic research". Guest editorial in Cancer Chemotherapy Reports, part 1, vol. 58, no. 5, Sept./Oct., 1974, pp. 623-626.

11. Gehan, E.A. and Freireich, E.J (1974), "Non-randomized controls in cancer clinical trials", New England Journal of Medicine, 290: 198-203.

12. Hill, A.B. (1962), Statistical Methods in Clinical and Preventive Medicine, Oxford University Press, New York.

13. Holland, J.S., Frei, E. (1973), Section VIII, "Experimental designs in clinical trials", Cancer Medicine, Lea and Febiger, Philadelphia.

14. Lee, E. (1974), "A computer program for linear logistic regression analysis", Computer Programs in Biomedicine, 4: 80-92.

15. Lonsdale, D., Gehan, E.A. et al (1975), "Interrupted vs. continued maintenance in childhood acute leukemia", Cancer, 36, 2: 341-351.

16. Luce, J.K., Gambel, J.F., Wilson, H.E., et al (1971), "Combined cyclophosphamide, vincristine and prednisone therapy in malignant lymphoma", Cancer, 28: 306-317.

17. Meier, P. (1975), "Statistics and medical experimentation", Biometrics, 31, 2: 511-530.

18. Sutow, W. et al (1970), "Comparison of survival curves, 1956 vs. 1962, in children with Wilm's tumor and neuroblastoma", Pediatrics, 45: 800-811.

Table 1

Observed and Predicted Responses from Logistic Regression Equation on 10-Day OAP and CIAL Studies

	CIAL			10-Day OAP		
Predicted Probability of Response	Total No. Obs.	Observed No. (PC) Responding	Expected No. Responding	Total No. Obs.	Observed No. (PC) Responding	Expected No. Responding
0 - .19	5	0(0)	.805	8	0(0)	1.385
.20 - .39	35	16(46)	10.075	44	13(30)	13.749
.40 - .59	27	17(63)	13.489	50	30(60)	24.878
.60 - .79	26	20(77)	18.051	50	37(74)	34.946
.80 -1.00	13	12(92)	10.857	3	3(100)	2.530
Total	106	65(61)	53.277 (50.761)	155	83(54)	77.488 (49.994)

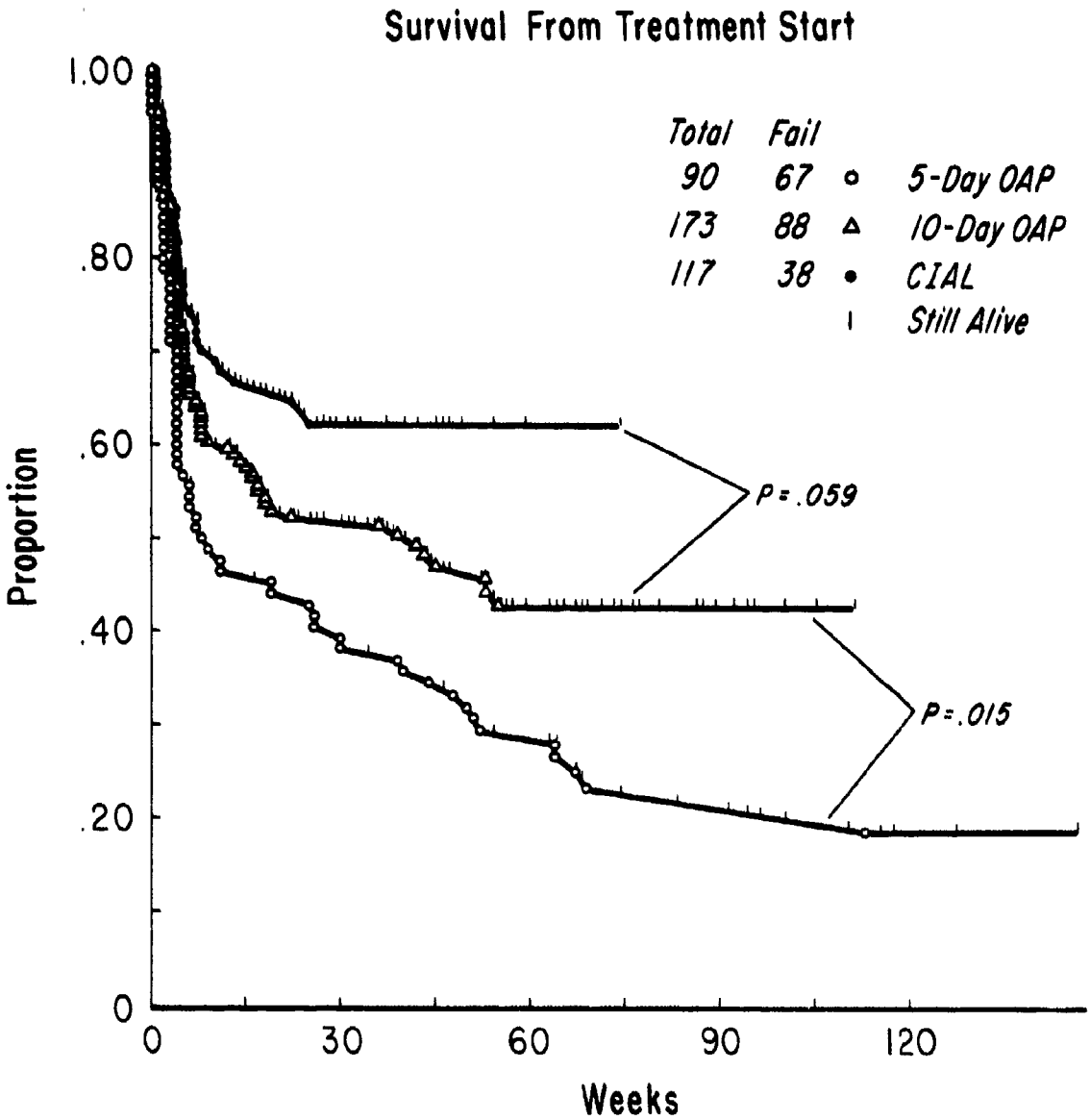

Survival From Treatment Start
Total Fail
90 67 5-Day OAP
173 88 10-Day OAP
117 38 CIAL
Still Alive
P=.059
P=.015
Proportion
1.00
.80
.60
.40
.20
0
0
30
60
90
120
Weeks

Figure 1

EMPIRICAL COMPARISON OF CRITERION REFERENCED MEASUREMENT MODELS

Kenneth I. Epstein
Frederick H. Steinheiser, Jr.
Army Research Institute for the
Behavioral and Social Sciences
Arlington, Virginia

ABSTRACT. The Army needs information about how well an individual can perform the tasks necessary for him to do his job. This information is often gathered by means of a "criterion-referenced test," a test made up of items directly related to the job of interest. The test results can be used in two ways. The first way is to sort individuals into two groups, one made up of those who can perform their job satisfactorily and the other made up of those who do not meet minimal job requirements. A second use of the test results is to estimate the "true" capability of the examinees to do the task being tested. These two uses are clearly related. If one can precisely estimate an individual's capability, then forming the two groups is not a problem. On the other hand, it may be possible to effectively form the two groups without getting good estimates of "true" capability.

Several psychometric models are available for grouping the individuals and/or for estimating "true" scores. For example, one may simply calculate the proportion of items correctly answered and use that proportion as an estimate of "true" capability. Alternatively, a binomial error model for deriving the expression for the regression of "true" score on observed score can be used and a "true" score calculated for each individual. Other possible models include a Bayesian Model II approach and a latent trait model such as the Rasch one parameter logistic model. Each of these models yields a somewhat different estimate of "true" capability for any given individual. It follows that the makeup of the job ability groups will vary from model to model. The purpose of this research is to empirically study the models referred to above. What is needed is an appropriate statistic (or statistics) and research design for comparing each model against all others given the same test data.

I. INTRODUCTION. The purpose of this paper is to elaborate on some technical details and to highlight specific statistical and research problems introduced in a previous paper by one of the authors (Epstein, 1975).

Epstein described four procedures for estimating true scores from observed scores. The first uses the observed proportion correct as an estimate of the true proportion correct. This procedure is straightforward and familiar. Hence, discussion of it will be reserved until

The views expressed in this paper are those of the authors and do not imply endorsement by the U. S. Army

the problem of comparing the models is developed. The other three procedures are 1) a binomial error model, 2) a Bayesian model, and 3) the Rasch logistic model. Each will be discussed in detail.

2. BINOMIAL ERROR MODEL. The binomial error model (Lord and Novick, 1968, pp. 508-529) is based on the assumption that the conditional distribution of observed score for given proportion correct true score (T) is the binomial distribution.

$$h(x|T) = \binom{n}{x} T^x (1-T)^{n-x}$$

x=0,1...n is the number of correct responses observed and n is the total number of items on the test.

It is assumed that items are scored dichotomously, that total score for an examinee is the numbe. of items answered correctly, that items are locally independent, and that items are equally diffitult for a given examinee.

The relationship between the observed score distribution and the underlying true score distribution can be written as follows:

$$\phi(x) = \binom{n}{x} \int_0^1 g(T)\, T^x (1-T)^{n-x}\, dT, \quad x=0,1,\ldots n, \text{ where } \phi(x) \text{ is}$$

the distribution of observed scores and g(T) is the unknown distribution of true scores.

It can be shown that if the regression of true score on observed score is linear then the distribution of observed score, symbolized h(x) to distinguish this special case from the general case $\phi(x)$, is negative hypergeometric.

$$h(x) \equiv \frac{b^{[n]}}{(a+b)^{[n]}} \frac{(-n)_x}{(-b)_x} \frac{(a)_x}{x!} \qquad x = 0,1\ldots n,$$

where

a and b are parameters to be determined and

$$n^{[x]} \equiv n(n-1)\ldots(n-x+1),$$

$$(a)_x \equiv a(a+1)\ldots(a+x-1), \quad n^{[0]} \equiv (a)_0 \equiv 1.$$

The parameters, a and b, can be expressed in terms of moments of the observed score distribution

$$a = (-1+1/\alpha_{21})\, \mu_x$$

$$b = -a-1+n/\alpha_{21}$$

$$\alpha_{21} \equiv \frac{n}{n-1} \left[1 - \frac{\mu_x(n-\mu_x)}{n\, \sigma_x^2}\right]$$

The discussion thus far has outlined an internal check of the appropriateness of this model for any given data set. That is, if one can show adequate fit to the negative hypergeometric distribution by the observed scores then it is reasonable to continue with this model assuming linear regression. If adequate fit is not obtained then either the more general nonlinear regression approach must be used or alternative models must be identified.

It can be shown that if the observed score distribution is negative hypergeometric, the true score distribution is either the two parameter beta distribution, or some other distribution having identical moments up through order n. In either case, the regression of true score on observed score is given by the linear equation

$$E(T|x) = \frac{a_{21}x}{n} + \frac{(1-a_{21})\mu_x}{n}, \quad x = 0,1,\ldots n.$$

3. BAYESIAN MODEL. The Bayesian model used to evaluate these data is described by Lewis, Wang, and Novick (1973). The procedure transforms the binomial test score data via an arc sine transformation. The resulting score is assumed to be a sample from a normal population with its mean value at the individual's transformed true ability. Distributions for the prior mean and variance of the examinee group's transformed scores are specified and posterior values calculated. Finally, the posterior marginal distributions for the transformed scores are obtained and estimates of individual true abilities on the original (proportion correct) scale are calculated. The mathematical details are outlined below.

The Freeman-Tukey transformation for binomial data is used in this procedure:

$$g_j = \frac{1}{2}\sin^{-1}\sqrt{\frac{x_j}{n+1}} + \sin^{-1}\sqrt{\frac{x_j+1}{n+1}}, \quad x_j = 1,2,\ldots n = \text{the}$$

number of correct responses. The g_j are assumed to be normally distributed with mean $\gamma_j = \sin^{-1}\sqrt{\pi_j}$ and variance $v = (4n+2)^{-1}$, where γ_j is the transformed value of the true proportion of correct responses, π_j. The validity of the assumption of normality and the suitability of the transformation for the procedures to follow can be shown to be adequate for examinee groups of at least 15 persons and for tests at least 8 items long.

The set of transformed variables, γ_j, is assumed to be a random sample from a normal distribution with mean μ_γ and variance ϕ_γ. μ_γ and ϕ_γ are further assumed to be independent and to have a uniform and inverse chi-square distribution respectively. Explicit expressions for the prior and posterior density functions are given in the Lewis, et al. paper.

The desired result of an analysis of this kind is the marginal posterior density function for γ_j. Unfortunately, an explicit expression for it is not obtainable from the joint posterior probability density function of the γ_j vector given the g_j vector. Lewis et al. show methods for obtaining the marginal means and variances for the γ_j using numerical integration. However, they indicate that for large sample sizes, the conditional posterior distribution of γ_j given ϕ_r and the g_j vector provides an acceptable approximation. The conditional approximation was used for the analysis of the data reported in the Epstein paper.

The conditional distribution of γ_j given ϕ_r and the g_j vector can be shown to be normal with mean

$$E(\gamma_j | \tilde{\phi}_r, \underset{\sim}{g}) = \frac{\tilde{\phi}_r \, g_j + v\bar{g}.}{\tilde{\phi}_r + v} ,$$

and variance

$$\text{var}(\gamma_j | \tilde{\phi}_r, \underset{\sim}{g}) = \frac{v(\tilde{\phi}_r + m^{-1}v)}{\tilde{\phi}_r + v} ,$$

where

$j = 1,2...m$ = the number of examinees,

$\underset{\sim}{g}$ = the vector of transformed scores, and

$\tilde{\phi}_r$ = the mode of ϕ_r given $\underset{\sim}{g}$.

$\tilde{\phi}_r$ can be obtained by solving the following equation:

$$(m + v + 1)\,\tilde{\phi}_r^3 + [(m + 2v + 3)\,v - \sum_i (g_j - \bar{g}.)^2 - \lambda]\,\tilde{\phi}_r^2$$

$$+ [(v + 2)\,v^2 - 2\lambda v]\,\tilde{\phi}_r - \lambda v^2 = 0 .$$

In the above equation, v is the degrees of freedom for the prior inverse chi-square distribution of ϕ_r. Lewis, et al. recommend that a value of eight be used for most practical applications. λ is the scale factor for the inverse chi-square distribution. It can be calculated by using the formula

$$\lambda = \frac{v - 2}{4(t+1)}$$

where t is interpreted as the number of test items that the prior information is considered to be equivalent to.

Once the γ_j have been calculated, the last step in the procedure is to calculate the estimates for the true proportion correct. This is accomplished by applying the following equation:

$$\Pi_j = (1 + \frac{1}{2n}) \sin^2 \gamma_j - \frac{1}{4n}$$

4. RASCH MODEL. The Rasch one parameter logistic model (Wright and Panchapakesan, 1969) assumes that the observed response a_{ni} of person n to item i is governed by a binomial probability function of person ability Z_n and item easiness E_i. The probability of a correct response is:

$$P(a_{ni} = 1) = \frac{Z_n E_i}{1+Z_n E_i} .$$

The probability of a wrong response is:

$$P(a_{ni} = 0) = 1 - P(a_{ni} = 1) = \frac{1}{1+Z_n E_i} .$$

These equations may be combined to yield

$$P(a_{ni}) = \frac{(Z_n E_i)^{a_{ni}}}{1+Z_n E_i} .$$

If we let $b_n = \log Z_n$ and $d_i = \log E_i$,

then

$$P(a_{ni}) = \frac{\exp(a_{ni}(b_n + d_i))}{1 + \exp(b_n + d_i)}$$

The number of correct responses to a given set of items is the only information needed to estimate person ability. All persons who get the same score will be estimated to have the same ability. Hence, in terms of score groups,

$$P(a_{ni}) = \frac{\exp(a_{ni}(b_j + d_i))}{1 + \exp(b_j + d_i)}$$

where j = score of person n, and all persons with a score j are estimated to have the same probability governing their responses to item i.

The equations obtained when the condition of a maximum likelihood is satisfied for the model described in the preceding equation are:

$$a_{+i} = \sum_{j}^{k-1} (r_j \exp(b_j^* + d_i^*)/(1+\exp(b_j^*+d_i^*))), \quad i = 1,2,\ldots k$$

$$j = \sum_{i}^{k} (\exp(b_j^* + d_i^*)/(1+\exp(b_j^* + d_i^*))), \quad j = 1,2,\ldots k-1$$

where a_{+i} = number of persons who get item i correct

j = the total test score, an ability estimate is obtained for each score

r_j = number of persons in score group j.

b_j^*, d_i^* = estimates of b_j and d_i

The method consists of computing d_i^* and b_j^* from the implicit equations above. The equations are handled as two independent sets and solved accordingly.

An approximation of a standard error for item estimates can be obtained by assuming that the variance of the item estimate is due primarily to the uncertainty in the item score a_{+i}. To a first approximation this gives:

$$V(d_i^*) = (\partial d_i/\partial a_{+i})^2 \, V(a_{+i})$$

which leads to:

$$V(d_i^*) = 1/\sum_{j}^{k-1} (r_j \exp(b_j^*+d_i^*)/(1+\exp(b_j^* + d_i^*))^2).$$

The major contribution to the error variance of the ability estimate comes from the variance in scores produced by a given individual. This part of the error variance depends upon the number of items and their easiness range.

An approximation of the variance of the abili y estimate b* is given by

$$V^*(b^*) = \{1/C(b^*)\exp(b^*)\} + \{1/C^2(b^*)\} \cdot \sum_{i}^{k} (V(d_i)\{\exp(d_i)/(1+\exp(d_i+b^*))^2\}^2)$$

where $C(b^*) = \sum_{i}^{k} (\exp(d_i)/(1+\exp(b^*+d_i))^2)$,

$V(d_i)$ is the variance of the item calibration d_i.

The first term in the denominator of the V*(b*) equation is due to the variance in the score, and the second term is due to the imprecision of item calibration. The first term is always larger than the second.

5. DISCUSSION OF THE PROBLEM. One characteristic of a useful model is that it has a small error of measurement. That is, the distribution of estimated scores for a given true score is closely clustered around the true score. The extent of the measurement error that can be expected with a given model is dependent on the variance of the estimated true score. For example, in the proportion correct model, the variance of the estimated true proportion correct is equal to $p(1-p)/n$. In this case the variance of the estimate will decrease as the number of observations increases. Thus it would seem that any level of precision could be obtained by simply adding observations. Unfortunately, for the number of items that are usually practical on a test, the level of precision possible is not completely satisfactory. It would be useful to compare the variance of the true score estimates obtained with the other models to the proportion correct model.

Therefore the question of how to derive an expression for the variance of the estimated true scores for the other models must be addressed. An expression for the binomial error model has been derived. Since the binomial error model results in a regression equation it seems reasonable to base the derivation on the general form of the error of estimation, $\sigma_E^2 = \sigma_T^2 \sqrt{1 - \rho_{xT}^2}$. The ratio of the variance of true scores to the variance of observed scores equals the reliability coefficient, $\frac{\sigma_c^2}{\sigma_x^2} = \alpha_{21}$, where σ_c is the variance of the true number correct. Since the true number correct equals the true proportion correct times the number of items, $C = nT$, one may write $\sigma_c^2 = n^2 \sigma_T^2$. Substituting, $\sigma_T^2 = \sigma_x^2 \alpha_{21}/n^2$. The reliability of a test equals the square of the correlation between true and observed scores, $\alpha_{21} = \rho_{xT}^2$. Hence, the variance of the estimated true score can be written

$$\sigma_E^2 = \frac{\sigma_x^2 \alpha_{21} (1 - \alpha_{21})}{n^2}$$

For the Bayesian and Rasch models expressions for the variances of the estimated true scores were not derived. In the case of the Bayesian model the output is in terms of the arc sine of the true proportion correct. While the sampling distribution of the transformed variable is known, the variance of the estimated true proportion correct itself was not determined. A similar problem exists for the Rasch model. The sampling distributions of the ability and item difficulty indices

are known as well as the explicit equation for calculating the proportion correct from those values. But an expression for the estimated true proportion correct has not been derived. In short, the problems are:
(1) For the Bayesian model, given the variance of α_j and the equation

$\Pi_j = (1 + 1/2n) \sin^2 \gamma_j - 1/4n$, what is the variance of Π_j ; and

(2) For the Rasch model given the variances of b* and d* and the equation $p\ (\text{correct}) = \frac{\exp(b^* + d^*)}{1 + \exp\ (b^* + d^*)}$, what is the variance of p?

As a result of the discussion during the session a solution to the above mathematical problems seems to be available. It was pointed out that methods exist for deriving standard errors of functions of random variables. One promising approach outlined in Kendall and Stuart (1969, p. 231) involves evaluating terms of a Taylor expansion. Using the Kendall and Stuart procedure it should be possible to derive expressions for the standard error of measurement for each of the models. This will allow for formal comparison of the models without real or simulated data.

The discussion then considered whether it was possible to compare the models by obtaining an estimate of "true score" and comparing it to the "real" true score. The problem lies in obtaining an acceptable true score. Three approaches were considered and are expected to provide a basis for future research. The first is to base model comparisons on Monte Carlo simulation studies. Monte Carlo studies provide an unambiguous true score but suffer from their lack of generalizability to practical applications. A second approach is to define true score as the score obtained on an instrument consisting of a large number of items. The models would then be used to estimate the true score using a smaller and more realistic number of items. This approach is empirical and more directly oriented to practical applications where testing time and the number of items that may be included in an instrument are limited. Although this approach suffers from the fact that the defined true score is not error free, the amount of error is not likely to be significant for practical purposes. The third approach would investigate the possibility of applying Geisser's predictive sample reuse method (Geisser, 1975) to the comparison of the models. Geisser's method may provide a more formal empirical approach to model comparison than the second approach discussed above, however, it has not been determined whether or not it is applicable to this research.

Four models for estimating true scores were presented and methods for comparing their outputs were discussed. Procedures for comparing the statistical properties of the models are available and relatively straightforward. Future research will be concerned with establishing the empirical validity of the models and their applicability to solving practical measurement problems.

References

Epstein, K.I. An empirical investigation of four criterion-referenced testing models. Presented at 17th Annual Conference of the Military Testing Association, Indianapolis, IND, Sept., 1975.

Geisser, S. The predictive sample reuse method with applications. Journal of the American Statistical Association, 1975, 70, 320-328.

Kendall, M.G. & Stuart, A. The advanced theory of statistics: 3rd Edition (Vol. 1). New York: Hafner Publishing Company, 1969.

Lewis, C., Wang, M.M., and Novick, M.R. Marginal distributions for the estimation of proportions in m groups. ACT Technical Bulletin 13. Iowa City, Iowa: American College Testing Program, 1973.

Lord, F.M. & Novick, M.R. Statistical theories of mental test scores. Reading, MASS: Addison-Wesley, 1968.

Wright, B., & Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

NON-RANDOMIZED FACTORIAL DESIGNS CHARACTERIZED BY TREND ELIMINATION AND A MINIMUM NUMBER OF FACTOR LEVEL CHANGES

Les Lancaster and Steve Reynolds
U.S. Army Operational Test and Evaluation Agency
Falls Church, Virginia

ABSTRACT. An admissible set of run orders is developed for 2^p factorial designs restricted to trend elimination. The best design is then selected from this admissible set having the minimum number of factor level changes. The procedure is developed for p=5 where admissible sets are generated between various mixtures of linear, quadratic, and cubic trend elimination and main effects, first order interactions, and second order interactions. The number of factor level changes is used to generate the admissible set.

1. INTRODUCTION. The design of two-level factorial experiments robust against time trends will be illustrated in this paper. In fact designs with zero time trends will be displayed that also keep the number of factor level changes form run to run small. Both of these features are essential in operational testing due to resource problems. Operational cost effectiveness is achieved by minimizing the number of factor level changes. Soldier learning and selection is controlled by an elimination of time trends in the experimental designs. Thus, these designs are characterized by specifying the run orders prior to running the tests. A combinatorial technique is developed for generating these desirable designs.

In the planning of an experiment costs can be reduced by a multi-phase design. The first phase would be the design of all controllable factors at their low and high levels. Additional phases would be adaptive. That is, the results of the first phase would be decisive for determining the design for the additional phases. Thus, forcing the complex overall design to be developed in the real time mode. However, the possible options at each phase are planned and designed a priori and the results of the previous phase trigger the design decisions for the next phase. This report will be concerned with the first phase where p factors are varied, each at two levels.

A method for the selection of run orders spaced at equal time intervals is developed whereby a subset of possible or admissible run order choices is restricted to trend elimination. The designer then has the option to randomize on this admissible set or else he can select the run order with a minimum number of factor level changes. With respect to trend elimination Figure 1 summarizes seven admissible subsets which will be studied in Chapter 6. However, cases two and three admit empty sets and are included for academic purposes.

FIGURE 1. Cases to be Considered

Case Number	Highest Restriction On		
	Main Effects	1st Order Interactions	2nd Order Inter
1	L	L	L
2	Q	Q	-
3	C	L	-
4	C	-	-
5	Q	L	-
6	L	L	-
7	Q	-	-

In Figure 1 the following notation is used:

L = Linear
Q = Quadratic and linear
C = Cubic, quadratic, and linear

The different cases can be expressed in vector notation by writing each case as (i, j, k). For example, case 5 can be expressed as (Q, L -). Utilizing this notation, the coordinate denotes where the restriction is to be placed and the coordinate value deontes the type of restriction. This will become clearer in Chapter 6.

The options left to the test designer for each of the cases are very flexible. In certain situations the choice for a run order may be dictated by other criteria such as engineering judgement with respect to some of the factor interactions. For example, some of the factor interactions or treatment combinations may be null or of no importance to the experimenter. For these situations the chosen run order can have a smaller number of factor level changes as a tradeoff for a higher time trend for the null treatment combinations.

The developed method is an alternative to full randomization. Some experimenters often use blocks to gain sensitivity at the expense of full randomization by reducing time trends to an average variation within blocks. However, if the blocks contain many runs, then the average trend within a block may still cause a disturbing effect. In the developed method randomization is restricted to the admissible set of runs whereby a price tag can even be attached to each ordered sequence of runs in the admissible set. Selection is then based on the set with the total number of factor level changes minimized. Procedures for partial randomization with respect to equivalence classes is left as an option to the designer.

2. REVIEW OF PERTINENT LITERATURE. In this paper admissible sets are restricted to zero time trends where the optimal run order is chosen which has a minimum number of factor level changes. Other work has restricted to admissible sets having the minimum number of factor level changes where the optimal run order is chosen which has a minimum (non-zero) simple or multiple correlation with time. In this paper the admissible sets have zero simple and multiple correlations with time. Thus far in the literature and including this paper only two-level factors have been studied.

Addelman (1) briefly summarizes the state-of-the-art up to March 1972.

Daniel and Wilcoxon (2) analyze full fractional factorial designs with respect to linear and quadratic time trends. Their approach is extended in this paper. They do not consider factor level changes in their run orders.

Draper and Stoneman (4) were the first to consider the tradeoff between factor level changes and linear time trends. However, they look mostly at the combinatorials and it appears that they use search techniques to display their run orders.

Tiahrt and Weeks (6) consider the selection of run orders with respect to factor level changes plus randomization on equivalence classes.

Dickinson (3) restricts to the minimum number of factor level changes and then selects his run orders having minimum simple and multiple correlations with linear time trends. He uses a computer search technique to find a few of the many possible run orders.

Thomas (5) considers run orders with the minimum number of factor level changes and applies the procedure to sensitivity analysis of parameters in large scale deterministic computer models.

3. METHOD OF DESIGN SELECTION. The method will be illustrated by application to a 2^5 factorial design with N = 32 runs. That is, a full factorial design. The extension to designs with p > 5 will be obvious from the illustration.

A 2^p factorial design is characterized by $N = 2^p$ runs of p factors; with each factor at two levels. For p = 5, Figure 2 displays the design matrix of $\pm$ 1's (1's are omitted for ease of typing) in standard Yates notation for the 32 runs and the 32 treatment combinations where "T" denotes the total treatment combination which is omitted in the selection criterion.

The Yates algorithm will be used for computing polynomial trend of factors at two levels. Daniel and Wilcoxon (reference 1) have applied the Yates algorithm to the integer linear and quadratic Tchebycheff orthogonal polynomials given in Figure 3. The Yates solution is equivalent to performing the matrix product between the design matrix (plus and minus ones as given by Figure 2) and the polynomial vector. The Yates solution is much faster than the matrix product. The Daniel-Wilcoxon procedure is applied here where we extend up to the (p-2)th order of the polynomial. Further, the method developed in this paper will take into account the number of factor level changes. In fact, it turns out that the number of factor level changes for each factor characterizes and complements the standard Yates design.

In Figure 3 only the first 16 numbers are arrayed. The second set of 16 numbers is found by reflecting each column downward and reversing the sign for the linear and cubic column. For example, the 32nd number for each column will be -31, 155, and 899.

For p = 5, Figure 4 gives the Yates solution performed on the Tchebycheff orthogonal polynomials (Figure 3) up to the third order. In Figure 4 the ordering of the treatment combinations has been changed from the standard Yates ordering to a more convenient ordering for the method to be developed in this paper. It turns out that this new ordering groups the various types of treatments with either sets of zeros or sets of non-zeros.

FIGURE 2. Standard Yates Notation for The Design Matrix for 32 Runs

	TABA B	CABA CCB C	DABA DDB D	CABA DCCB DDC D	EABA EEB E	CABA ECCB EEC E	DABA EDDB EED E	CABA DCCB EDDC EED E
1	+	−	−	+	−	+	+	−
2	+	−	−	+	−	+	+	−
3	+	−	−	+	−	+	+	−
4	+	−	−	+	−	+	+	−
5	+	+	−	−	−	−	+	+
6	+	+	−	−	−	−	+	+
7	+	+	−	−	−	−	+	+
8	+	+	−	−	−	−	+	+
9	+	−	+	−	−	+	−	+
10	+	−	+	−	−	+	−	+
11	+	−	+	−	−	+	−	+
12	+	−	+	−	−	+	−	+
13	+	+	+	+	−	−	−	−
14	+	+	+	+	−	−	−	−
15	+	+	+	+	−	−	−	−
16	+	+	+	+	−	−	−	−
17	+	−	−	+	+	−	−	+
18	+	−	−	+	+	−	−	+
19	+	−	−	+	+	−	−	+
20	+	−	−	+	+	−	−	+
21	+	+	−	−	+	+	−	−
22	+	+	−	−	+	+	−	−
23	+	+	−	−	+	+	−	−
24	+	+	−	−	+	+	−	−
25	+	−	+	−	+	−	+	−
26	+	−	+	−	+	−	+	−
27	+	−	+	−	+	−	+	−
28	+	−	+	−	+	−	+	−
29	+	+	+	+	+	+	+	+
30	+	+	+	+	+	+	+	+
31	+	+	+	+	+	+	+	+
32	+	+	+	+	+	+	+	+

FIGURE 3. Orthogonal Polynomials

Linear	Quadratic	Cubic
31	155	-899
29	125	-551
27	97	-261
25	71	-25
23	47	161
21	25	301
19	5	399
17	-13	459
15	-29	485
13	-43	481
11	-55	451
9	-65	399
7	-73	329
5	-79	245
3	-83	151
1	-85	51

FIGURE 4. The Yates Solution

Treatment	Level Changes	Linear	Quadratic	Cubic
A	31	32	0	1088
B	15	64	0	2144
C	7	128	0	4032
D	3	256	0	6016
E	1	512	0	4352
AB	16	0	32	0
AC	24	0	64	0
AD	28	0	128	0
AE	30	0	256	0
BC	8	0	128	0
BD	12	0	256	0
BE	14	0	512	0
CD	4	0	512	0
CE	6	0	1024	0
DE	2	0	2048	0
ABC	23	0	0	128
ABD	19	0	0	256
ABE	17	0	0	512
ACD	27	0	0	512
ACE	25	0	0	1024
ADE	29	0	0	2048
BCD	11	0	0	1024
BCE	9	0	0	2048
BDE	13	0	0	4096
CDE	5	0	0	8192
ABCD	20	0	0	0
ABCE	22	0	0	0
ABDE	18	0	0	0
ACDE	26	0	0	0
BCDE	10	0	0	0
ABCDE	21	0	0	0

The factor level changes are also given in Figure 4. Note that the number of factor level changes vary from 1 to 31. The main effect for A has the maximum number of factor level changes. For determining the number of factor level changes for any design only the level changes for the main effects are summed. Therefore, the standard Yates design is characterized by 57 factor level changes. Thus, as the references show, the standard Yates design is undesirable with respect to factor level changes. Also, the standard Yates design has large correlations with time, again an undesirable characteristic. Thus, optimal designs will be found in this paper having admissible properties.

The time counts for each treatment are the same as the Yates solution given in Figure 4. Note that for the standard Yates design the main effects have zero quadratic time trend. The first order treatment interactions have zero linear and zero cubic time trend. The second order treatment interactions have non-zero cubic time trend. The third order treatment interactions have all zero time trend. These observations are utilized to construct admissible run orders for the cases given in Figure 1.

The method consists of developing a new algebra whereby each of the 31 treatments is denoted by the number of factor level changes. In effect the new algebra permutes the 31 columns of Figure 2 into an optimal design. In the next section the development will be presented via illustration.

In Chapter 6 admissible sets of run orders for various cases will be constructed. In these cases whenever the designer has the option to randomize, it is to be understood that he can also randomize with respect to two equivalence classes.

One equivalence class is defined on the factor names. That is, the names (for example, A, B, C, D, or E) can be chosen at random for the admissible set. There are p! elements in this equivalence class.

A second equivalence class is defined on the choice of the high and low levels for one or more factors. That is, the designer can choose the plus and minus signs for each main effect at random. There are N elements in this equivalence class.

4. ALGEBRA. Multiplication of any two of the 31 treatments defined by Figure 2 entails pariwise multiplication of the 32 elements making up each of the columns of Figure 2. The classical method of multiplication will be utilized, whereby numbers, rather than letters, will be used to denote the treatment names. These numbers are the number of factor level changes for that particular treatment. That is, in Figure 4 instead of denoting the treatments by column one, column two will be used to denote the treatments as assigned by the standard Yates notation. As an example, the classical multiplication given as follows:

AC * ABD = BCD

is represented in the new algebra as follows:

24 * 19 = 11

Note that this triplet can be represented in three different ways as follows:

(i) 24 * 19 = 11

(ii) 19 * 11 = 24

(iii) 24 * 11 = 19

Figure 5 displays the 155 possible unique triplets as representation (iii) in a two-way table. To read off any product from Figure 5, note that the maximum value is the row, the minimum value is the column, and the value in between is the element of the matrix or body of the table. In Figure 5 all $\binom{31}{2}$ or 465 different triplets could have been displayed by filling in the blanks. However, by filling in only representation (iii) as defined above a pattern emerges. On extension to higher level designs, this pattern can be taken into account in developing a recursive method.

Figure 5. The 155 Possible Multiplications

MAXIMUM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1															
2															
3	2														
4															
5	4														
6		4	5												
7	6	5	4												
8															
9	8														
10		8	9												
11	10	9	8												
12				8	9	10	11								
13	12			9	8	11	10								
14		12	13	10	11	8	9								
15	14	13	12	11	10	9	8								
16															
17	16														
18		16	17												
19	18	17	16												
20				16	17	18	19								
21	20			17	16	19	18								
22		20	21	18	19	16	17								
23	22	21	20	19	18	17	16								
24								16	17	18	19	20	21	22	23
25	24							17	16	19	18	21	20	23	22
26		24	25					18	19	16	17	22	23	20	21
27	26	25	24					19	18	17	16	23	22	21	20
28				24	25	26	27	20	21	22	23	16	17	18	19
29	28			25	24	27	26	21	20	23	22	17	16	19	18
30		28	29	26	27	24	25	22	23	20	21	18	19	16	17
31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16

MINIMUM

5. SIEVE. In order to generate optimal or admissible designs the procedure entails development and utilization of a technique which shall be called a sieve. The first step of the sieve is formed by displaying the information from Figure 5 in Figure 6 for all 465 possible triplets. In Figure 6 each one of the 31 treatments is determined by any one of the corresponding 15 pairs. That is, the pairs are choices for the two main effects A* and B* and the product is the choice for the treatment AB*. The superscript * denotes the treatments belonging to a possible candidate for an optimal or admissible design. Further, in Figure 6, the symbols "-", "L", "Q", or "C" are taken from Figures 1 and 4 and displayed as an aid for sifting out the desired restrictions for the various cases of Figure 1. The idea is to sequentially search down each of the 31 blocks of Figure 6 and sift out the desired candidates for an admissible design. After this first step of the sieve, the designer will have possible candidates for A*, B*, and AB*.

Figure 6. Choices For A*, B*, and AB*

1 \ -	2	3	-
L	2	3	-
L	4	5	Q
L	6	7	-
L	8	9	Q
C	10	11	Q
L	12	13	Q
L	14	15	-
L	16	17	Q
C	18	19	Q
C	20	21	C
C	22	23	Q
L	24	25	Q
C	26	27	Q
L	28	29	Q
L	30	31	-

2 \ L	1	3	-
-	1	3	-
L	4	6	L
Q	5	7	-
L	8	10	C
Q	9	11	Q
L	12	14	L
Q	13	15	-
L	16	18	C
Q	17	19	Q
C	20	22	C
C	21	23	Q
L	24	26	C
Q	25	27	Q
L	28	30	L
Q	29	31	-

3 \ -	1	2	L
-	1	2	L
L	4	7	-
Q	5	6	L
L	8	11	Q
Q	9	10	C
L	12	15	-
Q	13	14	L
L	16	19	Q
Q	17	18	C
C	20	23	Q
C	21	22	C
L	24	27	Q
Q	25	26	C
L	28	31	-
Q	29	30	L

4 \ L	1	5	Q
-	1	5	Q
L	2	6	L
-	3	7	-
L	8	12	L
Q	9	13	Q
C	10	14	L
Q	11	15	-
L	16	20	C
Q	17	21	C
C	18	22	C
Q	19	23	Q
L	24	28	L
Q	25	29	Q
C	26	30	L
Q	27	31	-

Figure 6. Choices For A*, B*, and AB* (continued)

5 \ Q			
-	1	4	L
L	2	7	-
-	3	6	L
L	8	13	Q
Q	9	12	L
C	10	15	-
Q	11	14	L
L	16	21	C
Q	17	20	C
C	18	23	Q
Q	19	22	C
L	24	29	Q
Q	25	28	L
C	26	31	-
Q	27	30	L

6 \ L			
-	1	7	-
L	2	4	L
-	3	5	Q
L	8	14	L
Q	9	15	-
C	10	12	L
Q	11	13	Q
L	16	22	C
Q	17	23	C
C	18	20	C
Q	19	21	C
L	24	30	L
Q	25	31	-
C	26	28	L
Q	27	29	Q

7 \ -			
-	1	6	L
L	2	5	Q
-	3	4	L
L	8	15	-
Q	9	14	L
C	10	13	Q
Q	11	12	L
L	16	23	Q
Q	17	22	C
C	18	21	C
Q	19	20	C
L	24	31	-
Q	25	30	L
C	26	29	Q
Q	27	28	L

8 \ L			
-	1	9	Q
L	2	10	C
-	3	11	Q
L	4	12	L
Q	5	13	Q
L	6	14	L
-	7	15	-
L	16	24	L
Q	17	25	Q
C	18	26	C
Q	19	27	Q
C	20	28	L
C	21	29	Q
C	22	30	L
Q	23	31	-

Figure 6. Choices For A*, B*, and AB* (continued)

9 / Q			
-	1	8	L
L	2	11	Q
-	3	10	C
L	4	13	Q
Q	5	12	L
L	6	15	-
-	7	14	L
L	16	25	Q
Q	17	24	L
C	18	27	Q
Q	19	26	C
C	20	29	Q
C	21	28	L
C	22	31	-
Q	23	30	L

10 / C			
-	1	11	Q
L	2	8	L
-	3	9	Q
L	4	14	L
Q	5	15	-
L	6	12	L
-	7	13	Q
L	16	26	C
Q	17	27	Q
C	18	24	L
Q	19	25	Q
C	20	30	L
C	21	31	-
C	22	28	L
Q	23	29	Q

11 / Q			
-	1	10	C
L	2	9	Q
-	3	8	L
L	4	15	-
Q	5	14	L
L	6	13	Q
-	7	12	L
L	16	27	Q
Q	17	26	C
C	18	25	Q
Q	19	24	L
C	20	31	-
C	21	30	L
C	22	29	Q
Q	23	28	L

12 / L			
-	1	13	Q
L	2	14	L
-	3	15	-
L	4	8	L
Q	5	9	Q
L	6	10	C
-	7	11	Q
L	16	28	L
Q	17	29	Q
C	18	30	L
Q	19	31	-
C	20	24	L
C	21	25	Q
C	22	26	C
Q	23	27	Q

Figure 6. Choices For A*, B*, and AB* (continued)

	13 / Q		
-	1	12	L
L	2	15	-
-	3	14	L
L	4	9	Q
Q	5	8	L
L	6	11	Q
-	7	10	C
L	16	29	Q
Q	17	28	L
C	18	31	-
Q	19	30	L
C	20	25	Q
C	21	24	L
C	22	27	Q
Q	23	26	C

	14 / L		
-	1	15	-
L	2	12	L
-	3	13	Q
L	4	10	C
Q	5	11	Q
L	6	8	L
-	7	9	Q
L	16	30	L
Q	17	31	-
C	18	28	L
Q	19	29	Q
C	20	26	C
C	21	27	Q
C	22	24	L
Q	23	25	Q

	15 / -		
-	1	14	L
L	2	13	Q
-	3	12	L
L	4	11	Q
Q	5	10	C
L	6	9	Q
-	7	8	L
L	16	31	-
Q	17	30	L
C	18	29	Q
Q	19	28	L
C	20	27	Q
C	21	26	C
C	22	25	Q
Q	23	24	L

	16 / L		
-	1	17	Q
L	2	18	C
-	3	19	Q
L	4	20	C
Q	5	21	C
L	6	22	C
-	7	23	Q
L	8	24	L
Q	9	25	Q
C	10	26	C
Q	11	27	Q
L	12	28	L
Q	13	29	Q
L	14	30	L
-	15	31	-

Figure 6. Choices For A*, B*, and AB* (continued)

17 \ Q			
-	1	16	L
L	2	19	Q
-	3	18	C
L	4	21	C
Q	5	20	C
L	6	23	Q
-	7	22	C
L	8	25	Q
Q	9	24	L
C	10	27	Q
Q	11	26	C
L	12	29	Q
Q	13	28	L
L	14	31	-
-	15	30	L

18 \ C			
-	1	19	Q
L	2	16	L
-	3	17	Q
L	4	22	C
Q	5	23	Q
L	6	20	C
-	7	21	C
L	8	26	C
Q	9	27	Q
C	10	24	L
Q	11	25	Q
L	12	30	L
Q	13	31	-
L	14	28	L
-	15	29	Q

19 \ Q			
-	1	18	C
L	2	17	Q
-	3	16	L
L	4	23	Q
Q	5	22	C
L	6	21	C
-	7	20	C
L	8	27	Q
Q	9	26	C
C	10	25	Q
Q	11	24	L
L	12	31	-
Q	13	30	L
L	14	29	Q
-	15	28	L

20 \ C			
-	1	21	C
L	2	22	C
-	3	23	Q
L	4	16	L
Q	5	17	Q
L	6	18	C
-	7	19	Q
L	8	28	L
Q	9	29	Q
C	10	30	L
Q	11	31	-
L	12	24	L
Q	13	25	Q
L	14	26	C
-	15	27	Q

Figure 6. Choices For A*, B*, AB* (continued)

	21 / C		
-	1	20	C
L	2	23	Q
-	3	22	C
L	4	17	Q
Q	5	16	L
L	6	19	Q
-	7	18	C
L	8	29	Q
Q	9	28	L
C	10	31	-
Q	11	30	L
L	12	25	Q
Q	13	24	L
L	14	27	Q
-	15	26	C

	22 / C		
-	1	23	Q
L	2	20	C
-	3	21	C
L	4	18	C
Q	5	19	Q
L	6	16	L
-	7	17	Q
L	8	30	L
Q	9	31	-
C	10	28	L
Q	11	29	Q
L	12	26	C
Q	13	27	Q
L	14	24	L
-	15	25	Q

	23 / Q		
-	1	22	C
L	2	21	C
-	3	20	C
L	4	19	Q
Q	5	18	C
L	6	17	Q
-	7	16	L
L	8	31	-
Q	9	30	L
C	10	29	Q
Q	11	28	L
L	12	27	Q
Q	13	26	C
L	14	25	Q
-	15	24	L

	24 / L		
-	1	25	Q
L	2	26	C
-	3	27	Q
L	4	28	L
Q	5	29	Q
L	6	30	L
-	7	31	-
L	8	16	L
Q	9	17	Q
C	10	18	C
Q	11	19	Q
L	12	20	C
Q	13	21	C
L	14	23	C
-	15	23	Q

Figure 6. Choices For A*, B*, and AB* (continued)

25 / Q			
-	1	24	L
L	2	27	Q
-	3	26	C
L	4	29	Q
Q	5	28	L
L	6	31	-
-	7	30	L
L	8	17	Q
Q	9	16	L
C	10	19	Q
Q	11	18	C
L	12	21	C
Q	13	20	C
L	14	23	Q
-	15	22	C

26 / C			
-	1	27	Q
L	2	24	L
-	3	25	Q
L	4	30	L
Q	5	31	-
L	6	28	L
-	7	29	Q
L	8	18	C
Q	9	19	Q
C	10	16	L
Q	11	17	Q
L	12	22	C
Q	13	23	Q
L	14	20	C
-	15	21	C

27 / Q			
-	1	26	C
L	2	25	Q
-	3	24	L
L	4	31	-
Q	5	30	L
L	6	29	Q
-	7	28	L
L	8	19	Q
Q	9	18	C
C	10	17	Q
Q	11	16	L
L	12	23	Q
Q	13	22	C
L	14	21	C
-	15	20	C

28 / L			
-	1	29	Q
L	2	30	L
-	3	31	-
L	4	24	L
Q	5	25	Q
L	6	26	C
-	7	27	Q
L	8	20	C
Q	9	21	C
C	10	22	C
Q	11	23	Q
L	12	16	L
Q	13	17	Q
L	14	18	C
-	15	19	Q

Figure 6. Choices For A*, B*, and AB* (continued)

	29 / Q		
-	1	28	L
L	2	31	-
-	3	30	L
L	4	25	Q
Q	5	24	L
L	6	27	Q
-	7	26	C
L	8	21	C
Q	9	20	C
C	10	23	Q
Q	11	22	C
L	12	17	Q
Q	13	16	L
L	14	19	Q
-	15	18	C

	30 / L		
-	1	31	-
L	2	28	L
-	3	29	Q
L	4	26	C
Q	5	27	Q
L	6	24	L
-	7	25	Q
L	8	22	C
Q	9	23	Q
C	10	20	C
Q	11	21	C
L	12	18	C
Q	13	19	Q
L	14	16	L
-	15	17	Q

	31 / -		
-	1	30	L
L	2	29	Q
-	3	28	L
L	4	27	Q
Q	5	26	C
L	6	25	Q
-	7	24	L
L	8	23	Q
Q	9	22	C
C	10	21	C
Q	11	20	C
L	12	19	Q
Q	13	18	C
L	14	17	Q
-	15	16	L

The second step of the sieve is concerned with finding the main effect C* given candidates A*, B*, and AB*. Since the main effects can be relabeled with respect to equivalence classes, the choice for C* can be subjected to the following constraint:

$$A^* < B^* < C^*$$

Now to choose C*, suppose that A* and B* are fixed at "5" and "9" respectively, then, for this example, Figure 7 displays 28 possible choices for C*. In Figure 7, for any choice of C*, the remaining three treatments in that same row are automatically determined and assigned as shown in Figure 8, for example, for the second row of Figure 7. That is, the treatments in each row of Figure 7 for C* can be permuted, but only these seven rows can be defined.

Figure 7. Choices For C*

A*	B*	AB*	C*			
5	9	12	10	15	3	6
5	9	12	11	14	2	7
5	9	12	13	8	4	1
5	9	12	16	21	25	28
5	9	12	17	20	24	29
5	9	12	18	23	27	30
5	9	12	19	22	26	31

Figure 8. Choices for AC*, BC*, And ABC*

A*	B*	AB*	C*	AC*	BC*	ABC*
5	9	12	11	14	2	7
5	9	12	14	11	7	2
5	9	12	2	7	11	14
5	9	12	7	2	14	11

In applying the sieve, the last two rows of Figure 8 can be crossed out, for the example, due to the ordering constraint on these three candidates for the main effects. This ordering constraint will also reduce the set of choices given in Figure 7. Case restrictions will further reduce the set of choices. Therefore, as the sequential search for candidates progresses, or as A* and B* increase in value, the set of possible choices for each new C* decreases. Usually, the possibilities need not be exhaustive as shown by the cases studied in Chapter 6.

At this stage of the sieve, for each possible candidate for an admissible design, it turns out that seven out of the 31 possible treatments are now fixed. The third step of the sieve is concerned with finding admissible choices for D* and E*. To continue the sequential search, the ordering constraint is extended as follows:

$$A^* < B^* < C^* < D^* < E^*$$

Suppose that the candidate under consideration at this step is given by the first row of Figure 8. The new candidates will be found from the blocks of Figure 6. For this example, the best candidate for D* is "13". Further, on checking the 13th block of Figure 6 and crossing out the seven pairs corresponding to the seven fixed treatments, the best candidate for E* is "16". These two candidate blocks are repeated from Figure 6 as Figure 9 but without any case restrictions. Also in Figure 9 the seven treatments for this example are circled. As a check on the validity of the chosen design, note that in Figure 9, each block has seven pairs that are eliminated. Case restrictions would eliminate more pairs. Due to the ordering constraint and since the sum of the factor level changes for the main effects is to be minimized, only one pair of D* and E* treatments need be found for each candidate up to this step of the sieve. However, the three main effects from step 2 will not have a sum that strictly increases or decreases as the sequential search progresses.

After all admissible designs are sufficiently searched and displayed the designer selects the optimal design with respect to the particular case under consideration. However, due to the ordering criterion and the fixed choice of the plus and minus signs in Figure 2, the above selection is up to an equivalence class. Therefore, at this point, the designer has the option to randomize.

Figure 9. Choices For D* And E*

13	
1	(12)
(2)	15
3	(14)
4	(9)
(5)	8
6	(11)
(7)	10
16	29
17	28
18	31
19	30
20	25
21	24
22	27
23	26

16	
1	17
(2)	18
3	14
4	20
(5)	21
6	22
(7)	23
8	24
(9)	25
10	26
(11)	27
(12)	28
13	29
(14)	30
15	31

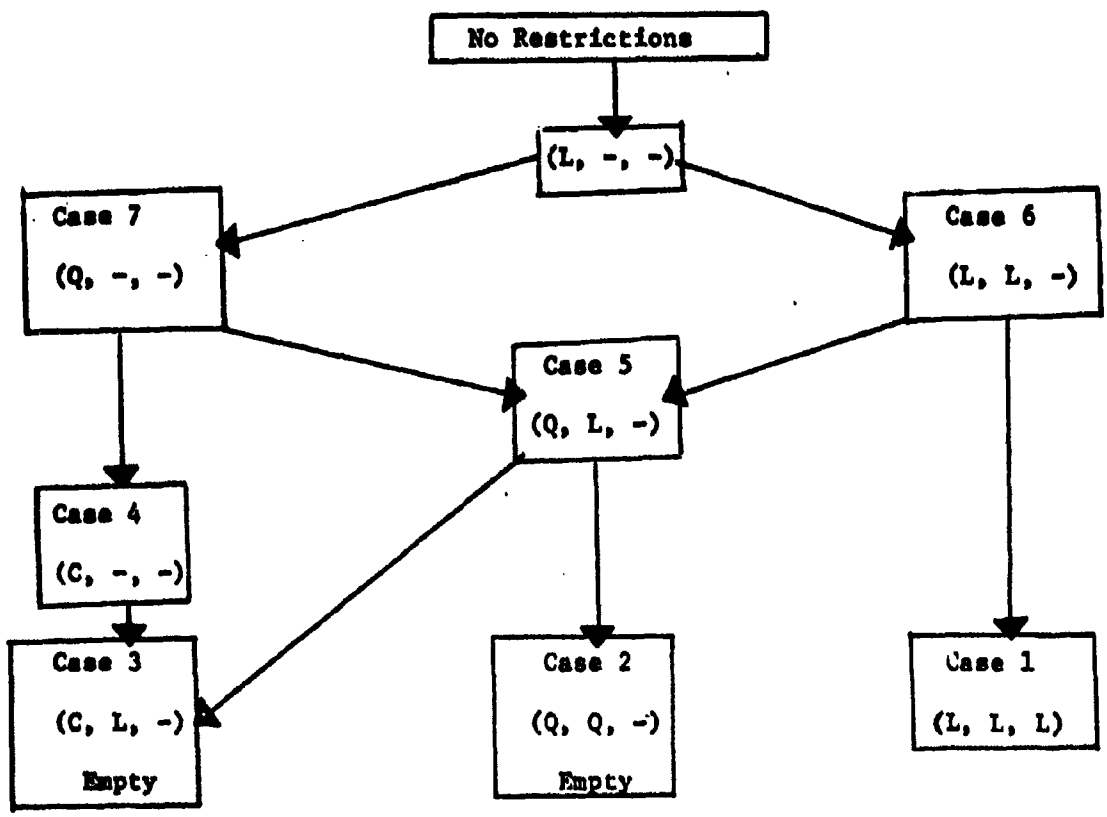
In order to be absolutely sure that the selected design is a valid design, the plus and minus signs of the main effects can be placed back through the standard Yates notation via the factor level changes as shown in Figure 10. In Figure 10 the design to be validated is given by the last row while the next to last row is the corresponding Yates notation from Figure 2. Here, a plus sign denotes a value of one and a minus sign denotes a value of zero. Thus, the Yates count is determined by writing the binary count of the five digit number of each row plus one. The Yates count for a valid design should include all numbers from 1 to 32.

Figure 10. To Validate The Selected Design

A*	B*	C*	D*	E*	Yates Count
+	+	+	-	+	30
+	-	-	+	-	19
-	-	-	+	-	3
-	+	+	-	+	14
-	+	-	+	-	11
-	-	+	-	+	6
+	-	+	-	+	22
+	+	-	+	-	27
-	-	-	+	+	4
-	+	+	-	-	13
+	+	+	-	-	29
+	-	-	+	+	20
+	-	+	-	-	21
+	+	-	+	+	28
-	+	-	+	+	12
-	-	+	-	-	5
-	-	+	+	-	7
-	+	-	-	+	10
+	+	-	-	+	26
+	-	+	+	-	23
+	-	-	-	+	18
+	+	+	+	-	31
-	+	+	+	-	15
-	-	-	-	+	2
+	+	-	-	-	25
+	-	+	+	+	24
-	-	+	+	+	8
-	+	-	-	-	9
-	+	+	+	+	16
-	-	-	-	-	1
+	-	-	-	-	17
+	+	+	+	+	32
BCDE	ABDE	ABCD	ABCDE	ABCE	
10	18	20	21	22	

6. CASE STUDY. Figure 1 summarizes seven cases with various time trend restrictions. Figure 11 shows how these cases or sets are included in each other. The case represented by (L, -, -) has a large number of elements or admissible designs as well as the case with no restrictions. Therefore, these two cases will not be analyzed but are shown in Figure 11 to complete the picture. As more restrictions are placed on the design, or as more arrows in Figure 11 are traced, the total number of factor level changes increases and the trade-off becomes a managerial decision. Note that Figure 11 is not drawn to any scale.

FIGURE 11. Inclusion of Cases



The logic for generating the admissible sets for the various cases has been programmed in FORTRAN. Table look-ups, "IF" statements, and "DO" loops simulate the sieve, the order constraints, and the restrictions and drive the sequential search.

CASE 1. (L, L, L). For this case the 5 treatments denoted "-" in Figure 6 must be designated as third or fourth order treatments. Therefore, up to an equivalence class, this set could have, at the most, 6 admissible designs. If 4 of the 5 possible third order interactions (treatments) are fixed then the fifth one is determined. Therefore, there are only 5 admissible designs and these five designs are displayed in Figure 12. In Figure 12 the 5 admissible designs are generated as follows. The first 4 treatments are fixed, thus determining the next 11 treatments. The treatments in line number 16 are fixed next, thus determining the rest of the treatments. The sum given in the last row characterizes each design and is found by adding the factor level changes or the values denoting the 5 main effects.

FIGURE 12. All Admissible Designs for Case 1

Line Number	Treatment	Design				
		1	2	3	4	5
1	ABCDE*	7	3	15	1	31
2	ABCD*	3	1	3	3	3
3	ABCE*	1	7	7	7	7
4	ABDE*	31	31	31	31	1
5	ABC*	5	5	11	5	27
6	ABD*	27	29	19	29	29
7	ABE*	25	27	23	25	25
8	CDE*	26	26	20	26	26
9	AB*	29	25	27	27	5
10	CD*	30	24	24	24	6
11	CE*	28	30	28	28	28
12	DE*	2	6	4	4	4
13	C*	24	28	16	30	30
14	D*	6	4	8	6	24
15	E*	4	2	12	2	28
16	ACDE*	15	15	1	15	15
17	B*	8	12	14	14	16
18	AE*	12	14	2	12	12
19	AD*	14	8	6	8	8
20	BC*	16	16	30	16	14
21	BDE*	10	10	10	10	20
22	BCE*	20	18	18	18	18
23	BCD*	22	20	22	22	22
24	A*	21	21	21	21	21
25	BCDE*	18	22	26	20	10
26	AE*	17	23	25	23	9
27	AD*	19	17	29	19	19
28	AC*	13	9	5	11	11
29	ADE*	23	19	17	17	17
30	ACE*	9	11	9	9	23
31	ACD*	11	13	13	13	19
Sum		63	67	71	73	119

CASE 2. (Q, Q, -). This case admits an empty set as shown as follows. Utilizing the first step of the seive, Figure 13 arrays the possible candidates as given by Figure 6 where each treatment of the triplet has an assigned Q or C. Using the ordering constraint, these triplets have been ordered in Figure 13. However, this ordering can be reversed if necessary. But the second step of the seive cannot be filled, since 6 of the 7 required treatments for each candidate at this step must be taken from Figure 13. Thus admitting an empty set.

FIGURE 13. Candidates for Case 2 from Step 1 of the Sieve

Triplets for A*, B*, and AB*		
5	17	20
5	18	23
5	19	22
9	18	27
9	19	26
9	20	29
10	17	27
10	19	25
10	23	29
11	17	26
11	18	25
11	22	29
13	20	25
13	22	27
13	23	27

CASE 3. (C, L, -). This case also admits an empty set. This can be shown in a similar fashion as shown in case 2 or by looking at the 5 treatments making up case 4 and putting on the further restriction on the first order interactions. To repeat the proof from case 2, Figure 14 arrays the possible candidates from the first step of the sieve. Note that in Figure 14 there are only 5 possible candidates for the main effects and the following product violates any possible designs:

$$10 * 18 * 20 * 22 = 26$$

That is, ABCD* and E* must be different. Thus showing that the set for case 3 is also empty.

FIGURE 14. Candidates for Case 3 from Step 1 of The Sieve

A*	B*	AB*
20	22	2
18	22	4
18	20	6
18	26	8
22	26	12
20	26	14
10	26	16
10	18	24
10	22	28
10	20	30

CASE 4. (C, -,-). For this case Figure 15 arrays the possible candidates from the first step of the sieve. Here there are only 6 possible candidates for the main effects, but one of these is inadmissible due to the following product violation:

$$10 * 18 * 26 = 2$$

$$20 * 22 = 2$$

This product violation is found an execution of steps 2 and 3 of the sieve. Figure 16 arrays the main effects and the first order interactions for the 5 admissible designs for this case along with the sum of the factor level changes. Figure 16 also shows that the set for case 3 is empty, since each design has at least one first order interaction that violates the further restriction imposed by going from case 4 to case 3.

FIGURE 15. Candidates for Case 4 from Step 1 of The Sieve

A*	B*	AB*
20	21	1
20	22	2
21	22	3
18	22	4
18	20	6
18	21	7
18	26	8
22	26	12
20	26	14
21	26	15
10	26	16
10	18	24
10	22	28
10	20	30
10	21	31

Figure 16. All Possible Designs for Case 4

Design	1	2	3	4	5
A*	10	10	10	10	18
B*	18	18	18	20	20
C*	20	20	21	21	21
D*	21	21	22	22	22
E*	22	26	26	26	26
AB*	24	24	24	30	6
AC*	30	30	31	31	7
AD*	31	31	28	28	4
AE*	28	16	16	16	8
BC*	6	6	7	1	1
BD*	7	7	4	2	2
BE*	4	8	8	14	14
CD*	1	1	3	3	3
CE*	2	14	15	15	15
DE*	3	15	12	12	12
SUM	91	95	97	99	107

Case 5. (Q, L, -). This case admits a very large set of admissible designs. Figure 17 displays some of these designs which were generated in a fraction of a second on the Univac 1108 computer along with the total sum of factor level changes. The designs with sums less than 70 were chosen to illustrate the possibilities.

Figure 17. Some Possible Designs for Case 5

A*	B*	C*	D*	E*	SUM
5	9	11	13	17	55
5	9	11	13	19	57
5	9	11	13	21	59
5	9	11	13	23	61
5	9	11	13	25	63
5	9	11	13	27	65
5	9	11	13	29	67
5	9	11	17	21	63
5	9	11	17	23	65
5	9	11	17	25	67
5	9	11	17	27	69
5	9	11	19	21	65
5	9	11	19	23	67
5	9	11	19	25	69
5	9	13	17	19	63
5	9	13	17	23	67
5	9	13	19	21	67
5	11	13	17	19	65
5	11	13	17	21	67

Case 6. (L, L, -). This case also admits a very large set of admissible designs, a set much larger than the set for case 5. Figure 18 displays some of these designs which were again generated in a fraction of a second on the Univar 1108 computer. The designs with sums less than 56 were chosen to illustrate the possibilities. The design with a sum of 43 is optimal. For comparitive purposes the standard Yates design has a sum of 57 plus non-zero time counts in the main effects.

Figure 18. Some Possible Designs for Case 6

A*	B*	C*	D*	E*	SUM
2	4	8	16	21	51
2	4	8	16	25	55
2	4	8	17	20	51
2	4	8	17	24	55
2	4	8	19	22	55
2	4	9	12	16	43
2	4	9	12	17	44
2	4	9	12	18	45
2	4	9	12	20	47
2	4	9	12	21	48
2	4	9	12	23	50
2	4	9	12	24	51
2	4	9	12	25	52
2	4	9	12	26	53
2	4	9	12	28	55
2	4	9	16	21	52
2	4	9	16	24	55
2	4	9	17	20	52
5	9	11	13	16	54
5	9	11	13	17	55
5	8	12	14	16	55

Case 7. (Q, -, -). This case is included for comparison purposes. Although it's much larger than cases 4 and 5, it turns out that it has the same optimal design as case 5 as given by the first design of Figure 17.

To compare these cases further, the optimal design for the case expressed by (L, -, -) is given as (2, 4, 5, 8, 16) with a sum of 35. Further, the case or set of designs having no restrictions is given as (1, 2, 4, 8, 16) with a sum of 31 or N-1 as shown by the references. However, on restricting to the standard Yates notation, as this paper has done, this is the only possible design up to an equivalence class, with a sum of 31. On relaxing the standard Yates restriction, as the references do, many designs can be found with a sum of 31, but with non-zero time counts.

7. APPLICATIONS. The application of the techniques presented in this paper to operational testing can best be shown by giving an example. For that purpose, an experimental design for an operational test of the hypothetical ZAP anti-tank weapon will be constructed.

After analysis of the system to be tested, five factors are chosen to be included in the design, each factor being taken at two levels, thus giving a 2^5 factorial experiment. The factors chosen and their associated levels are shown in Figure 19.

The importance of eliminating time trends in such a test can easily be seen. With so few factors being controlled, there exist the possibility that some uncontrolled and unmeasured factor is influencing test results. Such factors as weather, crew learning, and crew morale can, and usually do, change with time through the test.

Another consideration in designing this test is the ease of execution of the design. Quite often a penalty must be paid in time, money, and perhaps test validity for each factor level change which is made. For instance, changing the visability factor between day and night too often would greatly slow the test execution and destroy any attempt to portray a realistic combat scenario, as it would permit only a small number of firings during daylight and then delay further testing until night in order to achieve the desired factor level change. Similarly it may be difficult and time consuming to frequently move the test participants and test team from one location to another in order to achieve changes in the terrain factor. As a third example, frequent changes in the weapon factor may confuse the test participant and prevent him from performing as well as he might if he were allowed to stay with one weapon. For example, one weapon may require the soldier to lead a moving target while the other weapon does not. If the test participant is frequently switching back and forth, he may forget and lead when he should not or not lead when he should. Even if he does remember and does the right thing, he may not do it as proficiently as if he had been able to concentrate on developing a single skill instead of two.

With the foregoing constraints in mind, we can use the techniques presented in this paper to design a good test of our hypothetical anti-tank system.

If it is felt desirable to strongly protect the main effects, we could choose case five which eliminates linear, and quadratic time trends for the main effects and linear time trends for the first order interactions. To construct our design we select one of the admissible run orders found for case five, as given in Figure 17. This selection can either be made randomly or the one with the minimum total number of factor level changes can be chosen. For our example, let us choose the design which minimizes the factor level changes. We can then construct our experimental design by going back to the standard Yates notation and writing out the level changes

Figure 19. Operational Test of the ZAP Anti-tank Weapon

FACTORS	LEVELS
. Weapon	- Baseline (present anti-tank weapon)
	- Candidate (ZAP)
. Range	- Short
	- Long
. Visibility	- Night
	- Day
. Target Mode	- Stationary
	- Moving
. Terrain	- Open
	- Forrest

for the five factors as defined by the level change numbers given in Figure 17. This design is given in Figure 20. As with the selection of a design from the set of admissible run orders, the assignment of the five factors to the five columns can be done either randomly or by ordering the factors based on which factor should have the fewest level changes and which could have more level changes.

Suppose after examining Figure 20 we feel this design is not desirable because the number of factor level changes for visiblility, weapon, and terrain are excessive for the reasons discussed in paragraph 4 of this chapter. One alternative would be to relax the constraints on the elimination of higher order time trends. We could decide to select a design which eliminates only linear time trends for the main effects, and first order interactions. For this we can choose case six. Figure 18 gives admissible run orders for case six. Going through the same procedure as for case five, we come up with the design given in Figure 21.

Given that this design is determined to be staisfactory, it only remains to randomly assign a plus or minus to the actual level names for each factor. For ease of planning the conduct of the test, it may prove convenient to display the design information of Figure 21 in a more conventional format as shown in Figure 22 where the number in each cell gives the order of execution of each test event in filling out the full factorial design.

Figure 20. Case 5 Candidate Design for the ZAP Test

	A 5 TERRAIN	B 9 VISIBILITY	C 11 WEAPON	D 13 TGT MODE	E 17 RANGE
1	-	-	-	-	-
2	-	-	-	-	+
3	-	+	+	+	+
4	-	+	+	+	-
5	+	+	+	-	-
6	+	+	+	-	+
7	+	-	-	+	+
8	+	-	-	+	-
9	+	-	+	+	-
10	+	-	+	+	+
11	+	+	-	-	+
12	+	+	-	-	-
13	-	+	-	+	-
14	-	+	-	+	+
15	-	-	+	-	+
16	-	-	+	-	-
17	-	+	-	+	+
18	+	+	-	+	-
19	+	-	+	-	-
20	+	-	+	-	+
21	-	-	+	+	+
22	-	-	+	+	-
23	-	+	-	-	-
24	-	+	-	-	+
25	-	+	+	-	+
26	-	+	+	-	-
27	-	-	-	+	-
28	-	-	-	+	+
29	+	-	-	-	+
30	+	-	-	-	-
31	+	+	+	+	-
32	+	+	+	+	+

Figure 21. Case 6 Candidate Design for the ZAP Test

	A 2 TERRAIN	B 4 VISIBILITY	C 9 WEAPON	D 12 TGT MODE	E 16 RANGE
1	+	+	-	+	+
2	+	+	-	+	-
3	+	+	+	-	-
4	+	+	+	-	+
5	+	-	+	+	+
6	+	-	+	+	-
7	+	-	-	-	-
8	+	-	-	-	+
9	-	-	-	-	+
10	-	-	-	-	-
11	-	-	+	+	-
12	-	-	+	+	+
13	-	+	+	-	+
14	-	+	+	-	-
15	-	+	-	+	-
16	-	+	-	+	+
17	-	+	+	+	+
18	-	+	+	+	-
19	-	+	-	-	-
20	-	+	-	-	+
21	-	-	-	+	+
22	-	-	-	+	-
23	-	-	+	-	-
24	-	-	+	-	+
25	+	-	+	-	+
26	+	-	+	-	-
27	+	-	-	+	-
28	+	-	-	+	+
29	+	+	-	-	+
30	+	+	-	-	-
31	+	+	+	+	-
32	+	+	+	+	+

Figure 22. Case 6 Final Design Matrix for the ZAP Test

	TRN	FOREST (-)				OPEN (+)			
	TGT MODE	MOVING (-)		STATIONARY (+)		MOVING (-)		STATIONARY (+)	
VIS	RNG WPN	LONG (-)	SHORT (+)	LONG (-)	SHORT (+)	LONG (-)	SHORT (+)	LONG (-)	SHORT (+)
NIGHT (-)	BASE-LINE (-)	10	9	22	21	7	8	27	28
	CAND-IDATE (+)	23	24	11	12	26	25	6	5
DAY (+)	BASE-LINE (-)	19	20	15	16	30	29	2	1
	CAND-IDATE (+)	14	13	18	17	3	4	31	32

8. FUTURE WORK. The computer logic for recursively generating factorial designs having more than five factors would be desirable. Admissible designs with a mix of two and three level factors would be more realistic. Of further concern would be optimal fractional factorial designs.

9. REFERENCES.

(1) Addelman, Sidney, "Recent Developments in the Design of Factorial Experiments," Journal of the American Statistical Association, volume 67 (March 1972), 103-111.

(2) Daniel, Cuthbert, and Wilcoxon, Frank, "Factorial 2^{p-q} Plans Robust Against Linear and Quadratic Trends," Technometrics, 8 (May 1966), 259-278.

(3) Dickinson, A. W., "Some Run Orders Requiring a Minimum Number of Factor Level Changes for the 2^4 and 2^5 Main Effect Plans," Technometrics 16 (February 1974), 31-37.

(4) Draper, N. R., and Stoneman, D. M., "Factor Changes and Linear Trends in Eight-Run Two-Level Factorial Designs", Technometrics, 10 (May 1968), 301-311.

(5) Thomas, Jerry, "Application of Statistical Techniques to Model Sensitivity Testing," US Army Concepts Analysis Agency, TP-T4-10, (1974), 23 pages.

(6) Tiahrt, K. J., and Weeks, P.L., "A Method of Constrained Randomization for 2^n Factorials", Technometrics, 12 (August 1970), 471-486.

A METHOD OF ESTIMATING ERROR VARIANCE IN A NON-REPLICATED EXPERIMENT BY PARTITIONING AN INTERACTION TERM INTO NON-ADDITIVITY AND ERROR

1st Lieutenant L. Douglas Peirce
Instructor, School of Logistics Science
Army Logistics Management Center
Fort Lee, Virginia

H. Gill Hilton
Chairman, Department of Statistics
Brigham Young University
Provo, Utah

ABSTRACT. A method of estimating error variance in a non-replicated experiment by separating an interaction term into sums of squares of non-additivity and sums of squares pertaining to error was examined. A sequential procedure to test individual degrees of freedom of the interaction term for non-additivity was introduced. Five test statistics that could be applied to the sequential procedure are given. The critical values needed for each of the test statistics for $\alpha = 0.05$ and 0.15, for 10, 20, and 30 degrees of freedom respectively in the term being tested, and for three stages of the sequential procedure were estimated by Monte Carlo methods.

The five test statistics were compared as to their power and ability to estimate error variance when non-additive individual sums of squares were combined with individual sums of squares that estimated error variance. The results and recommendations as to which is the best test statistic are given. The data indicated that using a higher level of significance than 0.15 would better estimate error variance.

1. INTRODUCTION. Frequently, due to the nature of an experiment or through poor planning, a design is formed without replication. When this happens the experimenter has no estimate of experimental error in his data. This situation is illustrated in Table 1 taken from Fisher (1951). Since each entry in this table represents a single observation, there is no way to estimate experimental error. The usual

TABLE 1

Total Yields of Barley Varieties in Twelve Independent Trials

Place and Year		Manchuria	Svansota	Velvet	Trebi	Peatland
1	1931	81.0	105.4	119.7	109.7	98.3
	1932	80.7	82.3	80.4	87.2	84.2
2	1931	146.6	142.0	150.7	191.5	145.7
	1932	100.4	115.5	112.2	147.7	108.1
3	1931	82.3	77.3	78.4	131.3	89.6
	1932	103.1	105.1	116.5	139.9	129.6
4	1931	119.8	121.4	124.0	140.8	124.8
	1932	98.9	61.9	96.2	125.5	75.7
5	1931	98.9	89.0	69.1	89.3	104.1
	1932	66.4	49.9	96.7	61.9	80.3
6	1931	86.9	77.1	78.9	101.8	96.0
	1932	67.7	66.7	67.4	91.8	94.1

solution to this problem is to assume an additive model (no interaction) and to use the residual sum of squares as an estimate of error. In a model with two main effects this means renaming the two-way interaction as error. For the data in Table 1 the three-way interaction alone may be pooled into error or possibly the three-way and one or both of the two-way interactions may be pooled depending upon the experiment and the analyst. Having an estimate of the error the experimenter may now be able to test other terms in the model that weren't testable before pooling.

The problem with this procedure is that some of the pooled sums of squares may have estimated interaction and not error. If this happens, the estimate of the error will be too large giving the experimenter a less sensitive test of other terms in the model.

How, then, can it be determined if the mean square of an interaction term estimates error, interaction, or both? This paper examines five test statistics that are designed to answer this question. It will be restricted to fixed models with one observation per cell. The techniques developed can be applied to any or all interaction terms in any n-way model.

Using the Modified Abbreviated Doolittle (MAD) computer routine developed by Bryce (1970), the terms of a fixed model can be broken into single degree of freedom sums of squares. These single degree of freedom sums of squares form the building blocks of the five test statistics. The individual sums of squares of an interaction term are ranked and sequentially tested one at a time starting with the largest until non-significance is declared. At this point, the significant single degree of freedom sums of squares are pooled together as the part estimating interaction and the rest of the sums of squares and their corresponding degrees of freedom are pooled into error which is hopefully free of interaction.

This paper will compare the ability to find interaction when present, or power, of the five test statistics and the ability of each to estimate σ^2.

2. TEST PROCEDURE. The expected mean square of any interaction term can be broken into two parts. The first part contains the error variance, σ^2, and the second part contains the sum of the remaining different possible variance components. The number of terms in the second part would depend on the ANOVA model. If interaction exists, then the mean square of an interaction term estimates the sum of the two parts of the expected mean square; i.e., σ^2 plus the rest of the terms. However, if interaction does not exist, the mean square estimates only the error variance. If for a given model interaction is not present, it would be appropriate to pool the sums of squares and degrees of freedom associated with the interaction terms into the error term.

The sum of squares and n degrees of freedom of a term in the model can be partitioned into n sums of squares, each associated with one degree of freedom. If an interaction term is so partitioned, the resulting single degree of freedom sums of squares estimate either error variance or interaction. It would be desirable to extract the portion that estimates error only, thus giving an estimate of σ^2 and making it possible to test other terms in the model. This procedure assumes that some of the partitioned single degree of freedom sums of squares estimate σ^2 only and that not all estimate interaction.

The steps for the proposed sequential procedure for testing any interaction term and estimation of σ^2 are:

1. Separate the term with n degrees of freedom into n sums of squares containing one degree of freedom each.

2. Rank the n sums of squares.

3. Apply one of the test statistics to the largest sum of squares.

4. Check for significance using the appropriate values in the table for α and stage. (Stage is the number of the sequential test that is being performed on the individual sums of squares of an interaction term. For example, stage one is the test of the largest individual sum of squares, stage two the second largest and so on.)

5. If significance is declared, return to step three using the same test statistic and significance level to test the next largest sum of squares. If no significance is found, proceed to step six.

6. Pool the significant sums of squares and degrees of freedom into one interaction term.

7. Pool the remaining sums of squares with their appropriate degrees of freedom into error.

3. TEST STATISTICS. The proposed test statistics will be labeled F1, F2, F3, F4, and F5 for convenience and the sum of squares of a single degree of freedom interaction term will be written as S_i where $(S_1 < S_2 < \ldots < S_n)$. The stage in the sequential test procedure will be denoted by r and n will denote the degrees of freedom in the interaction term before testing.

The test statistics are:

$$F1 = \frac{\sum_{i=n-r+1}^{n} \frac{S_i}{r}}{\sum_{j=1}^{n-r} \frac{S_j}{n-r}}$$

$$F2 = \frac{S_{n-r+1}}{S_1}$$

$$F3 = \frac{S_{n-r+1}}{\sum_{i=1}^{n} S_i}$$

$$F4 = \frac{\sum_{i=n-r+1}^{n} \frac{S_i}{r}}{\sum_{j=1}^{n} S_j}$$

$$F5 = \frac{S_{n-r+1}}{\sum_{i=1}^{n-r} S_i}$$

F1 could be described as the sums of squares having been declared significant plus the test sum of squares (the individual sum of squares being tested for significance) averaged and divided by the average of the remaining sums of squares. F2 is the test sum of squares devided by the smallest sum of squares. F3 is the test sum of squares divided by the total sums of squares of the interaction term. F4 is a composite of F1 and F3. F5 is the numerator of F3 divided by the sum of the sums of squares less than the test sum of squares.

4. GENERATION OF CRITICAL VALUES. The sequential test procedure was developed to test the hypothesis of no interaction present in the single degree of freedom sum of squares of any interaction term. This would mean that each of the single degree of freedom interaction sum of squares estimate error and follow a central chi-square distribution with one degree of freedom. The null hypothesis for the test procefure at the first stage could be written

$$H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$$

where λ_i represents the non-centrality parameter of the chi-square associated with each of the ordered single degrees of freedom. If the test proceeds to the second stage the null hypothesis would be

$$H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_{n-1} = 0$$

and so on at other stages of the test.

Under the null hypothesis it is possible to generate the critical values for each test statistic using one degree of freedom central chi-squares. Two parameters affect the shape of the distribution of each test statistic; the stage of the test and the number of degrees of freedom in the interaction term under consideration. Using an electronic computer, the distributions of each of the test statistics were simulated for three stages and interaction terms of ten, twenty, and thirty degrees of freedom. The upper portion of the distributions were ordered and the five and fifteen percent points were found thereby giving an estimate of the 0.05 and 0.15 critical values under the null hypothesis.

The single degree of freedom chi-squares were formed by generating a standard normal value and squaring it. Each

standard normal was generated by the Box-Muller (1958) transformation using uniform values generated by the McGill Random Number Generator Package, supplied by McGill University. This method of generating standard normals was found satisfactory by Thomas (1975).

A more detailed explanation of how the critical values were found for stage one and ten degrees of freedom of interaction will now be given. Ten one-degree of freedom central chi-aquares were generated and ordered. A value for each of the five test statistics was calculated and saved. This process was repeated ten thousand times. The upper portion of the ten thousand values for F1 was ordered and the five percent and fifteen percent points were found. This gave the estimated critical values for a stage one test of an interaction term containing ten degrees of freedom using F1 as a test statistic. The critical values were found in the same manner for F2, F3, F4, and F5. This process was repeated for twenty and thirty degrees of freedom in interaction.

Stage two critical values for ten degrees of freedom interaction terms and $\alpha = 0.05$ were estimated by again generating values for the test statistics in the same manner as above. If generated numbers of the test statistics exceeded the 0.05 critical values with ten degrees of freedom for interaction at stage one, the test statistic for stage two was formed and saved. This was repeated until two thousand values at stage two were accumulated. The upper portion was ranked and the estimate of the 0.05 critical value for stage two was found. The same procedure was followed to find the table values for $\alpha = 0.15$ and so on for twenty and thirty degrees of freedom of interaction.

The calculation of stage three critical values is an extension of the stage two procedure. Critical values under the null hypothesis were calculated and if they exceeded the appropriate critical values of both stage one and stage two the test statistic for stage three was formed and saved until two thousand were accumulated. They were then ordered as before and the estimates of the five percent and fifteen percent critical values were found. The complete table of critical values generated is found in Table 2. The critical values do not extend past stage three because of the length of computer time that would be necessary to generate stage four critical values.

TABLE 2

CRITICAL VALUES FOR F1, F2, F3, F4, and F5

n is the total degrees of freedom associated with the interaction term being tested.

α is the level of significance.

TEST STATISTIC	n	α	Stage		
			1	2	3
F1	10	.05	13.7882	35.6391	108.8423
		.15	9.0107	19.6191	46.5067
	20	.05	12.0037	20.3695	32.4610
		.15	8.9826	13.8655	19.9743
	30	.05	11.9484	17.4037	23.5462
		.15	9.1645	12.7221	15.9907
F2	10	.05	84376.4338	14924046.1125	4099285578.0629
		.15	8119.5734	190156.7131	4723313.6463
	20	.05	421750.6897	157984650.5641	42188251909.4520
		.15	4273.3229	1462966.2431	51961752.0038
	30	.05	1060700.3502	330771314.1788	97140244926.5100
		.15	108524.4512	5007498.1783	183702308.8679
F3	10	.05	.6051	.2182	.0911
		.15	.5003	.2283	.1112
	20	.05	.3872	.2258	.1369
		.15	.3210	.2093	.1357
	30	.05	.2918	.1986	.1389
		.15	.2401	.1788	.1279
F4	10	.05	.6051	.4495	.3263
		.15	.5003	.4153	.3174
	20	.05	.3872	.3474	.2838
		.15	.3210	.3032	.2597
	30	.05	.2918	.2771	.2412
		.15	.2401	.2380	.2133
F5	10	.05	1.5231	1.7930	2.1168
		.15	.9985	1.1222	1.3075
	20	.05	.6365	.6677	.7151
		.15	.4741	.4849	.5207
	30	.05	.4129	.4315	.4443
		.15	.3177	.3190	.3302

5. CHOICE OF α. It may be desirable to make the test for interaction at a relatively small alpha rather than a large one. A small α under H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$, may lead to an inflated estimate of σ^2 by way of the sequential test because when no significance is found the test procedure is halted and the error sum of squares is calculated. A test using a small alpha may not find interaction when it is present thus leading to an inflated estimate of σ^2. Therefore, any tests of other factors in the model using the inflated error would be conservative. With this in mind, critical values for alpha equal to 0.05 and 0.15 were estimated.

It should be noted that the level of significance must remain the same at all stages of the test when using the critical values developed here. For example, it is not appropriate to test at stage one using $\alpha = 0.15$ and after finding significance to test at stage two using $\alpha = 0.05$.

6. GENERATION OF POWER DATA. Power in a sequential test is an elusive concept. For this reason, power at stage one is defined to be the probability of rejecting the null hypothesis, H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$, given the null hypothesis is false. Power at stage two is the probability of rejecting the null hypothesis, H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_{n-1} = 0$, given the null hypothesis is false.

Data generated to compare the power of the five test statistics were divided into two cases. Case one consisted of generating ten, twenty, or thirty standard normal deviates, adding a single non-centrality parameter, λ_i, to one of these at random, and squaring each. The result was one non-central and (n-1) central chi-squares. The sequential test procedure was then performed using one of the test statistics at a level of significance α. This was repeated one thousand times adding the same non-centrality parameter, λ_i, to a new set of standard normal deviates and keeping a record of the number of times significance was declared. An estimate of power for the test statistic, at α, n degrees of freedom for interaction, and λ_i at stage one was calculated by dividing the number of times significance was declared by one thousand. The above process was repeated for every possible combination of test statistics, levels of significance, number of degrees of freedom for interaction, and non-centrality parameters. The non-centrality parameters are

$\lambda_1 = 1.5$, $\lambda_2 = 2.5$, $\lambda_3 = 3.5$, and $\lambda_4 = 4.5$. The sequential test for power in case one was not carried past the first stage. The experiment was repeated once to form an estimate of experimental error.

A test for power at both stage one and stage two was performed in case two data. n random standard normal deviates were again generated and non-centrality parameters were added to two randomly selected standard normals before squaring. The sequential test was applied and the process repeated one thousand times keeping count of the total number of times significance was declared. Each time significance was found the test would proceed to stage two to test for significance and a tally was kept of the number of times the null hypothesis was rejected.

For a certain α, test statistic, n degrees of freedom of interaction, and set of non-centrality parameters, power at stage one was the number of times significance was found divided by one thousand while power at stage two equaled the number of times the null hypothesis was rejected at stage two divided by the total number of tests made. (The total number of tests made at stage two was the number of times significance was declared at stage one.)

The above power for case two was calculated independently for each combination of degrees of freedom of interaction, test statistics, levels of significance, and pairs of non-centrality parameters. As in case one, the experiment was replicated once. There were ten different pairings of λ_i, λ_j added to form non-central chi-squares. These are listed in Table 3.

Table 3

Pairings of Non-centrality Parameters
Added for Case Two Power

λ_1	λ_2
1.5	1.5
2.5	2.5
3.5	3.5
4.5	4.5
1.5	2.5
1.5	3.5
1.5	4.5
2.5	3.5
2.5	4.5
3.5	4.5

7. GENERATION OF MEAN SQUARE ERROR DATA. As the above procedure for power was being performed data for an analysis of the ability of the test statistics to estimate σ^2 was also being compiled.

As each set of ten, twenty, or thirty chi-squares was generated for case one data, the test procedure would check for significance at different stages until none was found. It would then tally the sum of squares and degrees of freedom to be pooled into error. This would proceed until all one thousand sets were tested. The estimate of σ^2 was then calculated by dividing the total sums of squares pooled into error by the pooled degrees of freedom. If significance was found at each of the first three stages in any of the one thousand sets, (n-3) degrees of freedom and the sums of squares not declared significant were added to error. Since these data were calculated simultaneously with the power there are two independent observations for all combinations of test statistics, degrees of freedom in interaction, non-centralities, and levels of significance. The case one mean square error data were calculated for five λ_1, four being the same as in the power analysis and the fifth being equal to zero.

Mean square error data for case two were generated simultaneously with case two power data. As both a stage one power test and stage two power test were performed for case two data, mean square error data were also collected at both the stage one power test and stage two power test. Case two mean square error data will be labeled and discussed in terms of stage of power test. This avoids the problem of thinking of the MSE data as "stage one MSE" and "stage two MSE" which carries the wrong connotation since both errors are estimated using the three-stage sequential procedure.

Mean square error data at stage one power test were collected as follows. The sequential (up to three stages) procedure was applied to each set of n single degree of freedom interaction sum of squares. If non-significance occurred at stage one all n sums of squares were pooled into the error estimate. When significance was declared at stage one but not at stage two (n - 1) sums of squares were pooled into the error estimate and with significance at stages one and two but not at stage three (n - 2) sums of squares were pooled into the error estimate. It was decided if significance was

found at all three stages that the remaining (n - 3) sums of squares would be pooled into error. Thus, each of the one thousand sets of n sums of squares contributed something to the estimate of error.

Mean square data at stage two power test were collected in a different manner than at stage one power test. The same three stage sequential procedure was applied, but only to those sets of n sums of squares which were declared significant at the stage one power test. If non-significance was observed at the stage one power test, then the set of n sums of squares did not become a part of the error estimate at the stage two power test. Thus fewer than one thousand sets of n sums of squares were used in the stage two power test estimate. One might say that the mean square error calculated at stage two power test is "adjusted" for those cases where non-significance was found at stage one power test.

This procedure was repeated for each combination of n, F, α, and pairings of λ_i, λ_j. The entire process was replicated so that two independent estimates of error were obtained at each design point.

The mean square error data at stage one power test are the values of interest in this paper. They will be larger than the mean square error values calculated at stage two power test because the sums of squares and degrees of freedom are pooled into the mean square error at stage two power test only if significance was found at stage one power test. This means that the largest, individual sum of squares that is not declared significant at stage one is never pooled into the mean square error at stage two power test. If one decided to estimate σ^2 only when significance was found at the first stage of the sequential procedure then the values of mean square error at stage two power test would give a picture of the results one might expect from the test statistics. However, if one wanted an estimate of σ^2 independent of significance being declared at stage one of the sequential procedure the mean square error data generated at stage one power test one will indicate which is the best test statistic.

8. METHOD OF ANALYSIS OF DATA. Analysis of variance was used to analyze the data generated for case one power. A four-way factorial model complete with all interactions was formed using degrees of freedom of interaction (n), test

statistic (F), non-centrality parameter (λ), and significance level, (α), for the four main effects. Degrees of freedom of interaction had three levels (ten, twenty, and thirty), test statistics had four levels (F1, F2, F3, and F5), non-centrality parameters had four levels (1.5, 2.5, 3.5, and 4.5), and alpha had two levels (0.05 and 0.15). F4 was left out of the analysis in case one because power wasn't extended past stage one and at stage one F3 and F4 are the same test statistic. The main effects for this model and for all models in this paper were considered fixed.

The dependent variable in the power analysis is a proportion. In case one data one thousand independent tests for power were made for each combination of n, F, α, and λ. The proportion was formed by dividing the number of times the null hypothesis was rejected by the total number of tests made.

Because of the range of non-centralities used to generate the data, it is possible that the assumption of homogeneous variance in each cell is violated. For this reason, the arc-sine transformation, as described by Snedecor and Cochran (1967), was used on the data but very little difference was found between the analysis of the raw data and that of the transformed data so the analysis of the raw data was used.

Case two power data were analyzed using a five-way factorial model. The five main effects were degrees of freedom for interaction (ten, twenty, and thirty), alpha (0.05 and 0.15), test statistic (F1, F2, F3, F4, and F5), non-centralities (the ten pairs in Table 3), and stage (stage one and stage two). The number of binomial results going into each observation of case two power data varied with stage. At stage one, one thousand binomial results went into each observation while at stage two the number of binomial results that went into each observation were the number of times significance was declared out of the one thousand trials at stage one. This is because the sequential test procedure doesn't proceed to stage two unless significance occurs at stage one. Analysis was performed on the raw data and also a weighted arc-sine transformation of the data, weighted by the number of binomial results making up each observation. Very little difference was found in the results between the two analyses and so only the analysis of the raw data will be considered here.

Before describing the method of analyzing the mean square error data, consideration of what would be the best estimate of mean square error by a test statistic in this paper will be made. Ideally, the test statistic would identify any single degree of freedom sums of squares that have interaction in them and pool into error only the sums of squares that truly estimate error. Each single degree of freedom that estimates error is a central chi-square with one degree of freedom and with expected value equal to one. Since the expectation of a sum of central chi-squares is equal to the sum of their degrees of freedom, the expected value of the pooled sum of squares of error when all interactions have been extracted by the test statistic is equal to the pooled degrees of freedom. The expected mean square error would then be equal to one. If the test statistic fails to remove all of the interaction the expected mean square would be greater than one. If the test statistic using the sequential procedure pools only part of the single degree of freedom sums of squares that estimate σ^2 into error the resulting mean square error would be less than one on the average. This is because the sums of squares of error left in interaction would be the largest sums of squares, not just any sums of squares selected at random, leaving the smaller for error thus decreasing the expected value of mean square error. Hence, for the data generated here, the ideal test statistic would yield an estimate of error having an expected value equal to one.

Analysis of variance was also used to analyze the mean square error data of case one and case two. Although heterogeneity of variance exists, since the observations are central or non-central chi-squares, Scheffe' (1959) notes that if an analysis is balanced the heterogeneity of variance has little consequence. This was seen in the analysis of the raw and transformed power data. The analysis of case one and case two mean square error data was performed on the untransformed dependent variable using the error estimate produced by replication to test terms in the model.

The ANOVA model for case one and case two mean square error were the same as for power with three exceptions. F4 was added to the levels of the main effect for test statistics in case one since it will estimate mean square error differently than F3. Zero was added to the levels of the main effect for non-centralities to investigate the ability of the test statistic to estimate σ^2 when no interaction is present.

The authors of this paper subscribe to the philosophy that when it is not desirable or possible to control main effects in an experiment it is proper to test for significance among the levels of main effects in the presence of interaction. This also applies to the testing of low ordered interactions in the presence of significant higher ordered interactions. The analyst must realize, however, that the main effects and low ordered interactions have been averaged over all other factors in the model and any interpretation of significance must be viewed in this light.

The analysis of the power and mean square error data will be discussed a case at a time instead of discussing power completely and then mean square error.

9. RESULTS AND DISCUSSION OF CASE ONE DATA. Table 4 is the analysis of variance table for case one power data and Table 5 is the table for case one mean square error data. Significance was found for almost every term.

The first thing to be considered is alpha. Figure 1 contains graphs of power and mean square error for the F by α interaction.

The graph of power in Figure 1 indicates that the power is better using a larger alpha which is not surprising, but the graph of mean square error shows that a better estimate of mean square error is obtained using $\alpha = 0.15$ since the line for $\alpha = 0.15$ is closer to one than that for $\alpha = 0.05$. Table 5 shows significance for main effect α which indicates that using $\alpha = 0.15$ for case one data gives a better estimate of mean square error.

Now consider Figure 2 which contains graphs for the power and mean square error of the F by λ by $\alpha = 0.15$ interaction term.

There is no significant difference between the power curves of F1, F3, F4, and F5 so power offers no help as to which test statistic is the best other than that the power of F2 is lacking. The graph of mean square error in Figure 2 shows that F2 also lacks in ability to estimate mean square error. There is no practical difference between the points of F1, F3, F4, and F5 for mean square error at $\lambda = 0$, 1.5, 2.5. At $\lambda = 3.5$, F3 is significantly higher than the

TABLE 4

ANOVA Table for Case One Power Data

Source	DF	MS	F
n	2	0.0030	23.5678
F	3	1.3084	9997.0462
nF	6	0.0009	7.3491
α	1	1.0141	7748.2736
nα	2	0.0026	20.5578
Fα	3	0.0001	0.8867*
nFα	6	0.0002	1.7236*
λ	3	3.0563	23351.5011
nλ	6	0.0010	8.1844
Fλ	9	0.2475	1891.2634
nFλ	18	0.0004	3.5115
αλ	3	0.0097	74.6679
nαλ	6	0.0002	2.1278*
Fαλ	9	0.0019	15.0068
nFαλ	18	0.0001	1.1544*
ERROR	96	0.0001	

* Indicates that the term was not significant at the .05 level. No * by the F value indicates significance was declared at the .05 level.

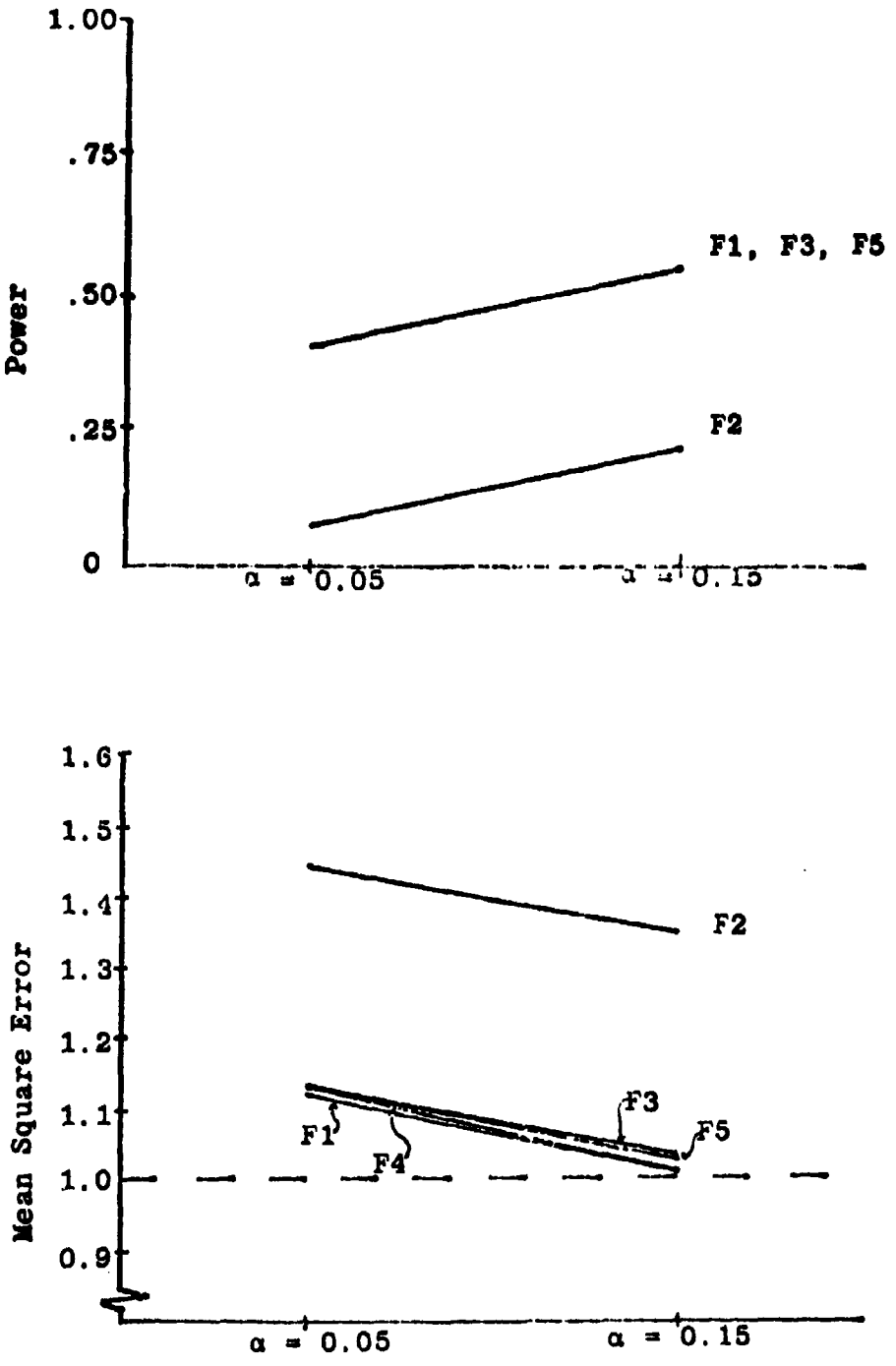


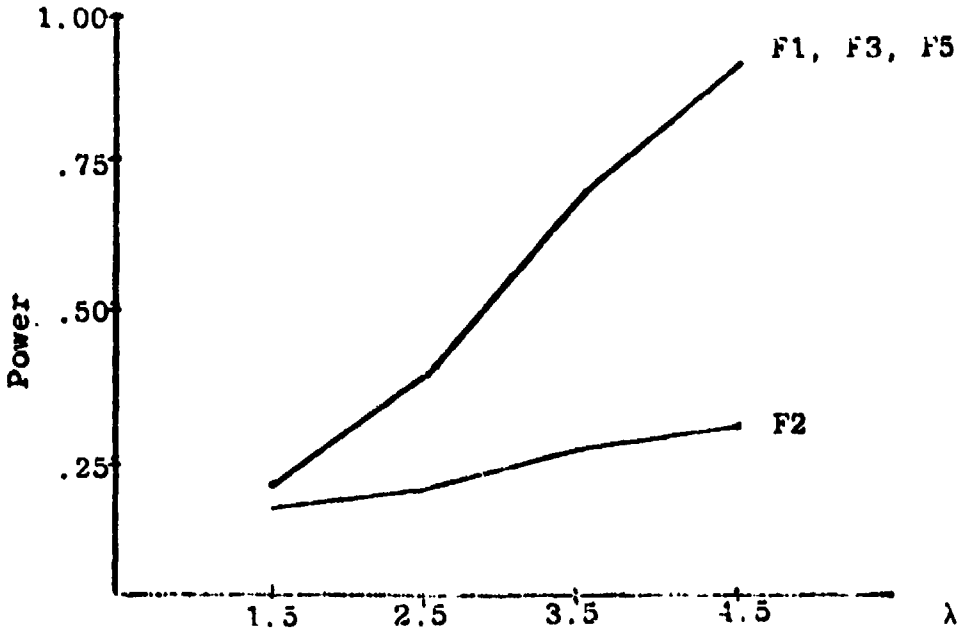
Figure 1. Power vs. $\alpha = 0.05$ and $\alpha = 0.15$ and MSE vs. $\alpha = 0.05$ and $\alpha = 0.15$ at Case One for F1, F2, F3, F4, and F5

TABLE 5

ANOVA Table for Case One Mean Square Error Data

Source	DF	MS	F*
n	2	1.1613	4570.3185
F	4	1.2276	4831.2999
nF	8	0.0866	341.0213
α	1	0.7622	2999.7058
nα	2	0.0759	299.0747
Fα	4	0.0010	4.2709
nFα	8	0.0005	2.1856
λ	4	0,8472	3334.1864
nλ	8	0.1453	571.9092
Fλ	16	0.3739	1471.8028
nFλ	32	0.0295	116.1555
αλ	4	0.0517	203.4774
nαλ	8	0.0077	30.3502
Fαλ	16	0.0026	10.4372
nFαλ	32	0.0005	2.0556
ERROR	150	0.0002	

* All tests are significant at the .05 level.



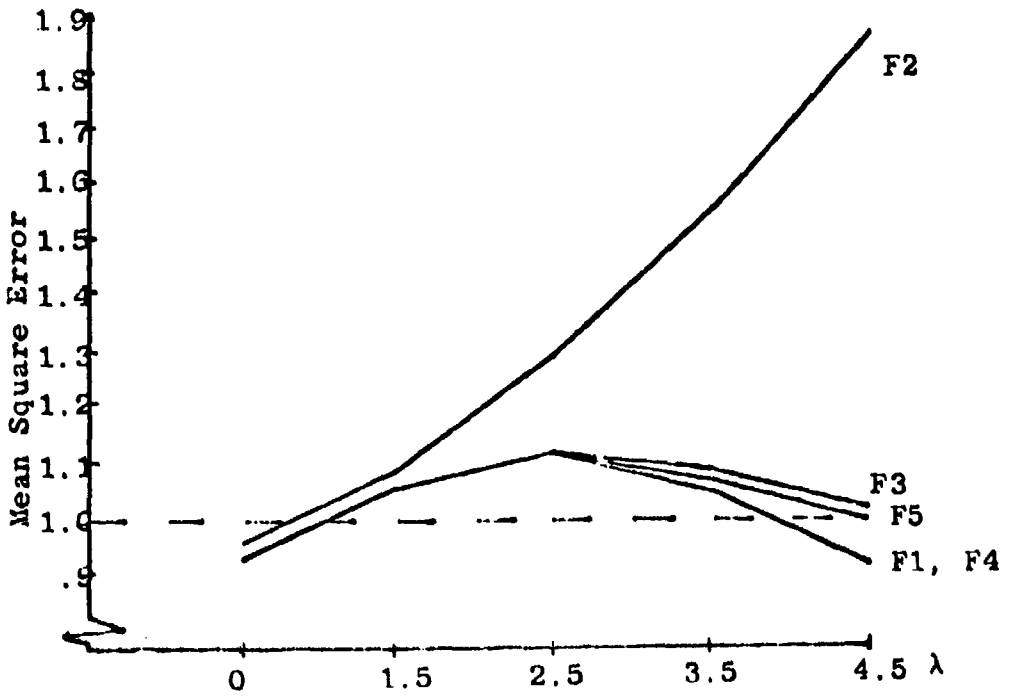


Figure 2. Power vs. λ and MSE vs. λ at α = 0.15 at Case One for F1, F2, F3, F4, and F5

other three and at $\lambda = 4.5$, F5 separates from F1 and F4. At $\lambda = 4.5$ F1 and F4 underestimate error while F3 overestimates error and F5 estimates error exactly.

The problem with F2 is that it will find significance if the smallest sum of squares is sufficiently small without regard to the size of the largest sum of squares. Even if the largest sum of squares is large it will not be declared significant unless the smallest sum of squares is sufficiently small. Thus, F2 has poor power and greatly overestimates mean square error.

At $\lambda = 4.5$, F3 estimates σ^2 to be 1.023. This is significantly different, using Scheffe''s test at $\alpha = 0.05$, compared to the F5 estimate of 1.000. As λ gets large, F3 tends to overestimate σ^2. This is due to the presence of the non-central chi-square in the denominator of F3.

F1 and F4 have the same numerator

$$\sum_{i=n-r+1}^{n} \frac{S_i}{r}$$

which leads to their underestimation of σ^2 at $\lambda = 4.5$. The test for mean square error in case one only goes as far as stage three. Any single degree of freedom sum of squares declared significant at stage one will remain in the numerator for the stage two test. One large single degree of freedom if interaction sum of squares could easily cause a type one error at stage two because of the inflated numerator of the test statistic. This would lead to an underestimation of σ^2.

To further investigate F1 and F4 consider the graph of n by F by $\alpha = 0.15$ interaction on mean square error which is shown in Figure 3.

The points of F1 and F4 for $n = 30$ are lower than one. As the number of individual sums of squares gets larger the probability of a large central chi-square being present increases. The numerators of F1 and F4 will be inflated at stage two with one significant individual sum of squares and a large central chi-square present. Thus a type one error at stage two and possibly at stage three could occur. This would keep large central chi-squares from being pooled into error and would cause an underestimate of σ^2.

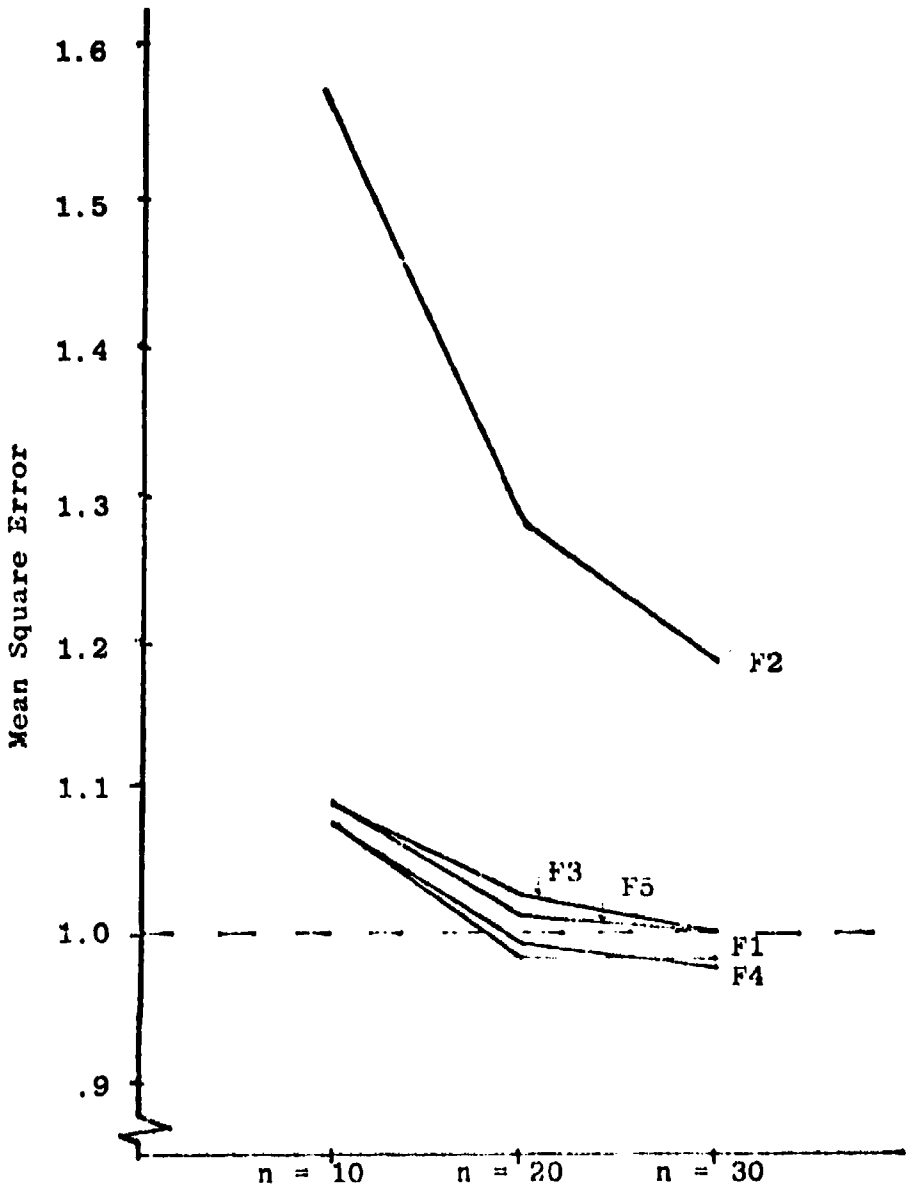


Figure 3. MSE vs. n = 10, n = 20, and n = 30 at Case One for F1, F2, F3, F4, and F5

10. RESULTS AND DISCUSSION OF CASE TWO DATA. Tables 6 and 7 contain the analysis of variance tables for case two power and mean square error data respectively. Significance was found for every term in both tables.

To find the better α for case two consider Figure 4 which is the F by α interaction on power and F by α by stage one power test of interaction on mean square error.

As in case one $\alpha = 0.15$ estimates mean square error better than $\alpha = 0.05$ but Figure 4 shows that the $\alpha = 0.15$ curve isn't as close to one as it was in case one data. This suggests that when two individual sum of squares associated with interaction are present, using a higher α will better estimate σ^2. Figure 4 also shows that F2 has poor power and greatly overestimates mean square error. For these reasons F2 will be dropped from any further discussion.

Figure 4 also shows that F1 and F4 have the best power of the five test statistics. This is further illustrated by Figure 5, a graph of F by λ at $\alpha = 0.15$ interaction on power.

The power of F1, F3, F4, and F5 are very close when pairs of λ are equal, but when the pairs of λ become unequal the pattern changes. As the difference between the non-centralities gets larger the difference in power between F1 and F4 compared to F5 and F3 also spreads. The reason for this becomes obvious after seeing Figures 6 and 7. Figure 6, which is F by λ by $\alpha = 0.15$ by stage one on power, shows no practical difference in power between F1, F3, F4, and F5, but Figure 7, which is F by λ by $\alpha = 0.15$ by stage two on power, shows wide differences in power.

The differences in Figure 5 originate in Figure 7 since Figures 6 and 7 make up Figure 5. Figure 7 is power at stage two or rejecting H_0: $\lambda_1 = \lambda_2 = \ldots = \lambda_{n-1} = 0$ when it is false. The real difference in power between F5 compared with F1 and F4 begins as the pairs of non-centralities start to spread. F1 and F4 have better power because the significant sum of squares at stage one is still in the numerator and when it combines with the smaller non-centrality significance is still found. At the same time F5 and F3 are testing the smaller non-centrality alone and not finding it significant as often as F1 and F4. As the smaller

TABLE 6

ANOVA Table for Case Two Power Data

Source	DF	MS	F*
n	2	2.0076	6662.1818
F	4	4.7802	15862.9815
nF	8	0.1806	599.5040
α	1	9.3725	31102.2565
nα	2	0.0699	231.9816
Fα	4	0.0824	273.6274
nFα	8	0.0035	11.7148
λ	9	2.8400	9424.7329
nλ	18	0.1017	337.8058
Fλ	36	0.1657	550.1733
nFλ	72	0.0078	25.8975
αλ	9	0.0291	96.6522
nαλ	18	0.0093	31.1496
Fαλ	36	0.0026	8.6435
nFαλ	72	0.0007	2.5451
† r	1	0.6061	21922.1602
nr	2	0.2618	868.7970
Fr	4	0.3465	1149.9805
nFσr	8	0.0427	141.7871
σr	1	0.0652	216.6943
nσr	2	0.0072	23.9867
Fσr	4	0.0349	116.1258
nFσr	8	0.0027	8.9653
λr	9	0.6665	2211.8393
nλr	18	0.0260	86.2968
Fλr	36	0.0540	179.3832
nFλr	72	0.0026	8.8321
αλr	9	0.0042	14.1705
nαλr	18	0.0021	7.1075
Fαλr	36	0.0024	8.2095
nFαλr	72	0.0004	1.5963
ERROR	600	0.0003	

† r represents stage of power test
* Each term is significant at the .05 level.

TABLE 7

ANOVA Table for Case Two Mean Square Error Data

Source	DF	MS	F*
n	2	37.5888	62218.3820
F	4	8.5351	14127.7110
nF	8	0.2126	351.9482
α	1	6.5986	10922.3722
nα	2	0.8028	1328.8303
Fα	4	0.0343	56.9299
nFα	8	0.0106	17.6582
λ	9	6.7756	11215.3054
nλ	18	2.3379	3869.8933
Fλ	36	0.3266	540.7006
nFλ	72	0.0078	13.0703
αλ	9	0.3890	643.9728
nαλ	18	0.0964	159.6656
Fαλ	36	0.0086	14.3329
nFαλ	72	0.0036	5.9810
† r	1	126.5534	209475.7356
nr	2	30.6301	50700.0991
Fr	4	0.5413	896.1425
nFr	8	0.0734	121.5168
αr	1	5.2837	8745.8149
Fαr	4	0.0951	157.5172
nFαr	8	0.0332	55.0646
λf	9	1.3656	2260.4472
nλr	18	1.0383	1718.7646
Fλr	36	0.1123	186.0435
nFλr	72	0.0308	51.0854
αλr	9	0.1382	228.7546
nαλr	18	0.0425	70.5036
Fαλr	36	0.0046	7.7365
nFαλr	72	0.0027	4.6175
ERROR	600	0.0006	

† r represents stage

* Each term is significant at the .05 stage.

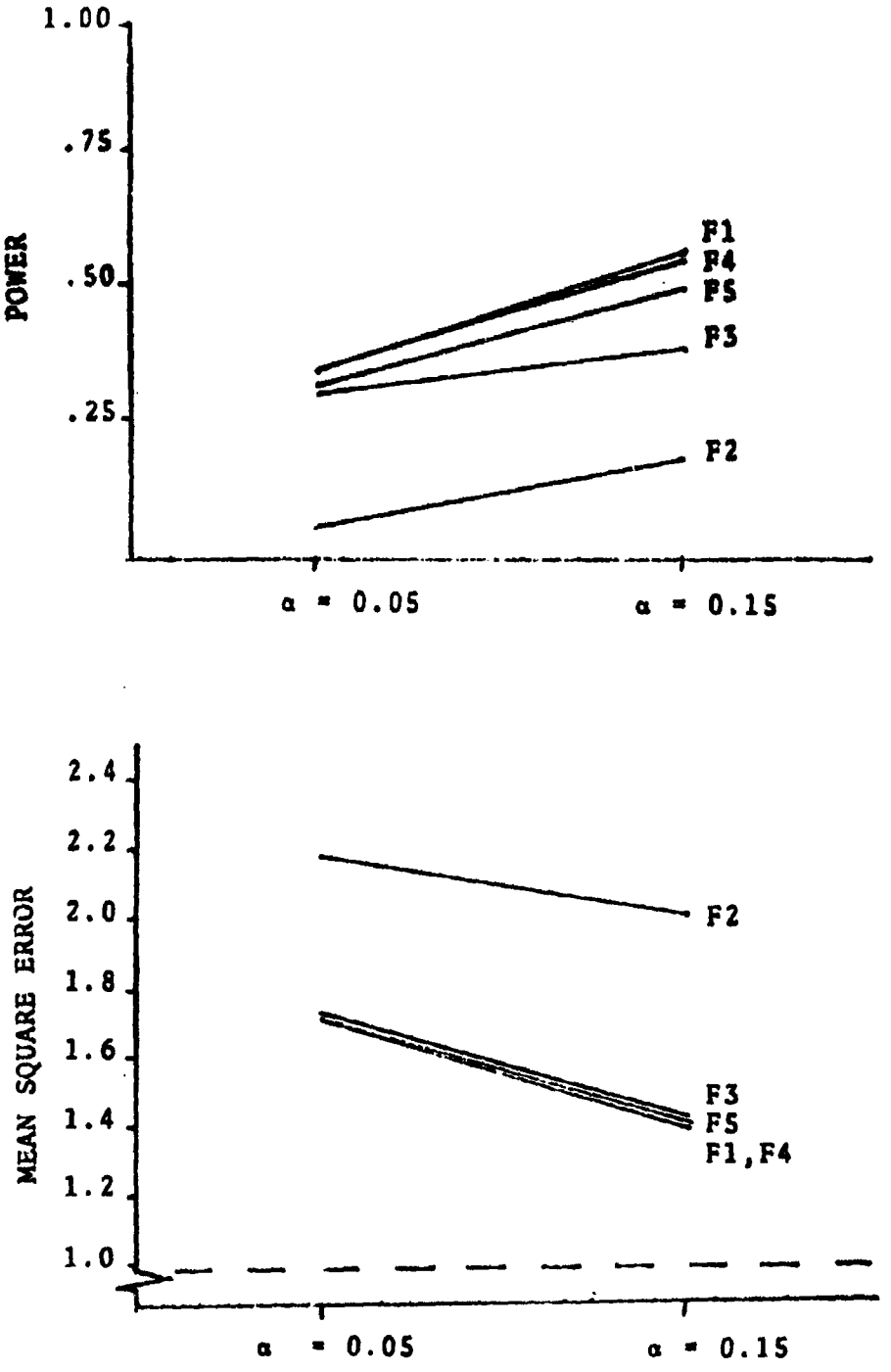


Figure 4. Power vs. $\alpha = 0.05$ and $\alpha = 0.15$ and MSE vs. $\alpha = 0.05$ and $\alpha = 0.15$ at Stage One Power Test and Case Two for F1, F2, F3, F4, and F5.

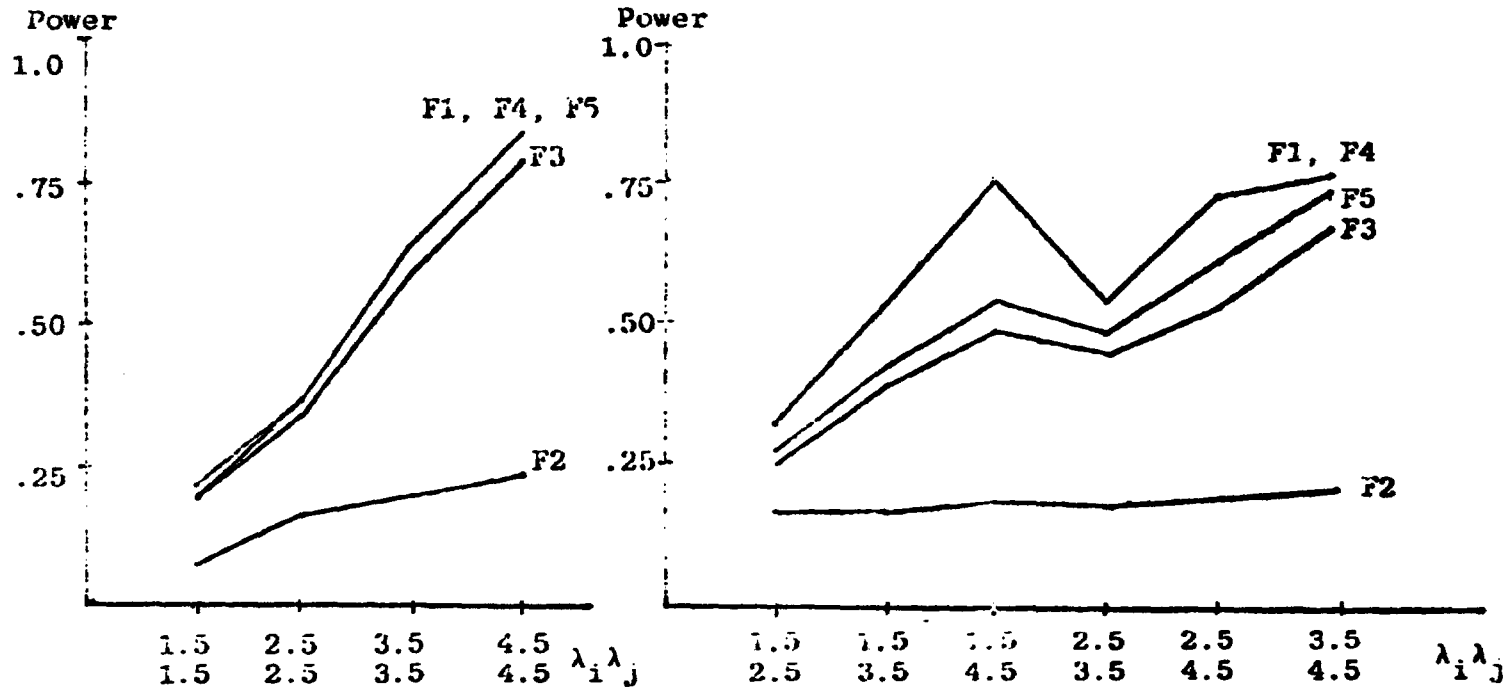


Figure 5. Power vs. λ at α = 0.15 and Case Two for F1, F2, F3, F4, and F5

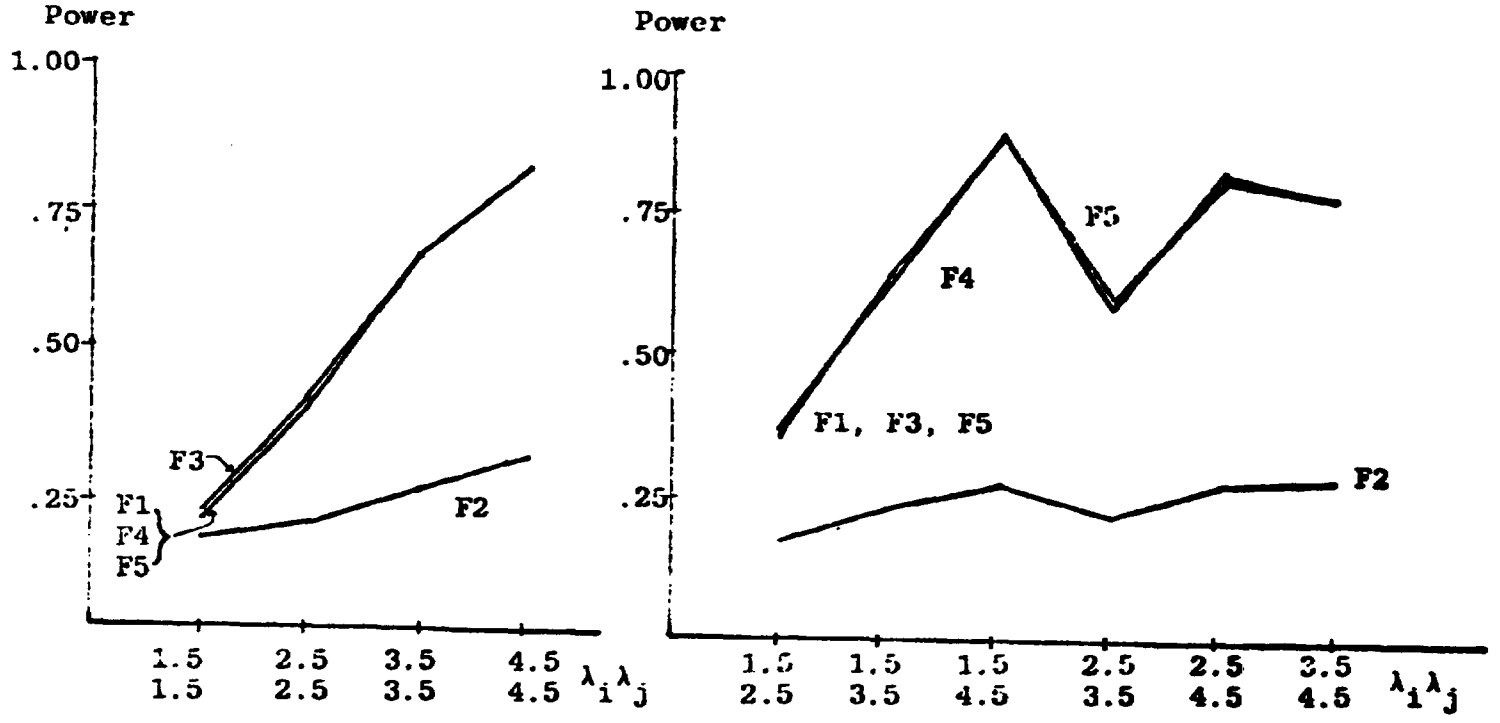


Figure 6. Power vs. λ at $\alpha = 0.15$, Stage One, and Case Two for F1, F2, F3, F4, and F5.

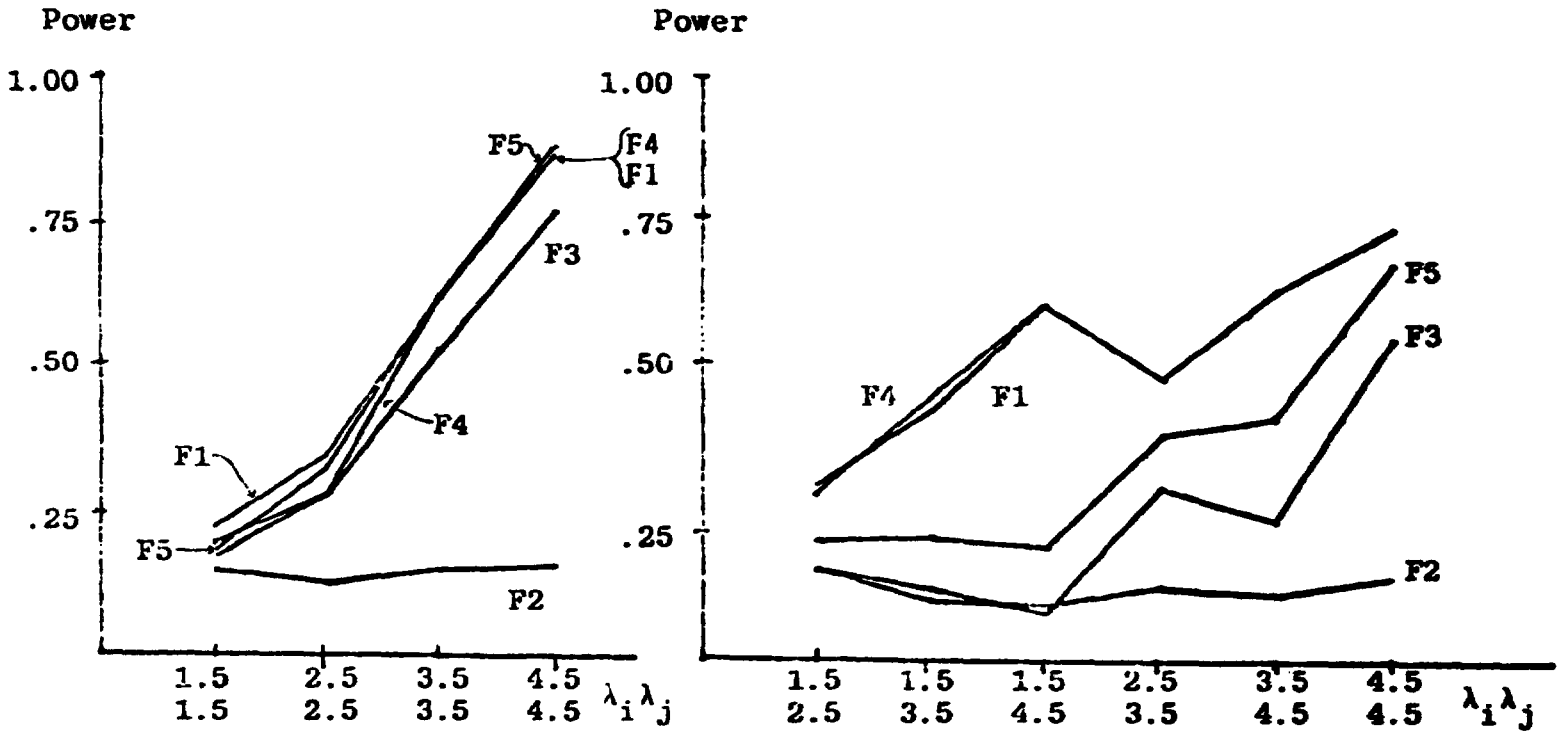


Figure 7. Power vs. λ at $\alpha = 0.15$, Stage Two, and Case Two for F1, F2, F3, F4, and F5

non-centrality gets larger the power of F3 and F5 also increases. This property of F1 and F4 builds their power but may not help their ability to estimate mean square error. Figure 8 is a graph of the F by λ by $\alpha = 0.15$ by stage one power test interaction on mean square error.

The only three places that mean square error of F1 and F4 are significantly closer to one than the mean square error of F5 are where the non-centralities are (1.5, 4.5), (2.5, 4.5), and (4.5, 4.5). This is due to the numerators of F1 and F4 being inflated with 4.5 while F5 is testing 1.5, 2.5, and 4.5 alone. This may be fine for a test using $\alpha = 0.15$, but if $\alpha = 0.25$ were being used, the structure of F1 and F4 could cause them to seriously underestimate σ^2, whereas F5 would not have an inflated numerator nor inflated denominator as F3. This is what happened when testing data with one non-centrality of 4.5 present in case one as illustrated in Figure 2. Figure 4 contains the points in Figure 8 averaged over non-centrality. From Figure 4 at $\alpha = 0.15$ the average mean square error values are 1.401 for F5, 1.391 for F4, and 1.396 for F1. These differences can be attributed to the differences observed in Figure 8 at points where the added non-centralities were (1.5, 4.5), (2.5, 4.5), and (4.5, 4.5). The differences in the ability of F1, F4, and F5 to estimate error variance averaged over everything except α and stage one power test are of no practical importance.

Figure 9 is analogous to Figure 3 in case one. It is the n by F by $\alpha = 0.15$ by stage one power test interaction for mean square error.

At $n = 10$ the value of F5 is significantly closer to one than F1 and F4. But as the sample size increases to $n = 20$ and $n = 30$, F1 and F4 are significantly closer to one than F5. This is because a large central chi-square is more likely to be present as the sample size increases. And the inflated numerators of F1 and F4 tend to declare a portion of the large central chi-squares significant whereas F5 does not. If a large α were being used, F1 and F4 may underestimate error whereas F5 may avoid this problem because of its structure.

11. ANALYSIS OF DATA IN TABLE 1 USING SEQUENTIAL PROCEDURE. Table 8 is an analysis of variance table of the data contained in Table 1.

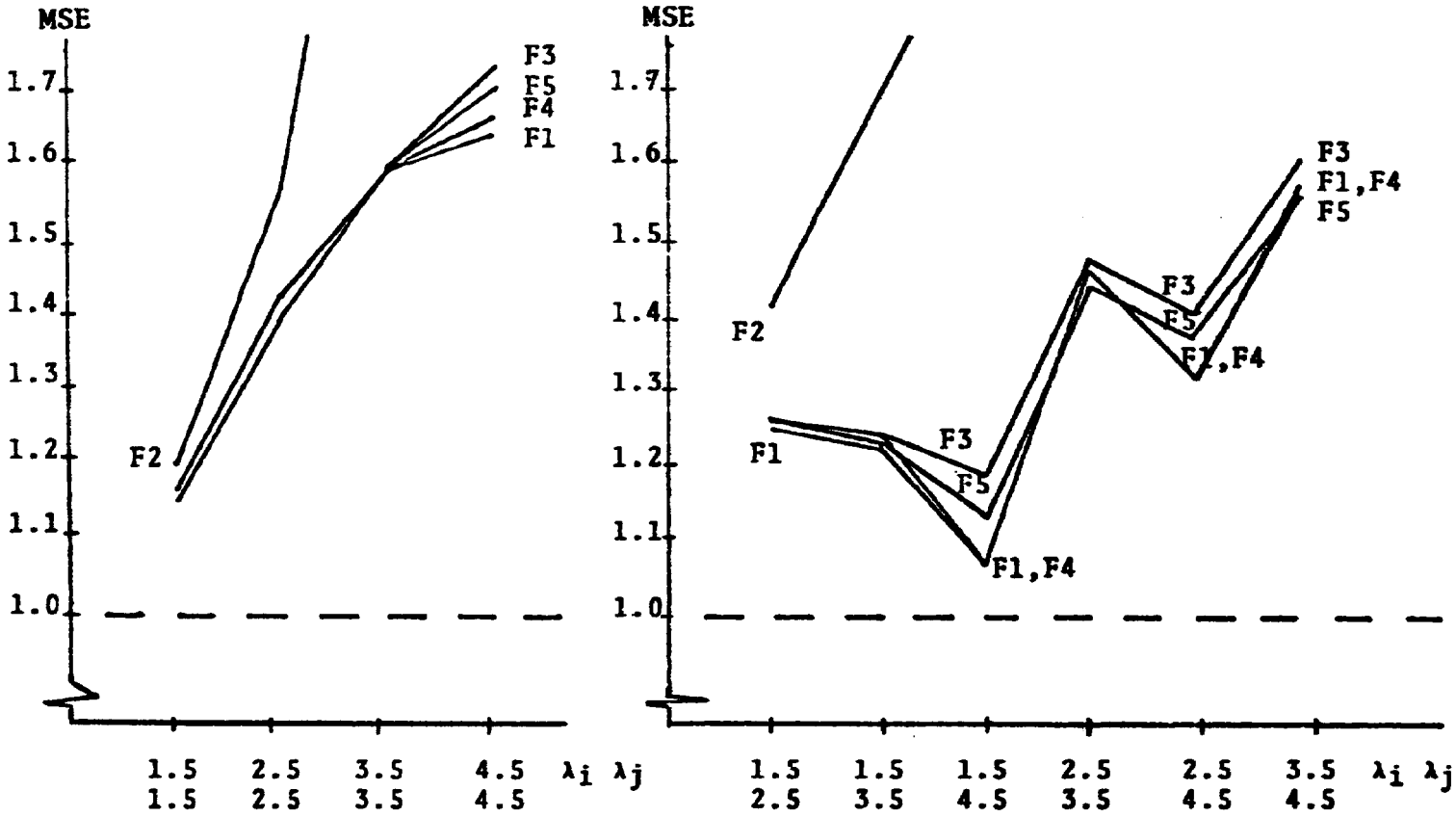


Figure 8. MSE vs. λ at α = 0.15, Stage One Power Test, and Case Two for F1, F2, F3, F4, and F5.

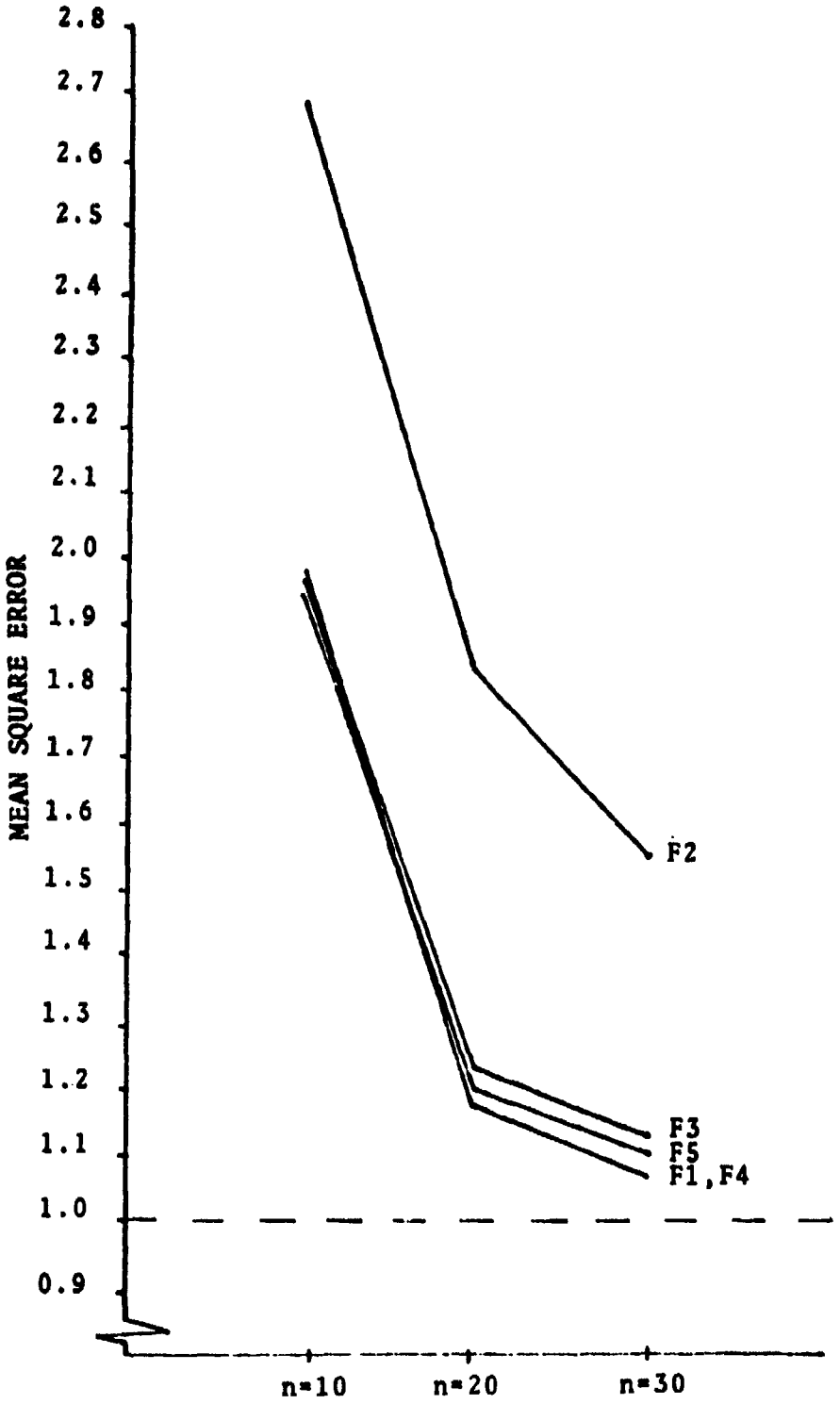


Figure 9. MSE vs. n=10, n=20, and n=30 at α = .15, Stage One Power Test, and Case Two for F1, F2, F3, F4, and F5.

TABLE 8

ANOVA Table for Data in Table 1

Source	DF	SS	MS
A	5	21221.0	4244.2
B	1	3798.5	3798.5
AB	5	6893.9	1378.8
C	4	5310.0	1327.5
AC	20	4433.0	221.7
BC	4	291.8	73.0
ABC	20	2784.2	139.2
ERROR	0	0.0	0.0
TOTAL	59	44732.4	

The AC and ABC interaction terms were partitioned into single degrees of freedom sums of squares and the sequential procedure using F1, F3, F4, and F5 was applied to the data. No indication of interaction was found using $\alpha = 0.15$ in either the AC or ABC term. Thus, both could be pooled into error giving an estimate of σ^2 equal to 180.43, however interaction could be present in most or all of the single degree of freedom sums of squares of AC and ABC, which may lead to a type two error using the sequential procedure.

12. CONCULSIONS. Based on the results of this paper, F1 and F4 may be as good a test statistic as F5 if the remaining sums of squares are pooled into error when significance is declared at stage three. F5 estimates σ^2 better in case one data then F1 and F4 but in case two data there

is no practical difference. If, however, more complete tables were available (higher significance levels and critical values for more than three stages) the authors would recommend F5 as the best of the five test statistics. F5 avoids the pitfalls of F1 and F4 which would probably manifest themselves in much greater detail if critical values for more stages and larger α were available.

As far as level of significance is concerned 0.15 is recommended over 0.05 because of the better estimate of σ^2 given. As the number of individual sums of squares associated with interaction increases a larger value of α will better estimate σ^2. This can be seen by comparing Figure 1 with Figure 4. The results indicate that with a higher α, perhaps 0.25, σ^2 would be estimated with less bias than at $\alpha = 0.15$.

These conclusions can only be strictly applied to the data analyzed in this paper. Any extension to three or more individual sums of squares containing interaction without further research is speculation.

13. SELECTED BIBLIOGRAPHY.

Box, G. E. P. and M. E. Muller. "A Note on the Generation of Random Normal Deviates," Annals of Mathematical Statistics, 29:610-611, 1958.

Bozivich, Helen, T. A. Bancroft, and H. O. Hartley. "Power of Analysis of Variance Test Procedures for Certain Incompletely Specified Models," Annals of Mathematical Statistics, 27:1017-1043, 1956.

Bryce, Gale Rex. "A Unified Method for the Analysis of Unbalanced Designs." Unpublished Master's Thesis, Brigham Yound University, 1970.

Fisher, Ronald A. The Design of Experiments. New York: Hafner Publishing Company, 1951.

Harter, H. Leon and Mary D. Lum. "An Interpretation and Extension of Tukey's One Degree of Freedom for Non-Additivity," Aeronautical Research Laboratory Technical Report, ARL 62-313, 1962.

Johnson, D. E. and F. A. Graybill. "Estimating σ^2 in a Two-Way Classification Model with Interaction," Journal of the American Statistical Association, 67:388-394.

_____. "An Analysis of a Two-Way Model with Interaction and No Replication," Journal of the American Statistical Association, 67:862-868, 1972.

Mandel, John. "Non-Additivity in Two-Way Analysis of Variance," Journal of the American Statistical Association, 56:878-888, 1961.

_____. "A New Analysis of Variance Model for Non-Additive Data," Technometrics, 13:1-18, 1971.

Milliken, G. A. and F. A. Graybill. "Extensions of the General Linear Hypothesis Model," Journal of the American Statistical Association, 65:797-807, 1970.

Scheffe´, Henry. The Analysis of Variance. New York: John Wiley and Sons, 1959.

Snedecor, George W. and William G. Cochran. Statistical Methods. Ames, Iowa: The Iowa State University Press, 1967.

Thomas, Richard W. "A Nonparametric Test for Normality in Analysis of Variance." Unpublished Master's thesis, Brigham Young University, 1975.

Tukey, John W. "One Degree of Freedom for Non-Additivity," Biometrics, 5:232-242, 1949.

_____. "The Future of Data Analysis," Annals of Mathematical Statistics, 33:1-67, 1962.

Ward, G. C. and I. D. Dick. "Non-Additivity in Randomized Block Designs and Balanced Incomplete Block Designs," New Zealand Journal of Science and Technology, 33:430-436, 1952.

PLANNING QUANTAL RESPONSE TESTS FOR ORDNANCE DEVICES: THE TWO-POINT STRATEGY

R. E. Little
School of Engineering
The University of Michigan
Dearborn, Michigan

ABSTRACT. This paper presents a small sample strategy that should prove to be useful in predicting high reliability (or high safety) for ordnance devices. The recommended "two-point" strategy was developed by the author for analogous use in estimating fatigue reliability.

Briefly, the "two-point" strategy incorporates the well-known up-and-down (Bruceton) strategy in its first stage to generate two (nonzero, nonunity probability) points along the assumed response distribution curve. Then, in its second stage, the strategy allocates the remaining specimens to the two corresponding stimulus levels such that the variance of the point estimate pertaining to the reliability (safety) of interest is minimized.

In essence, the issue is to find the specimen allocations which minimize the variance associated with extrapolation along the fitted response distribution to a point remote to the median. Optimally, this minimization requires testing certain specific proportions of the available specimens at carefully selected specific stimulus levels.

1. INTRODUCTION. The sensitivity of explosive devices to shock loading cannot be measured directly. Rather, the explosive device must be subjected to some arbitrary shock loading and if the given device explodes we know that the imposed shock loading exceeded its tolerance to shock loading. On the other hand, if the given device does not explode, then we know that the imposed shock loading did not exceed its tolerance to shock loading. Conducting similar shock loading tests at various (stimulus) levels generates the following quantal response test program:

Stimulus Level (e.g., drop height)	Number of Specimens Tested	Number Responding (e.g., exploding)
s_1	n_1	R_1
s_2	n_2	R_2
.	.	.
.	.	.
s_i	n_i	R_i
.	.	.
.	.	.
s_k	n_k	R_k

The problem of interest herein is how to select s_i and n_i such that we obtain the most precise estimate of the critical stimulus level s_p corresponding to a very low (high) probability of responding p, e.g., 0.001 or even 0.00001 (0.999 or even 0.99999). Specifically we shall describe our two-point test program and estimation method [1,2]. The two-point strategy requires considerably fewer specimens than current techniques such as the run down method [3].

2. OPTIMAL REGRESSION BACKGROUND. The following discussion is intended to serve as background material for the subsequent summary of the two-point strategy.

2.1. Simple Linear Regression Example. Consider the problem of most precise estimation of the slope β for the simple linear model

$$Y = \alpha + \beta x + \varepsilon \qquad (1)$$

Assuming a homoscedastic variance σ^2, the variance of $\hat{\beta}$ is given by the expression

$$\sigma^2_{(\hat{\beta})} = \frac{\sigma^2}{\Sigma n_i (x_i - \bar{x})^2} \qquad (2)$$

Elementary analysis (or intuition) shows the $\sigma^2_{(\hat{\beta})}$ takes on its minimum value when: (a) only two levels of x_i are used in testing, (b) these levels are spaced as widely apart as practical, and (c) $n_{total}/2$ specimens are tested at each of the two x_i levels, where n_{total} is the fixed number of specimens available for testing.

This elementary example illustrates the minimum variance strategy in planning test programs. Namely, select the stimulus levels and allocate the test specimens such that we minimize the variance of some estimate of direct interest. This minimum variance strategy may be applied to models with heteroscedastic variances and with time and/or cost constraints [2].

2.2. Optimal Regression Derivations for Linear Response Curves. We shall now discuss minimum variance estimation of a point on the linear response curve

$$\hat{y} = F^{-1}(\hat{p}) = \hat{\alpha} + \hat{\beta}s \qquad (3)$$

in which s refers to the stimulus level and $p = F(y)$ is the distribution of interest (e.g., normal, logistic, extreme value-smallest). The heteroscedastic binomial variance associated with sampling at a given stimulus level is

$$\sigma^2_{\hat{(p)}} = pq/n \qquad (4)$$

in which p is the true probability of responding, $q = (1 - p)$, and n is the number of specimens tested at the given stimulus level.

We may now use the variance expression for $\hat{p}$ to obtain a variance expression for the variate y, using the simple relation $\sigma^2(aX) = a^2\sigma^2(X)$ and the assumed distribution $p = F(y)$ to obtain dp/dy, viz.,

$$\sigma^2_{\hat{(y)}} = (dy/dp)^2[pq/n] \qquad (5)$$

Now by analogy with the simple linear regression example above, we conduct response tests at just two stimulus levels. Specifically, we test n_1 specimens at stimulus level s_1 and n_2 specimens at stimulus level s_2, where $n_1 + n_2 = n_{total}$ is specified prior to testing. We assume that r_1 specimens respond during the tests at s_1 and r_2 respond at s_2. Hence, the respective proportions responding are $\hat{p}_1 = r_1/n_1$ and $\hat{p}_2 = r_2/n_2$. These $\hat{p}_i$ values are then used to compute the corresponding $\hat{y}_i$ values using the relationship $\hat{y}_i = F^{-1}(\hat{p}_i)$, in which $p = F(y)$ is the distribution function assumed for the response curve.

The response curve of interest appears in Figure 1. Two parameter distributions plot as a straight line on appropriate probability paper, passing through the two points $[(\hat{y}_1, s_1), (\hat{y}_2, s_2)]$. Hence,

$$\hat{\alpha} = (\hat{y}_1 s_2 - \hat{y}_2 s_1)/(s_2 - s_1) \tag{6}$$

and

$$\hat{\beta} = (\hat{y}_2 - \hat{y}_1)/(s_2 - s_1) \tag{7}$$

Then, for any point along the line, say $(\hat{y}_0, s_0)$, we write

$$\hat{y}_0 = \hat{\alpha} + \hat{\beta} s_0 = \frac{\hat{y}_1 s_2 - \hat{y}_2 s_1}{s_2 - s_1} + \frac{\hat{y}_2 - \hat{y}_1}{s_2 - s_1} (s_0) \tag{8}$$

and, since $\hat{y}_1$ and $\hat{y}_2$ are independent, we see that

$$\sigma^2_{\hat{(y_0)}} = \left[\frac{\partial(\hat{y}_0)}{\partial(\hat{y}_1)} \right]^2 \sigma^2_{\hat{(y_1)}} + \left[\frac{\partial(\hat{y}_0)}{\partial(\hat{y}_2)} \right]^2 \sigma^2_{\hat{(y_2)}} \tag{9}$$

in which

$$\frac{\partial(\hat{y}_0)}{\partial(\hat{y}_1)} = (s_2 - s_0)/(s_2 - s_1) \text{ and } \frac{\partial(\hat{y}_0)}{\partial(\hat{y}_2)} = -(s_1 - s_0)/(s_2 - s_1) \tag{10}$$

Next, we substitute $\sigma^2_{\hat{(y_1)}}$ and $\sigma^2_{\hat{(y_2)}}$ into (9) and introduce the notation

$$n_i w_i = 1/\sigma^2_{\hat{(y_i)}} \tag{11}$$

to obtain

$$\sigma^2_{\hat{(y_0)}} = \frac{1}{(s_2 - s_1)^2} \left[\frac{(s_2 - s_0)^2}{n_1 w_1} + \frac{(s_1 - s_0)^2}{n_2 w_2} \right] \tag{12}$$

Our problem now is to minimize (12) by appropriate selection of n_1, n_2, s_1, and s_2.

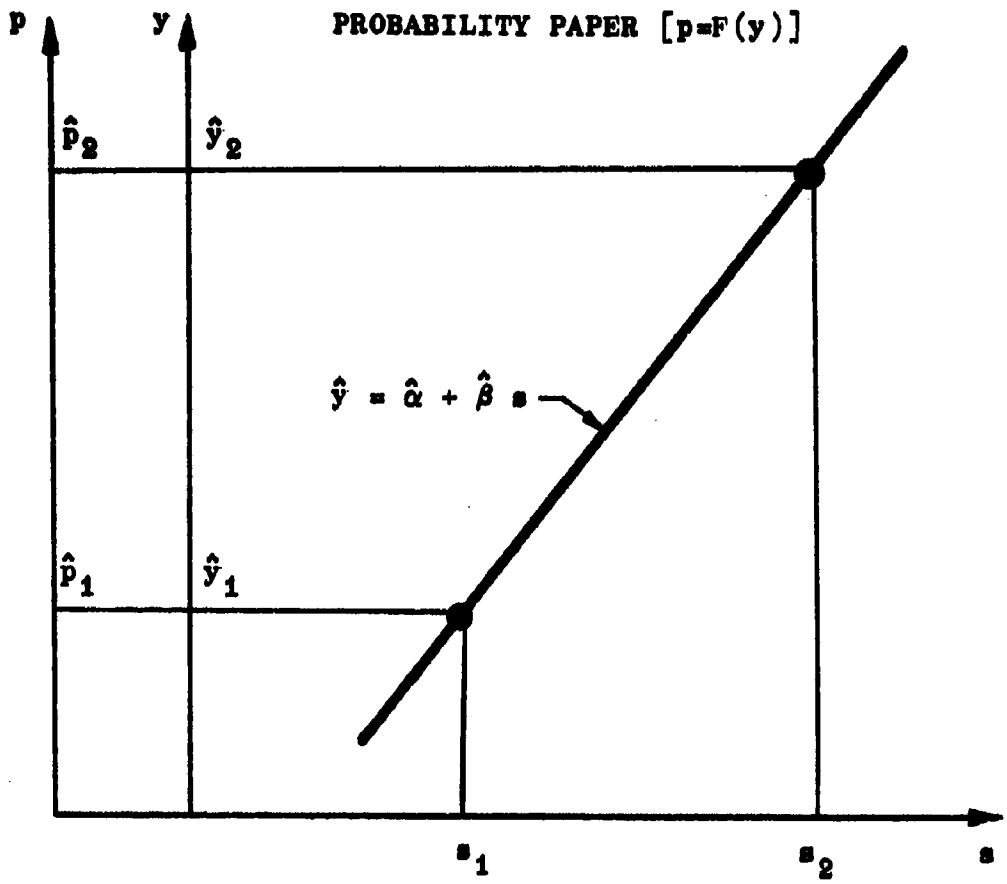


Figure 1. Response curve plotted on probability paper passing through the points $[(\hat{y}_1, s_1), (\hat{y}_2, s_2)]$.

First, consider optimum allocation of n_1 and n_2 for given values of s_1 and s_2. Substitute $n_1 = n_{total} - n_2$ into (12), and set the derivative of (12) with respect to n_2 equal to zero. We thus obtain the expression

$$\frac{\partial[\sigma^2_{\hat{(y_0)}}]}{\partial n_2} = 0 = \frac{1}{(s_2 - s_1)^2}\left[\frac{(s_2 - s_0)^2}{(n_{total} - n_2)^2 w_1} - \frac{(s_1 - s_0)^2}{n_2^2 w_2}\right] \qquad (13)$$

Equation (13) is satisfied when

$$\left(\frac{n_1}{n_2}\right) = \pm \left(\frac{w_2}{w_1}\right)^{\frac{1}{2}} \left(\frac{s_2 - s_0}{s_1 - s_0}\right) = \pm \left(\frac{w_2}{w_1}\right)^{\frac{1}{2}} \left(\frac{y_2 - y_0}{y_1 - y_0}\right) \qquad (14)$$

where the plus sign pertains to extrapolation and the minus sign pertains to interpolation.

Substituting (14) back into (12) gives (after some algebra)

$$\sigma^2_{\hat{(y_0)}} = \frac{1}{n_{total}(s_2 - s_1)^2}\left[\frac{(s_2 - s_0)}{\sqrt{w_1}} \pm \frac{(s_1 - s_0)}{\sqrt{w_2}}\right]^2 \qquad (15)$$

$$= \frac{1}{n_{total}(y_2 - y_1)^2}\left[\frac{(y_2 - y_0)}{\sqrt{w_1}} \pm \frac{(y_1 - y_0)}{\sqrt{w_2}}\right]^2$$

where again the plus sign pertains to extrapolation and the minus sign pertains to interpolation. This variance expression may now be minimized by appropriate selection of y_1 and y_2.

Taking the derivatives of (15) with respect to y_1 and y_2 and equating these derivatives simultaneously to zero shows that the optimum values of y_1 and y_2 are independent of the value of y_0 of specific interest. However, because of the complex nature of the w, p (w, y) relationship, the optimum values must be determined numerically, refer to Table 1.

Distribution	Optimum y		Optimum p	
	y_1	y_2	p_1	p_2
Normal	-1.575	+1.575	0.058	0.942
Logistic	-2.399	+2.399	0.083	0.917
Extreme Value - Smallest	-2.073	+1.269	0.118	0.971^+

Table 1. Optimum y and p values for minimum variance estimation of $\hat{y}_0$.

NOTE: Remarkably the optimum values also pertain to minimum variance estimation of $\hat{\beta}$, but the corresponding optimal allocations differ. The optimum allocations for minimum variance estimation of $\hat{\beta}$ satisfy $n_1/n_2 = (w_2/w_1)^{1/2}$.

Value of y_0	Variance Ratio (Normal Distribution)
- 1.575	1.000
- 2.0	1.16
- 3.0	4.6
- 4.0	63.5

Table 2. Ratio of transformed binomial variance $\sigma^2_{\hat{(y_0)}}$ for all n_{total} tests conducted at stimulus level s_0, to the optimal regression variance $\sigma^2_{\hat{(y_0)}}$. These example results pertain to the normal distribution.

2.2.1. Discussion of Results. It is helpful in understanding the results summarized in Equation (15) and Table 1 to plot w versus p. Refer to Figure 2. Here we see that the weight w approaches zero as p approaches zero or one (viz., as y approaches minus infinity or plus infinity). This w, p (w, y) relationship indicates that if we attempt to separate s_1 and s_2 too widely, the variance of $\hat{y}_0$ increases because w in the denominator of Equation (15) approaches zero. On the other hand, if we do not separate s_1 and s_2 enough, then the term $(s_2 - s_1)^2$ in the denominator is too small. Thus, there are unique values of s_1 and s_2 (independent of s_0) which minimize (15) -- not too far apart and not too close together.

It is also helpful in understanding the optimal (weighted) regression results herein to compare the variances of $\hat{y}_0$ associated with optimal regression and with direct testing at the single stimulus level s_0 corresponding to y_0, refer to Table 2. Here we see that optimal regression is much more efficient than direct testing. The reason for the increased efficiency is essentially that, as evident in Figure 2, direct testing at very low or very high p values is extremely inefficient because the weights w are almost zero (i.e., the transformed binomial variability is so large). The optimal regression strategy, on the other hand, allocates specimens to stimulus levels where the weights are not only much higher than the weights associated with direct testing at extreme values of p, but it also minimizes the increase in the variance of $\hat{y}_0$ associated with extrapolation. It is clear from the results summarized in Table 2 that optimal regression is remarkably suited to the problem of estimating stimulus levels corresponding to very high and to very low probability of response.

2.2.2. Application to Ordnance Problems. The optimum values of p in Table 1 are too close to zero and one to have direct application in ordnance problems. The difficulty lies in selecting s_1 and s_2 such that we do not obtain all response or all non-responses at either s_1 or s_2. If either situation occurs, we cannot establish the two y values required to specify the fitted distribution. Thus, to use the optimal regression results directly, we require very accurate initial estimates of α and β. This requirement is of course quite

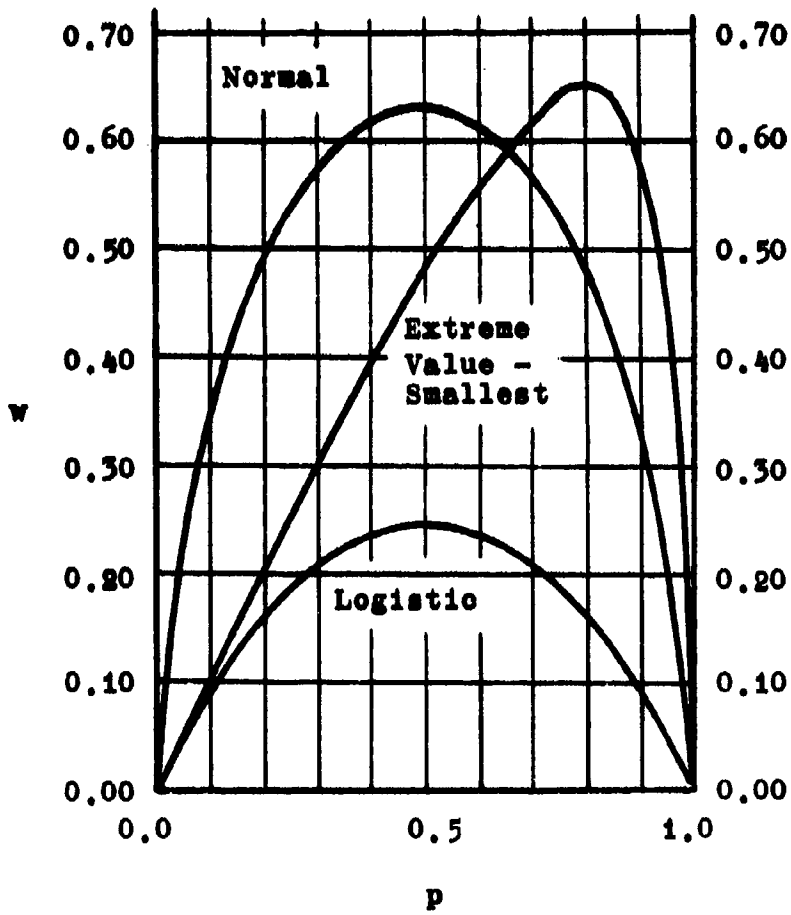


Figure 2. Plot of w, p relationships for the normal, logistic, and extreme value-smallest distributions.

impractical. Thus, we must modify the optimal regression strategy to make sure that we can always establish the two required $\hat{y}$ values. The modified procedure is termed the two-point strategy.

3. THE TWO-POINT STRATEGY. There are two versions of the two-point strategy, one for small samples, say fifty specimens or less, and one for large samples, say one hundred or more specimens.

3.1. Small Sample Procedure. The small sample procedure is as follows: (1) conduct the beginning portion of the test program using the up-and-down strategy illustrated in Figure 3, (2) change over to testing at only two stimulus levels s_1 and s_2 as soon as two finite values of $\hat{y}$ are established by the up-and-down portion of the test program, and (3a) allocate the test specimens to s_1 and s_2 as the test progresses using Equation (14) to decide between testing at s_1 or s_2, or (3b) proceed as in (3a) except test at the two stimulus levels corresponding to the optimum values of $\hat{p}$ in Table 1. (These two levels may be updated as the test progresses. The iterative procedure may be quite worthwhile when s_1 and s_2 are closely spaced[(a)].)

The up-and-down portion of the two-point test program should generally be undertaken with the uniform spacing between successive stimulus levels chosen to be approximately equal to the standard deviation of the underlying response distribution. If the spacing is too narrow, the resulting values of s_1 and s_2 in the two-point testing portion of the program will generally be too close together to permit precise estimation of y_0. On the other hand, if the spacing is too wide, the up-and-down portion of the test program tends to be quite long, with the successive test outcomes alternating back and forth between response and nonresponse. Thus, a reasonably accurate estimate of the standard deviation of the assumed underlying distribution is mandatory, viz., there

(a) Ideally the investigator has a computer program which records the given test outcome and provides the stimulus level for the next test. Otherwise, the computations may take place at convenient intervals as the test program progresses.

Code:	X denotes Response,	0 denotes Nonresponse

Stimulus Level	Test Number 1 2 3 4 5 6 7 8 9 10	
2.0	X	
1.7	X X X	
1.4	0 X 0 X	XXXX0
1.1	0 X	0000000000000000X0
0.8		
	Up-and-down Testing (a)	Two-point Testing (b)

Data Summary:

s_i	n_i	r_i	$\hat{p}_i$
2.0	1	1	1.000
1.7	3	3	1.000
1.4	9	6	0.667
1.1	20	2	0.100

Figure 3. The two-point test program consists of: (a) a beginning up-and-down series of tests to establish two finite $\hat{y}$ values ($\hat{p}$ values not equal to zero or one), followed by (b) tests conducted at two stimulus levels, s_1 and s_2, which specimens allocated to s_1 or s_2 as the overall test progresses such that text Equation (14) is satisfied.

NOTE: The up-and-down test strategy is as follows: The outcome of any given test determines the stimulus level used in the next test. For example, the second specimen responded (denoted X), thus the third specimen was tested at a lower stimulus level. On the other hand, the third specimen did not respond (denoted 0) and therefore the fourth specimen was tested at the next higher stimulus level. Uniform spacing between adjacent stimulus levels is used for convenience, but is not mandatory.

must be some preliminary testing or some prior experience to form a basis for selecting the spacing of the stimulus levels used in testing. Generally an estimate of the standard deviation σ that is accurate within plus or minus fifty percent is adequate, but it is preferable that the spacing d fall in the range $\sigma < d < (3\sigma/2)$. The advantage of the iterative procedure (3b) increases as d is decreased below σ.

Many readers will probably opt for the simplified test method and analysis. In this case we merely ignore the tests conducted at stimulus levels other than s_1 and s_2 (refer to Figure 3) and estimate the fitted distribution by drawing a straight line through the two points $[(\hat{y}_1, s_1), (\hat{y}_2, s_2)]$. The variance of $\hat{y}_0$ is then estimated using Equation (12) and reading $\hat{w}$ from Figure 2.

If it does not seem advisable to ignore tests at stimulus levels other than s_1 and s_2, the variance of $\hat{y}_0$ may be estimated using the general expression

$$\sigma^2_{(\hat{y}_0)} = \left[\frac{1}{\Sigma n_i w_i} + \frac{(s_0 - \bar{s}_w)^2}{\Sigma n_i w_i (s_i - \bar{s}_w)^2} \right] \qquad (16)$$

The w_i values in (16) may be approximated either by empirical weights (i.e., based on the observed $\hat{y}_i$ values), or fitted weights (e.g., based on maximum likelihood analyses [2]).

3.1.1. Numerical Example (Simplified Analysis). Given the quantal response data in Figure 3 (ignoring the tests at stimulus levels other than 1.4 and 1.1), viz.,

s_i	n_i	r_i	$\hat{p}_i$
1.4	9	6	0.667
1.1	20	2	0.100

estimate s_0 corresponding to p = 0.001 and sketch the lower 95% (asymptotic) confidence band. Assume an underlying normal distribution.

Solution. First, we shall check the allocation of n_1 and n_2, relative to the final values of $\hat{p}$. For $\hat{p}_1 = 0.100$, $\hat{y}_1$ from normal tables equals - 1.28; and for $\hat{p}_2 = 0.667$, $\hat{y}_2$ equals + 0.43. Moreover, for $p_0 = 0.001$, $y_0 = -3.09$. The corresponding values of $\hat{w}$ are 0.34 and 0.60 respectively. Thus, using (14)

$$\frac{n_1}{n_2} \approx \left(\frac{0.60}{0.34} \right)^{\frac{1}{2}} \left[\frac{+ 0.43 - (- 3.09)}{- 1.28 - (- 3.09)} \right] = 2.6$$

whereas the actual value is 20/9 = 2.2. This discrepancy means that if further tests were conducted, the first few additional tests should be conducted at $s_1 = 1.1$ -- unless of course the $\hat{p}$ values change markedly as the data accumulate.

The fitted response distribution passes through the points [(1.1, - 1.28), (1.4, + 0.43)], giving the response expression

$$\hat{y} = - 7.55 + 5.07s$$

Hence, $y_0 = - 3.09$ ($p_0 = 0.001$) corresponds to s_0 equal 0.78. In turn, using (12)

$$\sigma^2_{\hat{(y_0)}} \approx \frac{1}{(1.4 - 1.1)^2} \left[\frac{(1.4 - 0.78)^2}{20 \times 0.34} + \frac{(1.1 - 0.78)^2}{9 \times 0.60} \right]$$

Thus

$$\sigma^2_{\hat{(y_0)}} \approx 0.84$$

The corresponding lower 95% asymptotic confidence band appears in Figure 4. Note that we can be approximately 95% confident that 99.9% of all specimens will survive a stimulus level of 0.22.

3.2. Large Sample Procedure. The large sample procedure is based on information obtained by response tests conducted using the previous small sample procedure. Namely, approximately fifty specimens are tested using the small sample procedure to estimate s_1^* and s_2^* corresponding to the optimum p values in Table 1. Then, given this information,

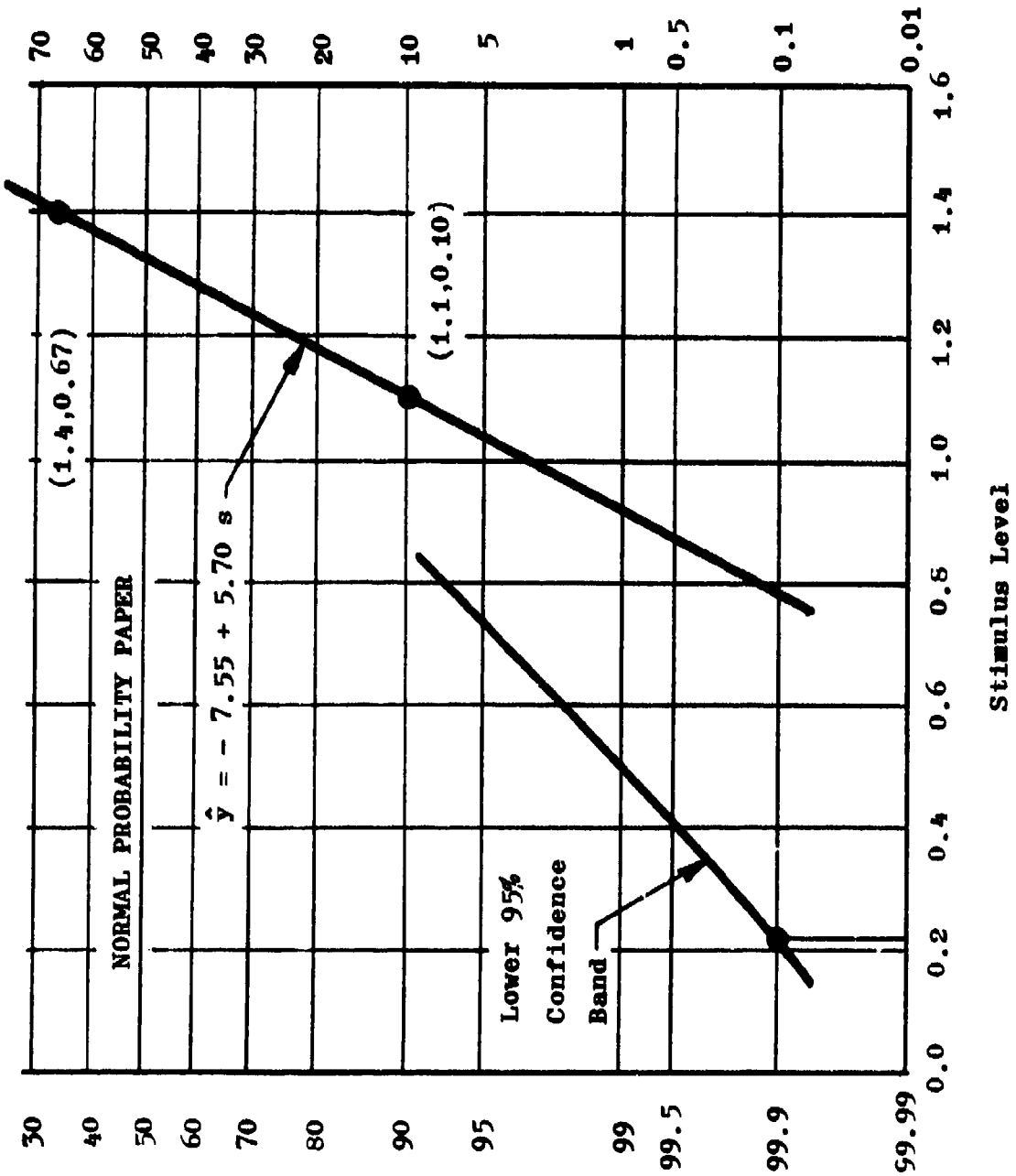


Figure 4. Plot of fitted response curve and the associated lower 95% confidence band for the test example.

the remaining specimens are tests at s^*_1 or at s^*_2 using (14) for appropriate allocation; or else each successive specimen may be tested at that stimulus level which minimizes (16) as the data accumulate. The latter iterative procedure is enhanced by a digital computer program compiled and placed in a file ready for execution by remote terminal.

4. SUMMARY. The procedure is straightforward: (a) select the appropriate values of the stimulus level, and (b) allocate the tests at these stimulus levels such that the variance of the desired point estimate is minimized. Usually the variance of the desired point estimate may be reduced markedly merely by considering a few alternative stimulus levels before testing (using Figure 2 and Equation 16). But the variance of the point estimate may be reduced even further by adopting certain minimum variance strategies.

REFERENCES

1. R. E. Little, Handbook for Up-and-Down Testing with Small Samples, unpublished ms.

2. R. E. Little and E. H. Jebe, Statistical Design of Fatigue Experiments, Applied Science Publishers, Ltd., London, 1975. (See also, R. E. Little and E. H. Jebe, Manual on the Statistical Planning and Analyses for Fatigue Experiments, ASTM STP 588, 1975.)

3. J. N. Ayres, L. D. Hampton and I. Kabik, "Explosive Safety and Reliability Estimates from a Limited Sample Size," Proceedings of the Tenth Conference on the Design of Experiments in Army Research, Development and Testing, ARO-D Report 65-3, October 1965, pp. 261-281. (See also, Department of Defense, MIL-STD-1512 (ASAF), March 1972, "Military Standard; Electroexplosive Subsystems, Electrically Initiated, Design Requirements and Test Methods.")

TECHNIQUES FOR STATISTICALLY DETERMINING FLIGHT SUITABILITY OF AN ARTILLERY PROJECTILE

Ronald Corn
Gertrude Weintraub
Ammunition Development and
Engineering Directorate
Picatinny Arsenal
Dover, New Jersey

ABSTRACT. The M483 155mm Projectile being tested at Nicolet, Canada, to evaluate aeroballistic performance at high air density exhibited flight instability. The authors were responsible for determining cause of problem, correcting the problem and developing the statistical technique necessary for predicting success. The projectile design modifications evolved successfully passed retesting at Nicolet and the projectile has been released for production. The induced yaw technique for disturbing projectiles as they exit the gun tube, developed during this program, is currently being used on other developmental projectiles and will be used to evaluate aerodynamic stability of all future Howitzer type projectiles.

The statistical techniques used to predict success which also permitted a minimal expenditure of projectiles were:

a. A Weibull mathematical model was selected and implemented to predict point estimates and confidence level estimates of reliability and percentage points based upon the maximum likelihood estimates of the parameters of a Weibull population. This model afforded excellent theoretical descriptive characteristics of the density and probability distributions of the empirical test data which were symmetrical and asymmetrical in form.

b. Automated computer programs especially adapted to the Weibull model were employed to derive density and probability distribution curves.

c. Probability plotting methods were implemented to describe the adequacy of the theoretical distributions to the empirical test data.

1. INTRODUCTION. The M483 projectile development which was completed in 1971 provided an important new 155mm capability to the US Army. Figure 1 depicts an M483 projectile alongside of the standard 155mm M107 projectile. Because of

the obvious increase in size and cargo volume, over 50% of the standard, the M483 configuration is being utilized for a variety of projectiles whose mission is to deliver cargo on to a target area (e.g. chemical, smoke, illuminating and submunition).

To accommodate the increased cargo, the M483 projectile is over 6 calibers in length and utilizes an aluminum ogive and base and fiberglass wrapped body to minimize weight and distribute it properly for aerodynamic considerations. Because of its unique shape, comparatively little knowledge of its aerodynamic characteristics was available prior to 1974 when surprisingly poor performance was exhibited in cold weather tests.

In 1974 a cold weather test program was conducted at Nicolet, Canada, located between Montreal and Quebec along the Saint Lawrence River (Figure 2). Nicolet provides an existing Canadian test facility which permits projectile firings at near Arctic conditions to evaluate aeroballistic performance at high air density (in excess of 110% of standard), which tends to amplify aerodynamic instabilities.

On 14 Feb 74, 20 each M483 projectiles were fired with a standard US Propellant charge whose weight was adjusted to obtain a velocity of Mach 0.93. At these Arctic conditions this Mach number was predicted to be the most severe aerodynamically. The impact point of 13 of those projectiles which exhibited normal flight performance is shown in Figure 3. These projectiles impacted on expected ranges of approximately 6300 meters. Seven of the twenty projectiles impacted between 2000 and 3300 meters short of the impact area as shown in Figure 4.

Production of the M483 was suspended as a result of the incident at Nicolet and an intensive program initiated to determine the cause of the erratic performance at Nicolet. Initially a fault tree was configured (Figure 5) and an investigative program was developed based upon fault tree elements.

To determine whether the cause of the problem was routed in interior or exterior ballistics, it was necessary to conduct a highly instrumented series of firings which for the first time, would obtain initial yaw characteristics of a statistical sample of in-flight projectiles, as well as projectile range information for those same projectiles. Figure 6 shows the test site at Yuma Proving Ground. Cameras and

yaw cards were used to independently measure launch angles of the projectile while radar and standard triangulation techniques were used to determine flight characteristics and down range impact points. Launch velocities were adjusted from standard US velocities to duplicate the critical mach number of the Nicolet tests by modification of propelling charges.

The results of the initial tests showed that the M483 problem was primarily an exterior ballistic problem and that in fact, the aeroballistic characteristics of the projectile were unsatisfactory. Ms. Weintraub's application of statistical techniques proved invaluable for predicting performance and follows in detail.

2. STATISTICAL TECHNIQUES USED IN A FLIGHT SUITABILITY INVESTIGATION. At the outset, I want to take this opportunity to express my gratitude to Mr. Corn and his associates in this stability investigation. They were open-minded and willing to draw upon statistical disciplines to assist them in resolving an engineering problem. The result of the cohesive union of engineering and statistics proved successful.

A complex problem was solved when a probabilistic approach was applied to analyze real world test data. Professor John Tukey of Princeton would probably refer to the statistician's efforts in our data analysis as exploratory and probabilistic and the end result as confirmatory. Our greatest gains in analyzing empirical data came from surprises, which I will explain a little later.

In this case, the engineering community succeeded in ferreting out the causes for short rounds (defined as those which fail to fly to full range) and redesigned the projectile to eliminate the occurance of short rounds.

As statistician, I entered the picture after the following events had occurred:

1. On 10 Feb 74, seven out of twenty standard M483 projectiles fired at critical Mach number (0.93) from the 109A1 Howitzer flew approximately half range.

2. The engineering community undertook an investigation by designing a test program to determine the cause of these short rounds. The program wasacomplex and ambitious one and sought to determine whether the problem was either interior or exterior ballistic related.

3. Aerodynamic knowledge at the start of the investigation supported the belief that the M483 was stable up to a first maximum yaw angle of 8°.

The first test conducted at Yuma Proving Grounds was with the standard M483 fired at critical mach number in order to correlate first maximum yaw angle with range. The yaw angles were obtained with yaw cards and cameras as back up, the test set up is shown in Figure 6. The yaw cards were set approximately 100 feet forward of the gun.

Figure 7 is a plot of the first maximum yaw angle vs. range and the first surprise of this test program was that the critical yaw angle was 5-6° and not 8° as previously predicted. Critical yaw angle is defined as the angle above which the projectile becomes aerodynamically unstable and does not fly full range.

The yaw angles generated from 20 tests conducted with the standard M483 projectile (varying its internal cargo, tubes and muzzle brakes) were presented for analysis. As had been done on other problems, a probabilistic design approach was used. Yaw angle was considered the continuous random variable and the problem was to examine the distribution of yaw angle. I chose to fit a Weibull distribution model since it afforded me a useful mathematical tool for describing the probability distribution function and the density function of symmetrical and asymetrical forms. Figure 8 shows a spectrum of distributional forms which can be described by a Weibull model (see Figure 9 for the pdf and density mathematical forms of the Weibull distribution).

In terms of a statistical probability distribution, the distribution of yaw angles for the standard M483 Projectile fired from a 50% worn tube at Yuma is seen on Figure 10. It was determined that this condition tube produced the highest first maximum yaw distribution and this tube was used for most of the testing.

Maximum likelihood estimates of the parameters of a Weibull population were determined based upon the iteration procedures for joint maximum likelihood estimation of the 3 parameters of the Weibull population described by Harter and Moore in their notes contained in Technometrics, Volume 7, No. 4, November 1965. The asymptotic variances and covariances of maximum-likelihood estimators were then employed in deriving confidence interval estimates for probabilities based upon the MLE estimates. The latter confidence interval estimates were derived with the assistance of Dr. Einbinder and members of the Computer Programming Facility at Picatinny Arsenal.

Based upon the maximum likelihood estimates of the 3 Weibull parameters, one could expect 33% of the standard M483 Projectiles fired from the 50% worn tube to exceed 5°. And, in fact, at Nicolet, Canada, 7 out of 20 (35%) fell short. This gave further credence to the low critical maximum yaw angle premise.

The fitted yaw distribution function also indicated that for the standard M483 to fly full range, its critical first maximum yaw angle must be greater than 13°. At this critical yaw angle one can expect no more than one short range projectile in a million rounds.

Thereafter, the investigative test program was directed to assessing the effects of system parameter changes on the yaw angle distribution and the design of modifications that would have high critical yaw angles. The system parameters investigated included: new tubes and worn tubes, with and without muzzle brakes, and cargo variation. It appeared that the greatest effect on yaw angle level was the presence or absense of a muzzle brake on the end of the gun tube.

Figure 11 shows how absence of a muzzle brake improves the yaw angle probability distribution of the standard M483. Now only 7 in 10,000 rounds are expected to exceed the 5° critical yaw angle in lieu of 33% with a muzzle brake. This frequency was also too high to be acceptable.

The real problem facing the engineering task team was to design a projectile modification whose critical angle exceeded 13°, since as previously shown no more than one short range round in a million would be expected at this critical yaw angle.

After many design modifications, and statistical analyses of these changes, two modifications of the standard M483 were built and tested: Figure 12 describes the modifications made to the standard M483; Figure 13 compares the yaw angle probability distribution functions obtained for Mods 1 and 2 when tested with the 50% worn tube with muzzle brake. For each Mod, it was found that one in a million rounds would exceed 8° first maximum yaw angle.

Since the modifications were designed to be more stable than the standard M483, a technique had to be devised for determining how much more stable they were and also their critical yaw angle.

Since it had been determined that muzzle brakes significantly effected yaw angles, modified muzzle brakes, Figure 13A,

were designed and tested as a means of inducing even greater yaw angles to evaluate design modifications. First maximum yaw angles of as high as 20° were obtained.

Figure 14 illustrates, visually, by means of a yaw card comparison, the large angle from which the modified rounds will still damp and fly normal ranges as compared to the original M483 projectile, Figure 15.

An interim Picatinny Report dated March 1975 has been published covering this work. Figures 16 and 17 show the adequacy of the Weibull model in describing the empirical distribution characteristics of test data for the standard M483 round and for design modification 2.

This probability plotting method was used to assess the goodness of fit of the theoretical Weibull model to the empirical test data.

Figures 18 and 19 show the density function for the standard M483 and design modification 2. Each of the distributions is right-skewed, but we can see that modification 2 shows a significantly smaller dispersion around the mean.

Summing up, therefore, what modification 2 accomplished is two-fold:

1. It yielded a significantly smaller dispersion of first maximum yaw angle around the mean, one in a million exceeds 8° vs. 33% exceeding 5° for the standard M483.

2. It produced a more stable projectile, critical angle greater than 18° vs. 5-6° for the standard M483.

3. CONCLUSION. A real world engineering problem was resolved with the assistance of probability methods. Statistical analyses were helped immeasurably by computer software programs which were available. These programs afforded rapid assessment of design modifications and comparisons. The efforts could not possibly have been accomplished in as short a time without the computer. The computer program of Drs. Harter and Moore of Wright-Patterson Air Force Base was used extensively to derive the maximum likelihood estimates of the Weibull parameters.[1] Software programs available at Picatinny

---

[1] As an aside, gratitude is extended to Dr. Badrig Kurkjian for introducing Picatinny Arsenal to the Harter Moore program which has proved to be invaluable in helping to solve many engineering problems.

Arsenal, specifically in the Concepts and Effectiveness Division, contributed greatly toward the successful evaluation of test data.

4. STATISTICAL CONTRIBUTION.

1. Statistical probability techniques fixed the critical yaw angle for the standard M483 Projectile.

2. Statistical analysis predicted the yaw angle probability distribution for many modifications and for different tubes. These distributions provided the engineering task team with essential information for directing their efforts toward projectile modification.

3. a. For the first time, probability design was used to predict projectile performance using a minimal number of rounds. Cost reduction and risk associated with future artillery development programs should follow.

b. The application of probability design served a twofold purpose:

(1) It predicted the probability of exceeding a given yaw for a specific design M483 Projectile.

(2) It afforded the engineering task team a goal, in this case, a 13° critical yaw angle; so that their efforts were directed toward achieving this goal in order to eliminate short rounds.

4. A Blue Ribbon Panel especially assigned to overview the stability investigation approved the efforts and findings of the investigative team and commended all members of the team for their analysis of and correction to the projectile flight problem. The panel further stated that "in the course of this program much has been learned that is of basic value in the ballistic design and development of projectiles." Further, the panel recommended that the "team can well undertake future new and interesting designs of special shells" and recommended that this project be well documented for future guidance.

CONCLUDING REMARKS:

As a result of the program and techniques just described, modifications 1 and 2 were extensively tested at Nicolet during the winter of 1975. Both modifications performed satisfactorily as predicted. Modification 2 was selected since it did not result in internal cargo volume loss and it was recently released for production as the M483A1.

REFERENCES:

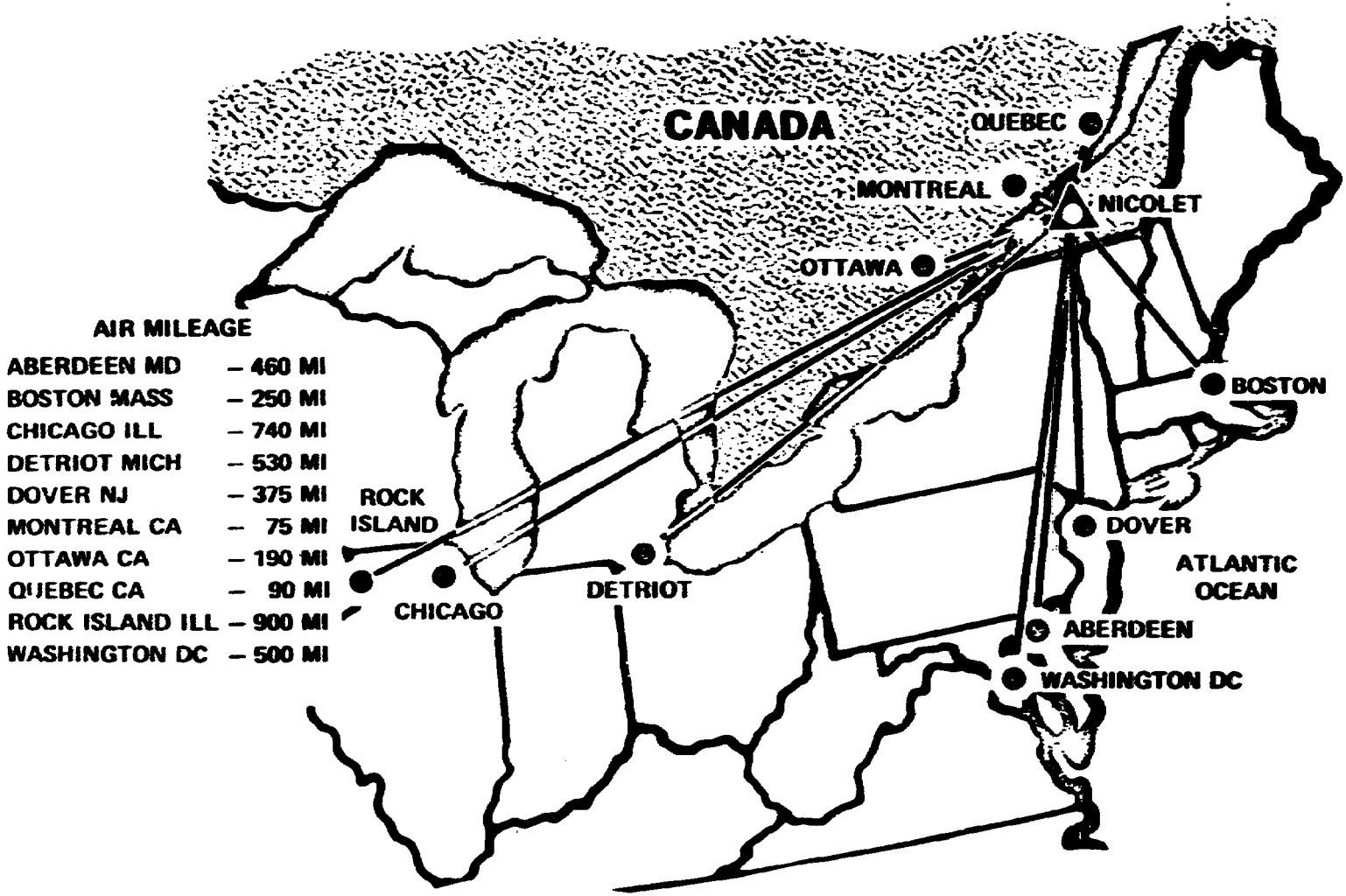
1. H. Leon Harter and Albert H. Moore, "Maximum - Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples", Technometrics, Vol. 7, No. 4, Nov 1965, p. 639-643.

2. H. Leon Harter and Albert H. Moore, "Asymptotic Variances of Maximum Likelihood Estimators, from Censored Samples of the Parameters of Weibull and Gamma Populations", Annals of Mathematical Statistics, Vol. 38, No. 2, April 1967, p. 557-570.

3. Seymour K. Einbinder, "One Shot Sensitivity Test for Extreme Percentage Points", ARO Report 74-1, Dept. of Defense, p. 369-386.

4. M. B. Wilk and R. Granadesikan, "Probability Plotting Methods for the Analysis of Data, "Biometrika, 1968, Vol. 55, p. 1-17.

FIGURE 1.



AIR MILEAGE

ABERDEEN MD	– 460 MI
BOSTON MASS	– 250 MI
CHICAGO ILL	– 740 MI
DETRIOT MICH	– 530 MI
DOVER NJ	– 375 MI
MONTREAL CA	– 75 MI
OTTAWA CA	– 190 MI
QUEBEC CA	– 90 MI
ROCK ISLAND ILL	– 900 MI
WASHINGTON DC	– 500 MI

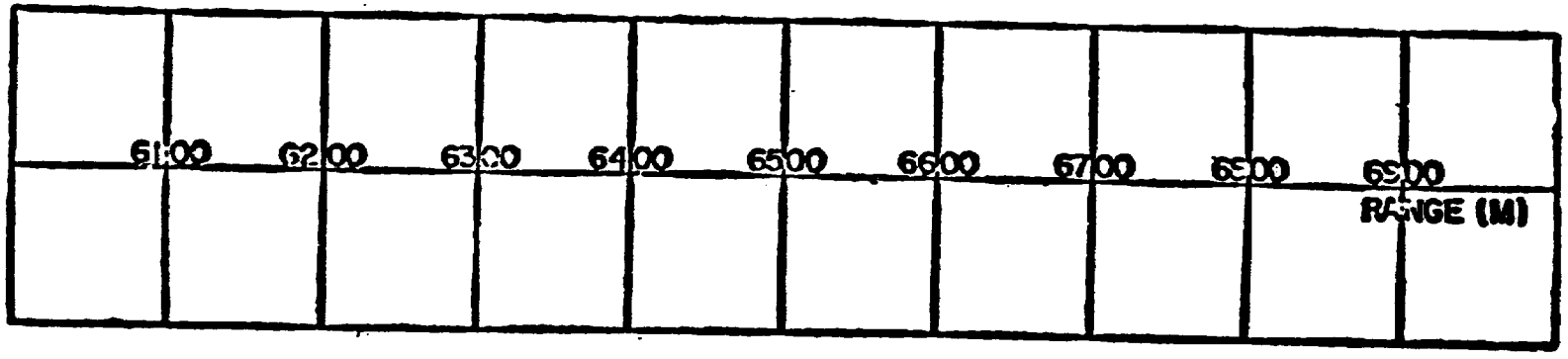
PRELIMINARY

14 FEB 74

M483-UK ZONE 3 (NORMAL ROUNDS)

QE=30°

T=-2° F ρ=115% STD



MEAN OBSERVED RANGE (W/O 7 SHORT ROUNDS)= 6300M
PROBABLE ERROR=22M=.4%R

Fall of Shot, M483 (13 Normal Rounds). 14 Feb 74

FIGURE 3

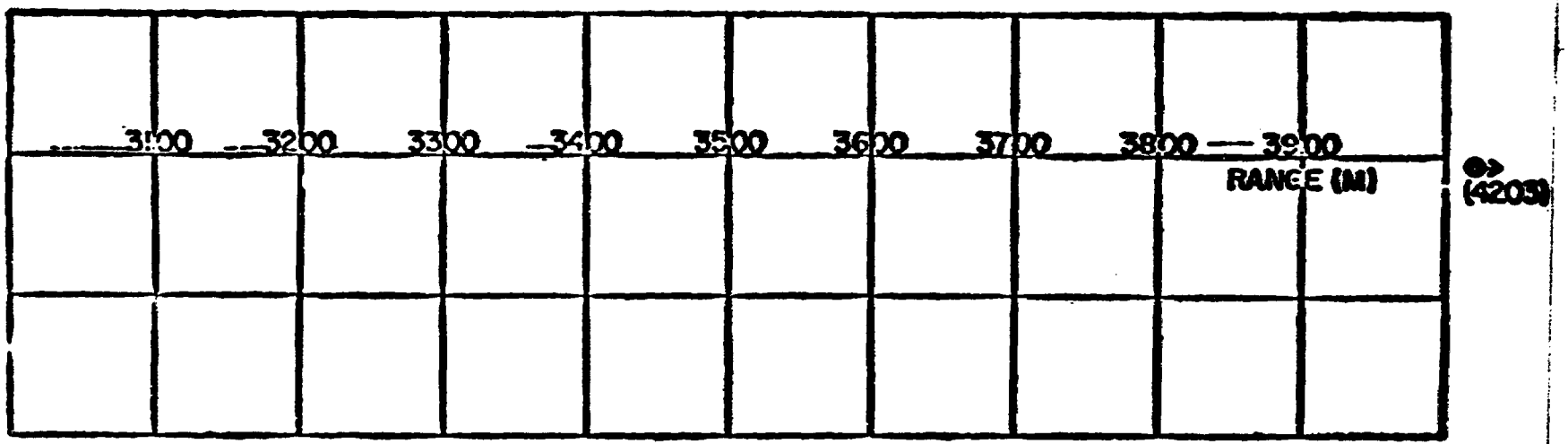
PRELIMINARY

14 FEB 74

M483-UK ZONE 3 (SHORT ROUNDS)

QE=30°

T=-2°F ρ=115% STD



178

MEAN RANGE (OBSERVED)=5387 M
PROBABLE ERROR=875 M =16.3% R

MEAN MUZZLE VELOCITY =293.8 M/SEC
PROBABLE ERROR = 0.6 M/SEC

Fall of Shot, M483 (7 Short Rounds) [illegible] 74

FIGURE 4

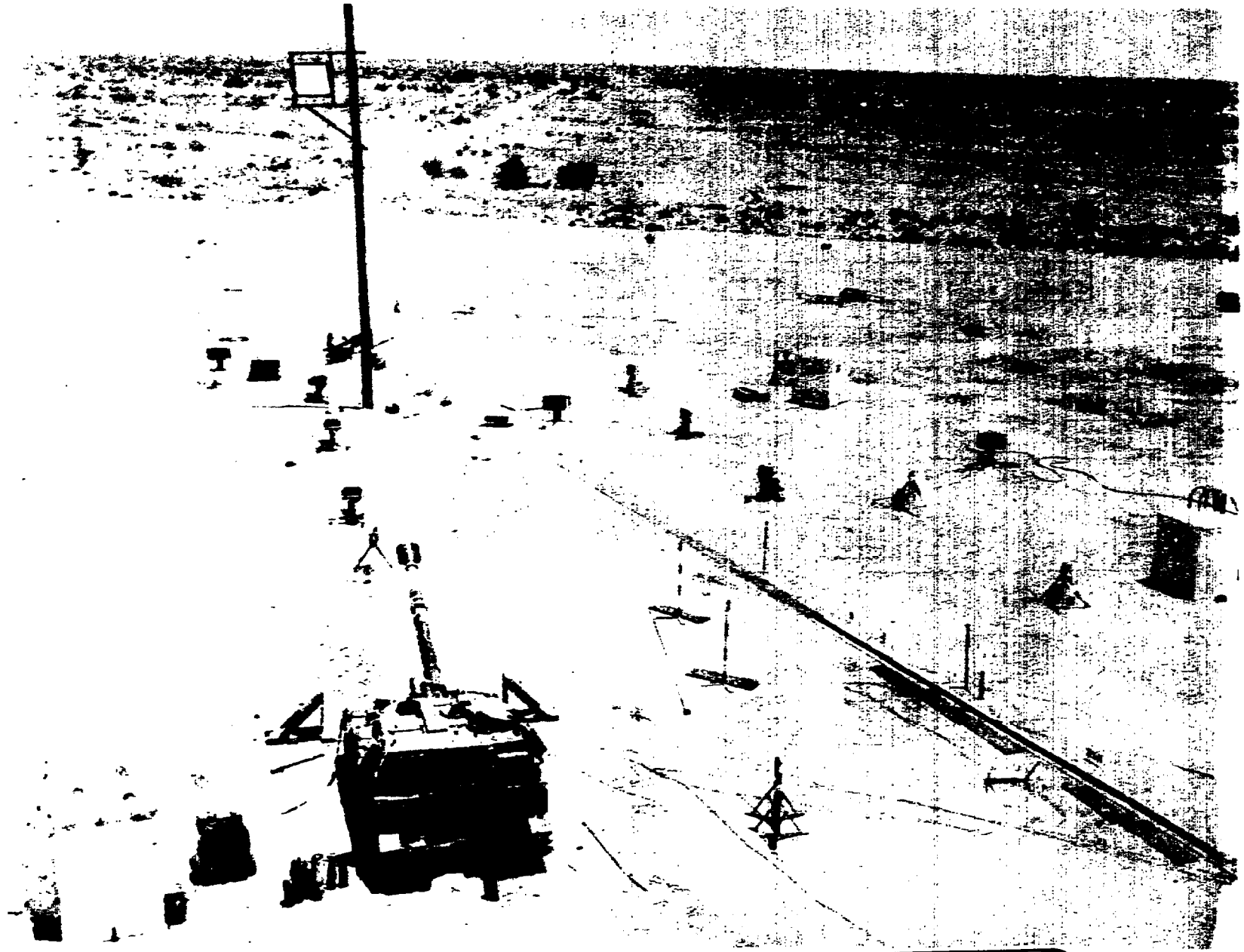
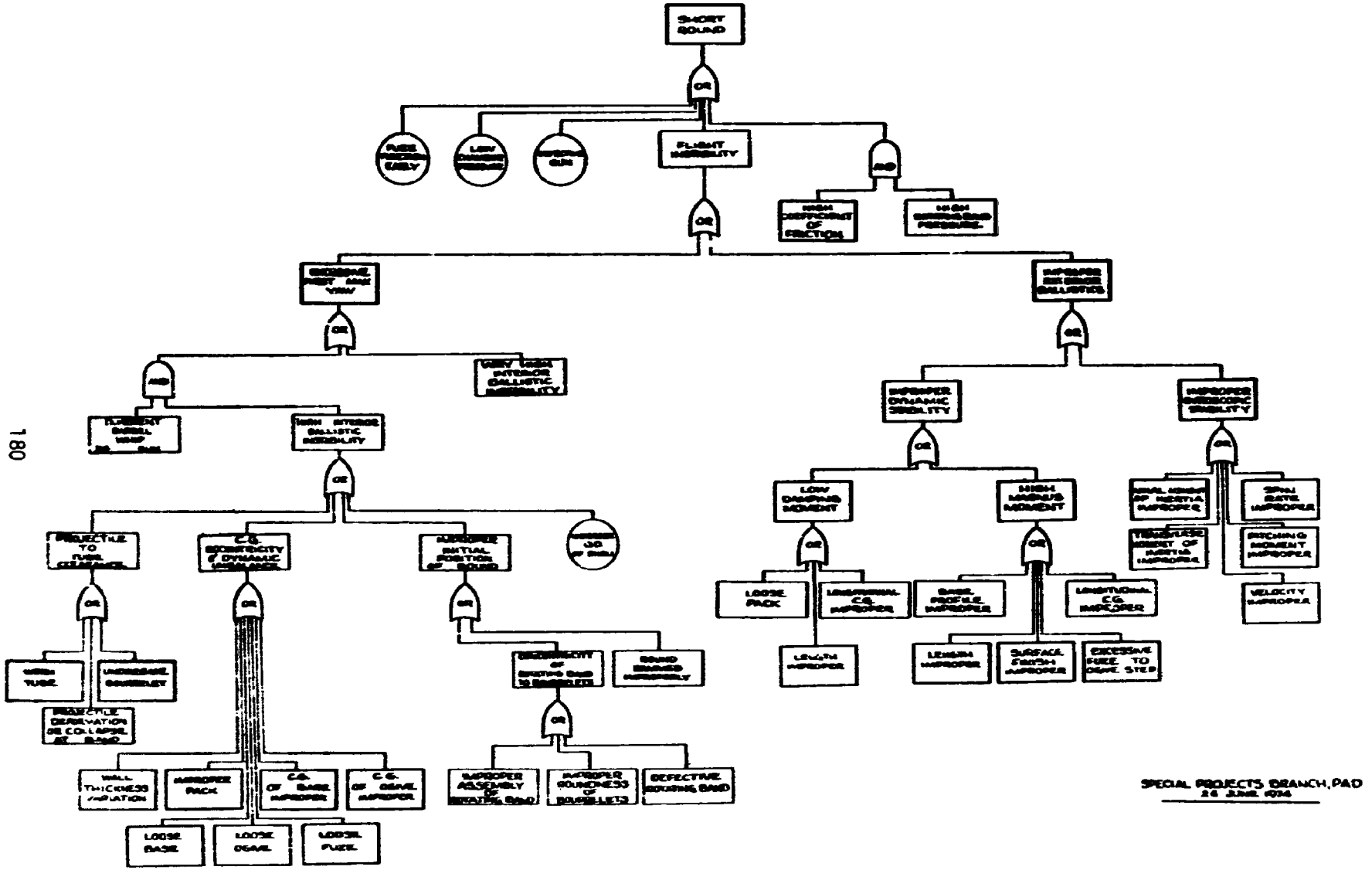


FIGURE 6

FAULT TREE FOR M483 PROJECTILE



SPECIAL PROJECTS BRANCH, PAD
24 JUNE 1974

FIGURE 5

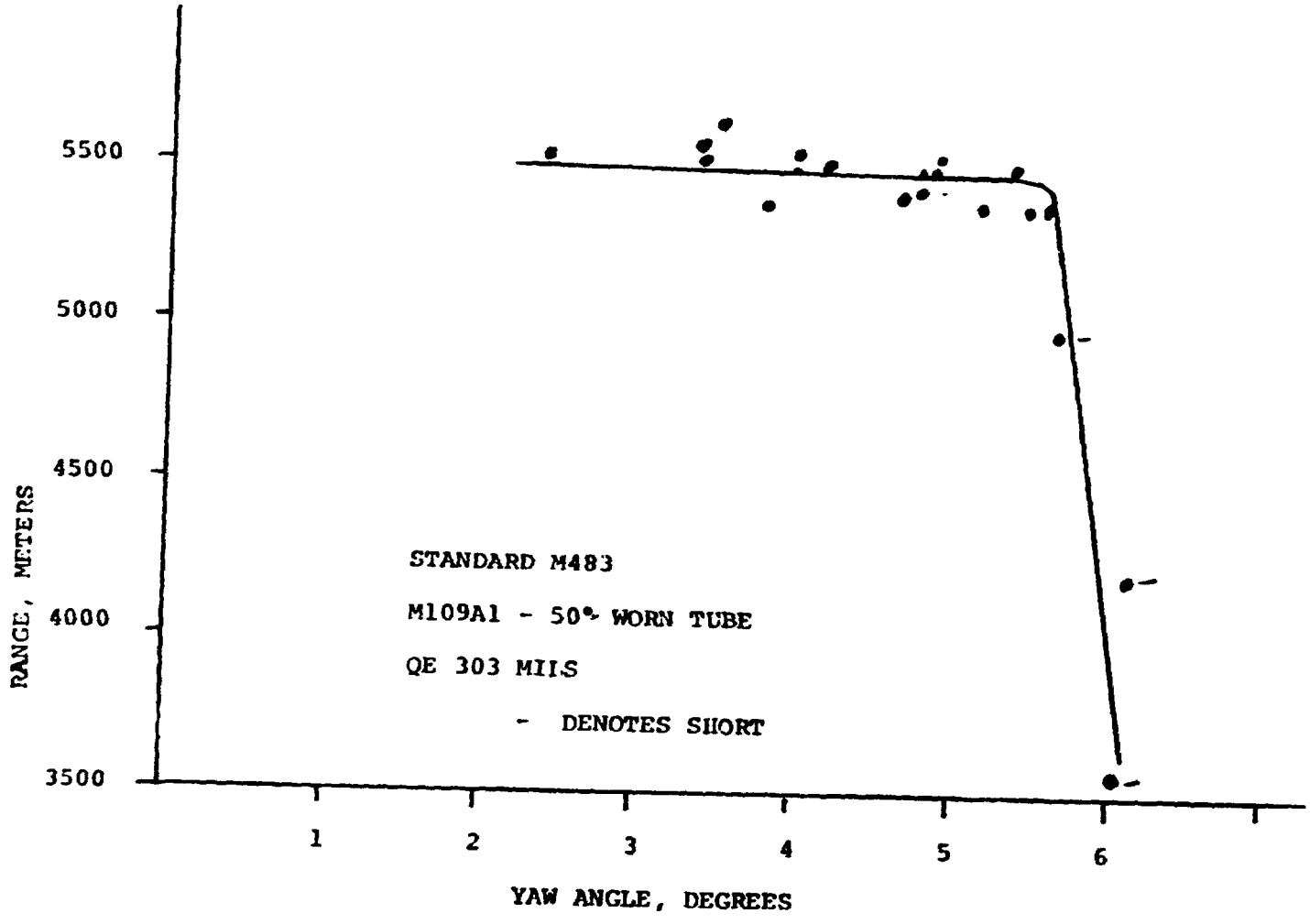

5500
5000
4500
4000
3500
RANGE, METERS
STANDARD M483
M109A1 - 50% WORN TUBE
QE 303 MILS
- DENOTES SHORT
1
2
3
4
5
6
YAW ANGLE, DEGREES

FIGURE 7

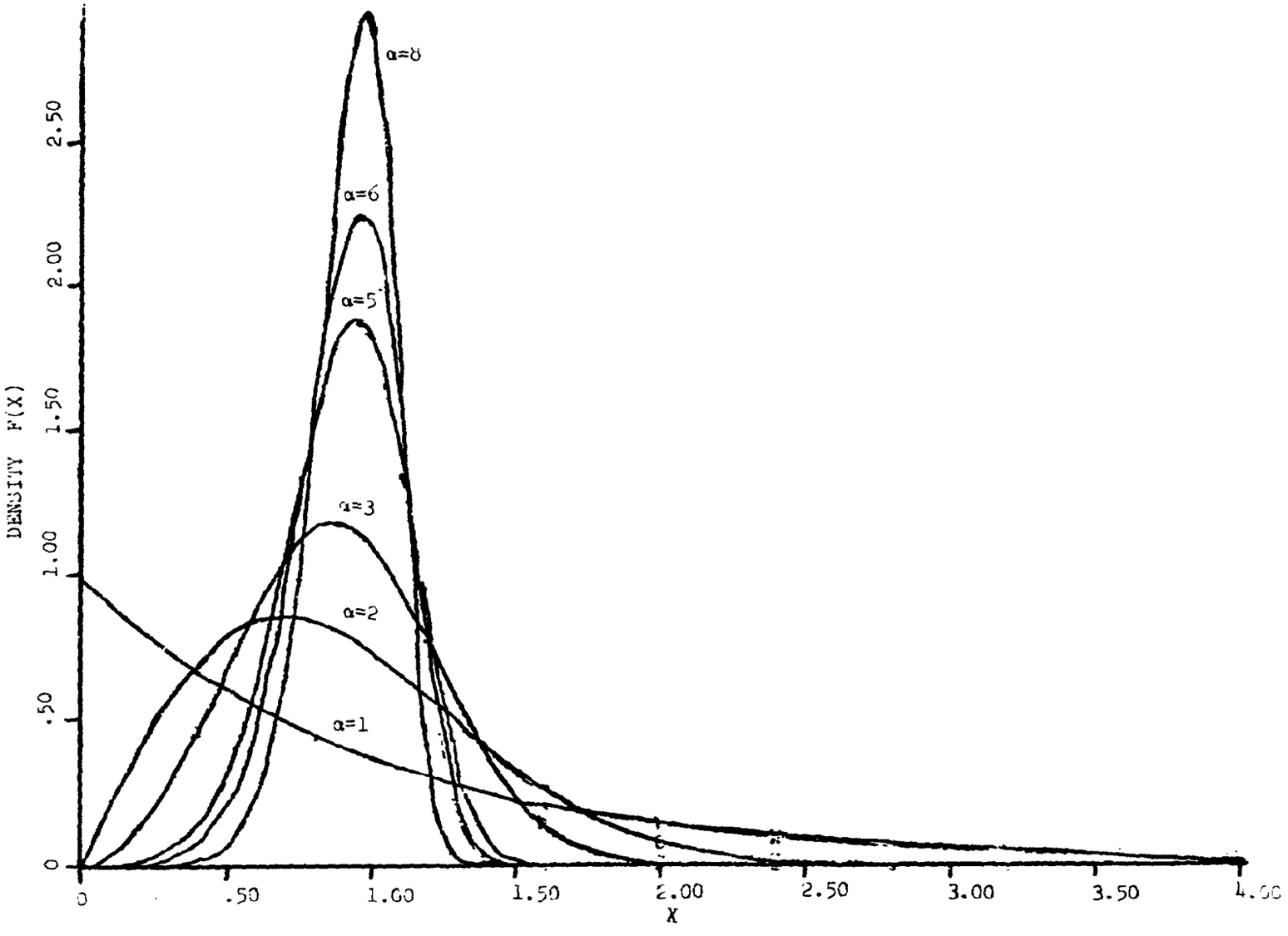

α=8
α=6
α=5
α=3
α=2
α=1
DENSITY F(X)
2.50
2.00
1.50
1.00
.50
0
0
.50
1.00
1.50
2.00
2.50
3.00
3.50
X

FIGURE 8

THE WEIBULL DISTRIBUTION FUNCTION IS:

$$F(X) = \begin{cases} 1-e^{-(X-c/\theta)^k}, & X>c \\ & \theta, k>0 \\ 0, & x<c \end{cases}$$

THE WEIBULL DENSITY FUNCTION IS:

$$f(x) = K(x-c)^{k-1}/\theta^k \, e^{-(X-c/\theta)^k}$$

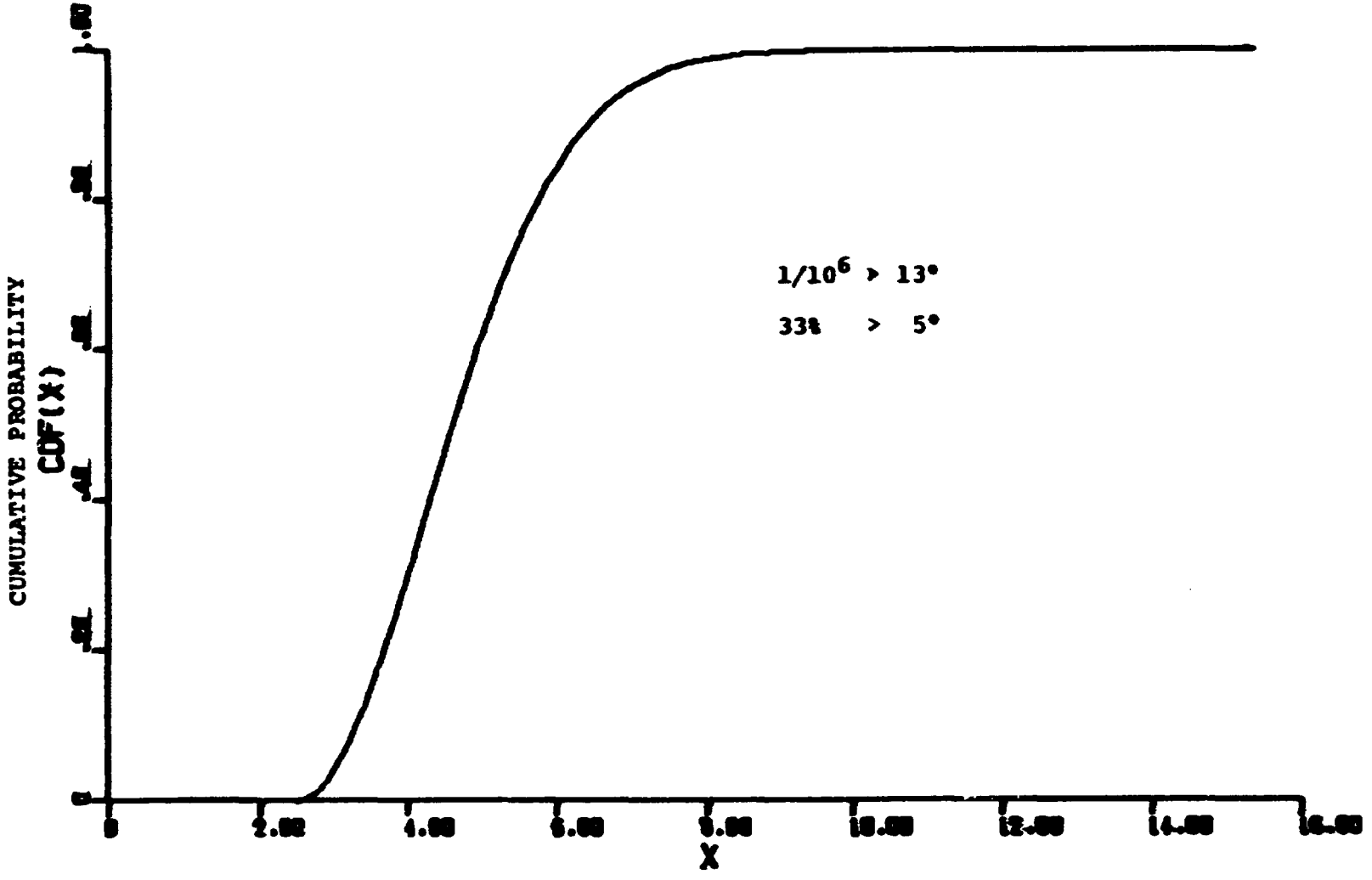
c = location parameter

θ = scale parameter

k = shape parameter

FIGURE 9

WEIBULL M483 STANDARD W/MUZZLE BRAKE



1st MAX. YAW ANGLE

FIGURE 10

WEIBULL M483 STD W/O MUZZLE BRAKE

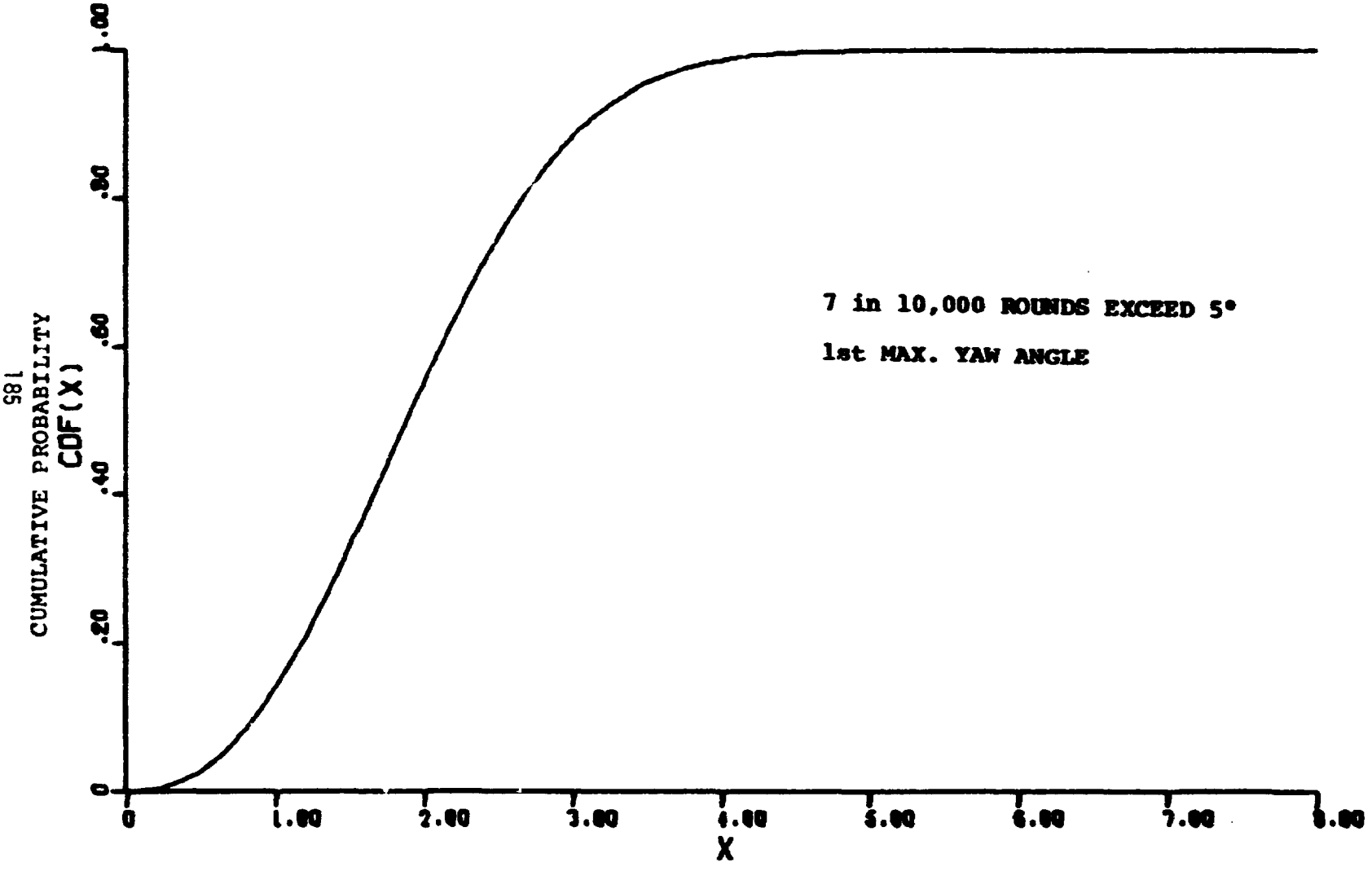


FIGURE 11

1st MAX. YAW ANGLE

FIGURE 12

M483 MOD. DESCRIPTIONS

MOD. I — STABILITY FACTOR, S_G = 2.28

- SHORT BODY (1.8 INCHES SHORT)
- NO FIBERGLASS
- CYLINDRICAL ADAPTERS
- SHORT BOATTAIL (1.5 INCHES SHORT)

MOD. II — STABILITY FACTOR, S_G = 1.74

- STANDARD LENGTH BODY
- MODIFIED FIBERGLASS
- CYLINDRICAL ADAPTERS
- SHORT BOATTAIL (1.5 INCHES SHORT)

WEIBULL FOR MOD I AND MOD II

M109A1 - 50% WORK TUBE

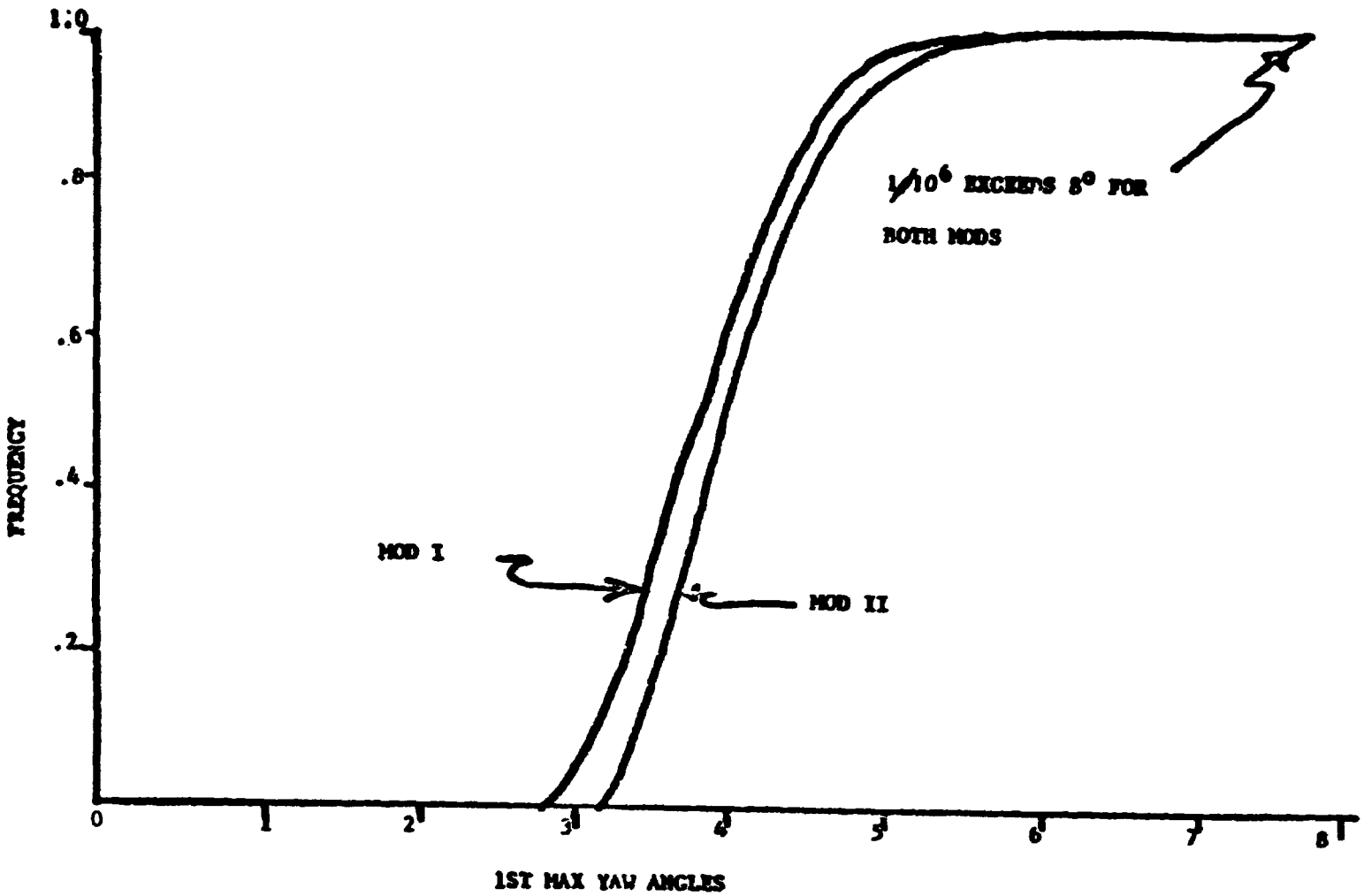


FIGURE 13.

155mm M109A1 STD BRAKE

155mm M109A1 YAW INDUCING ½ SIDE PLATE

155mm M109A1 YAW INDUCING FULL SIDE PLATE

155mm M109A1 YAW INDUCING ½ SIDE P

155mm M109A1 YAW INDUCING FULL SIDE PLATE

FIGURE 13A.

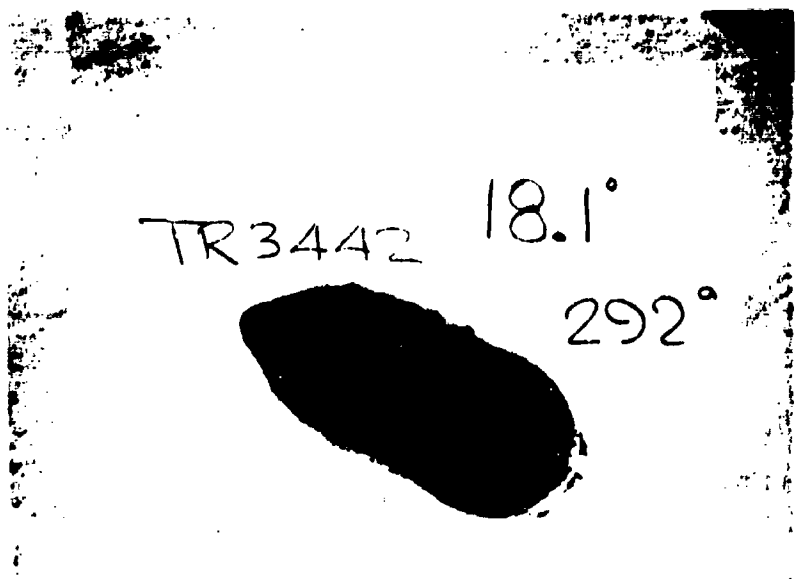


FIGURE 14

18.1° First Maximum Yaw Angle M483 MOD 2 Flies Full Range when Disturbed up to This Angle

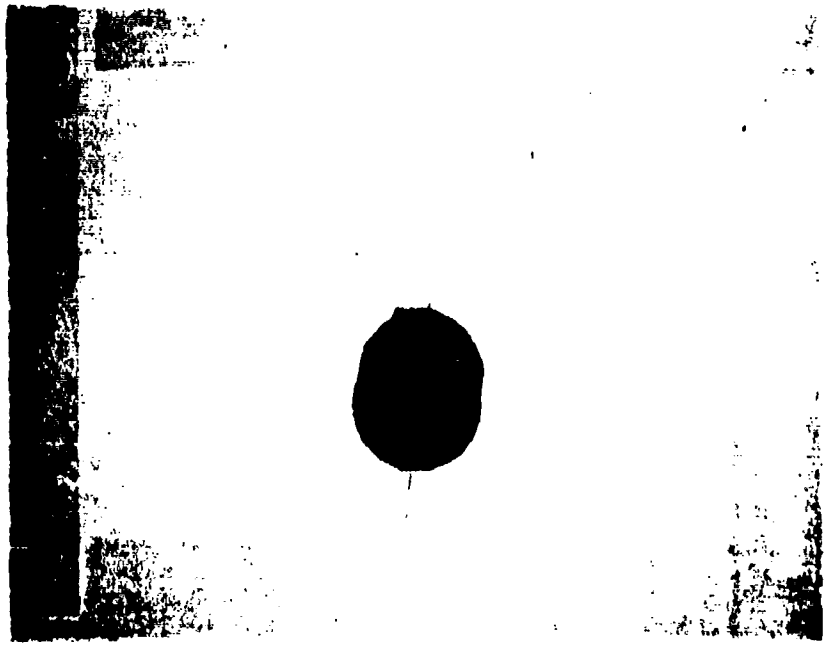


FIGURE 15

5.5° First Maximum Yaw Angle - Standard M483
Projectile Falls Short at This Angle

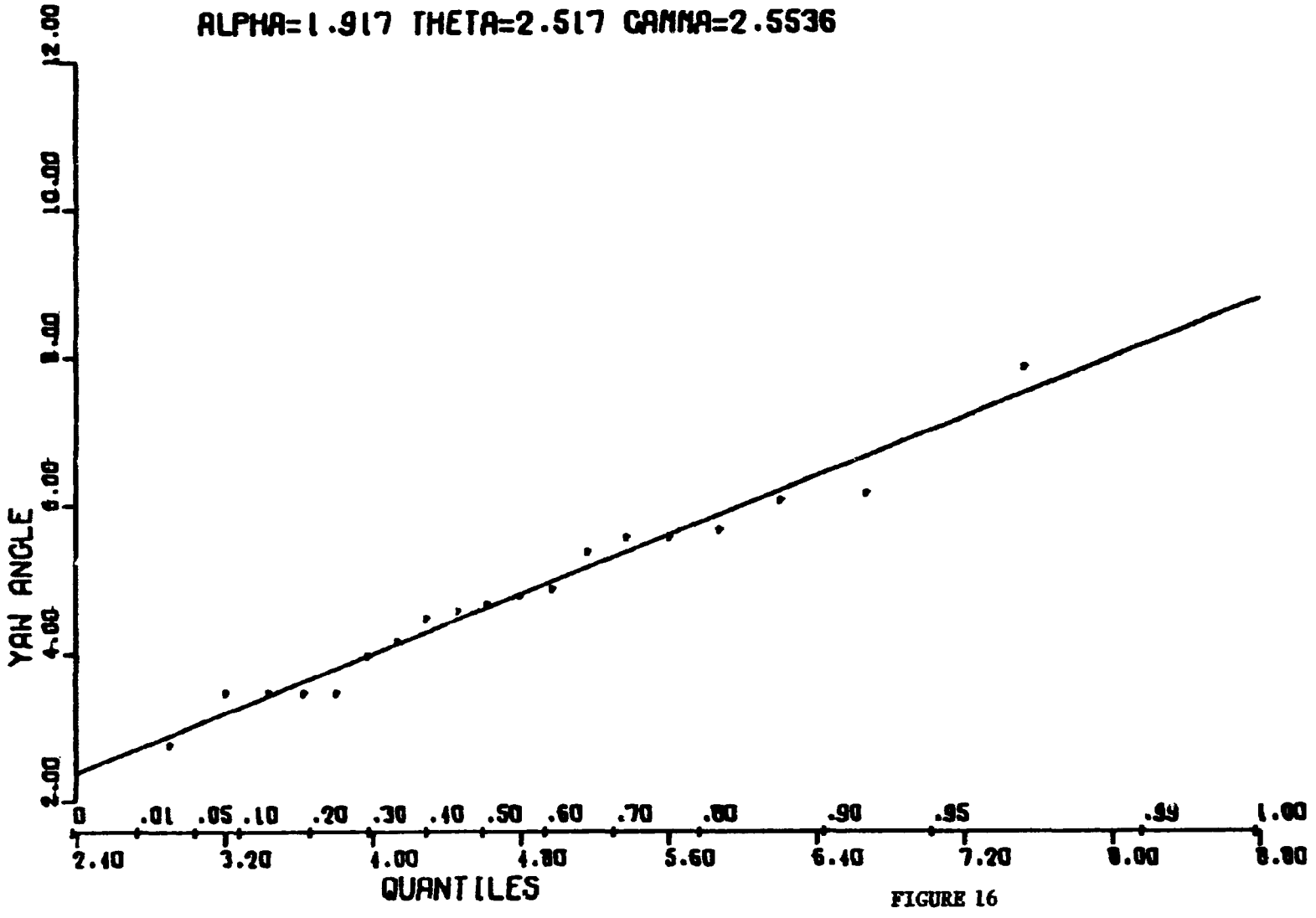

WEIBULL M109A1 STO M483 W/BRAKE
ALPHA=1.917 THETA=2.517 GAMMA=2.5536
YAW ANGLE
12.00
10.00
8.00
6.00
4.00
2.00
0
.01
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
.99
1.00
2.40
3.20
4.00
4.80
5.60
6.40
7.20
8.00
8.80
QUANTILES

FIGURE 16

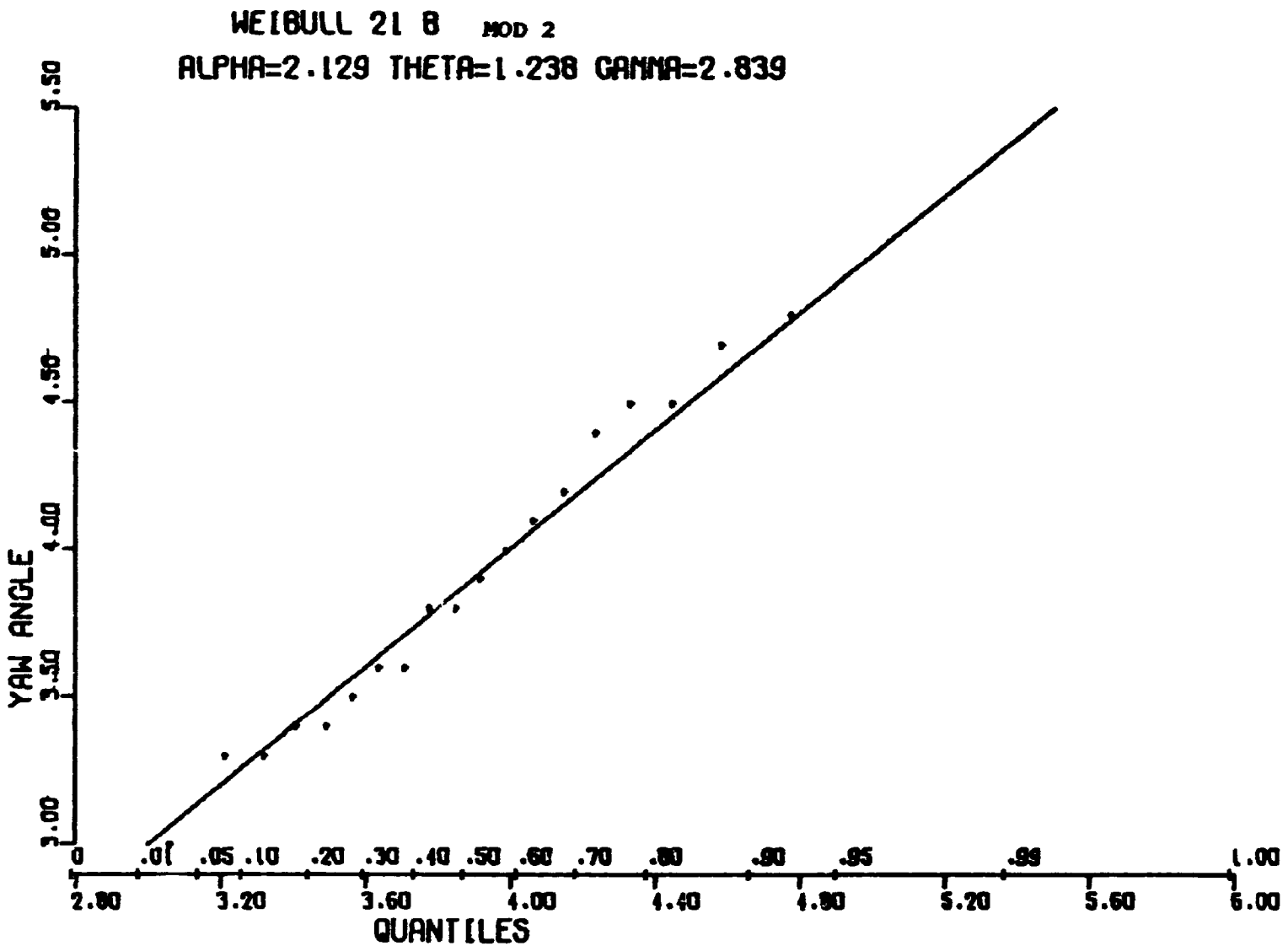

WEIBULL 21 8 MOD 2
ALPHA=2.129 THETA=1.238 GAMMA=2.839
YAW ANGLE
3.00
3.50
4.00
4.50
5.00
5.50
0
.01
.05
.10
.20
.30
.40
.50
.60
.70
.80
.90
.95
.99
1.00
2.80
3.20
3.60
4.00
4.40
4.80
5.20
5.60
6.00
QUANTILES

FIGURE 17

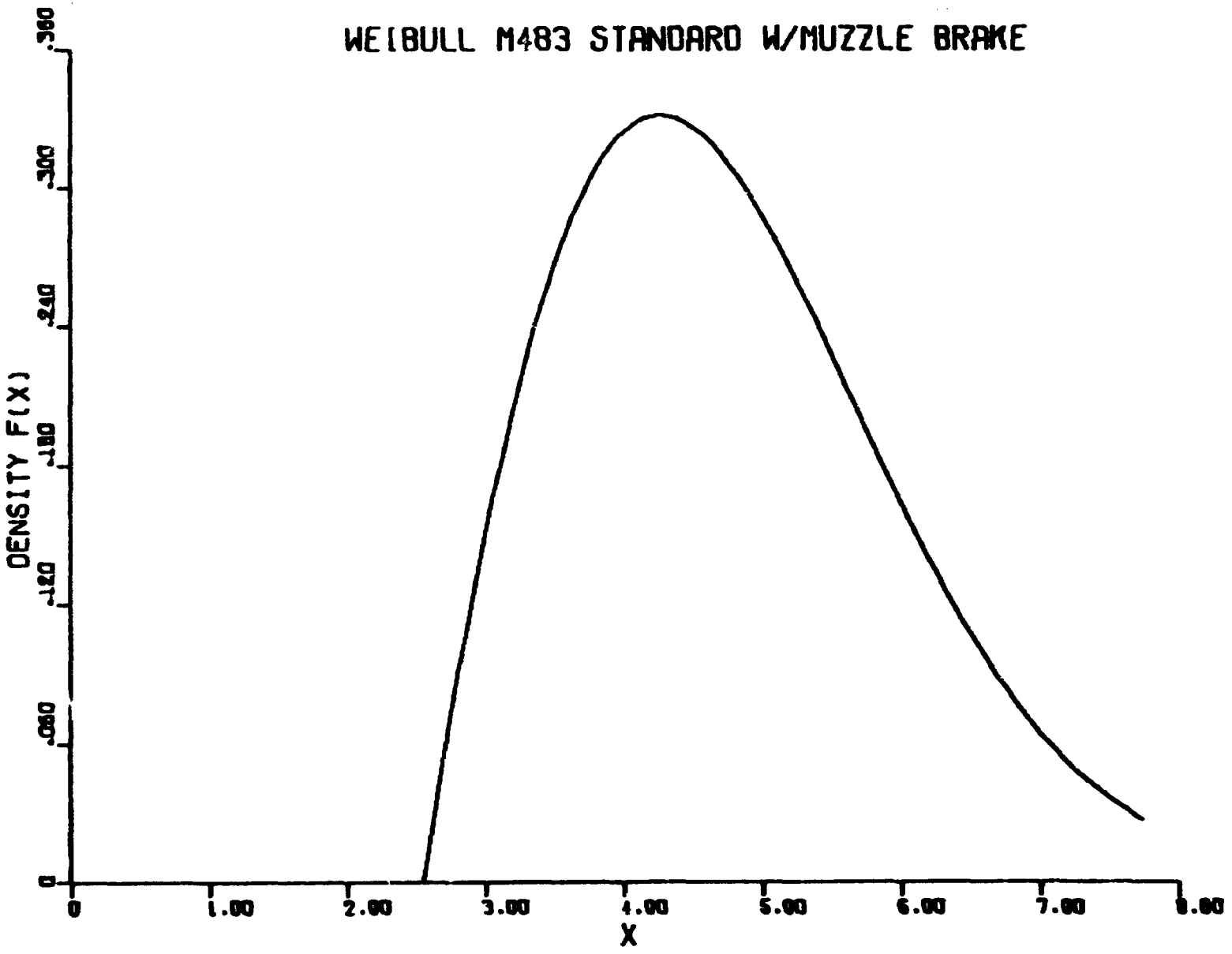

WEIBULL M483 STANDARD W/MUZZLE BRAKE
DENSITY F(X)
.360
.300
.240
.180
.120
.060
0
0
1.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
X
1st MAX. YAW ANGLE

FIGURE 18

WEIBULL M483 MOD 2

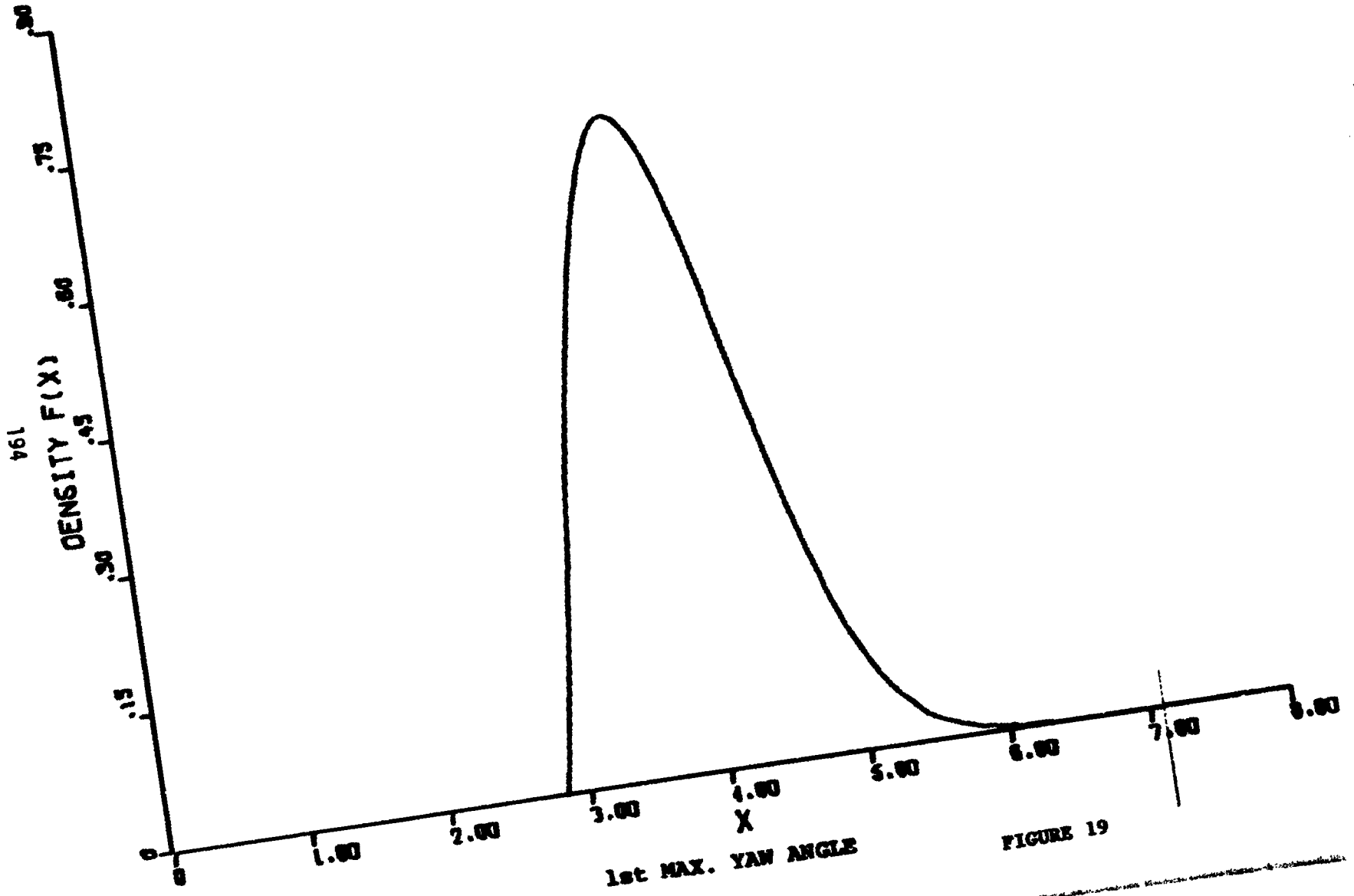


FIGURE 19

APPLICATION OF LIFE TESTING TECHNIQUES TO DETECTION DATA

Carl B. Bates and Jerry Thomas
Applications Group
Methodology and Resources Directorate
US Army Concepts Analysis Agency
Bethesda, Maryland

ABSTRACT. Life testing techniques for censored sample data are discussed. Singly and progressive censoring of type I and type II are defined. The detection phenomenon involving observers not always detecting targets is placed in the framework of progressively censored sampling. Maximum likelihood estimates for the parameters of the two-parameter Weibull distribution are given, and a test statistic is presented for comparing two Weibull distributions fitted to censored sample data. Weibull distributions of sample sizes 500, 250, and 100 having 0, 10, and 20 percent censored are simulated. The shape parameter is varied over the range 1.0 to 3.5 and equality of pairs of the distributions is tested. The relationships between Beta and the Beta difference that is distinguishable are given for each of the three sample sizes. For the largest sample size, at the 0.5-level of significance, the Beta difference that is distinguishable varied from 0.15 for small shape parameter values to 0.38 for large shape parameter values. For the 100 sample size distribu-tions, the Beta difference distinguishable varied from 0.30 to 0.73.

I. INTRODUCTION

The detection, identification, and localization of enemy targets is an integral part of many US Army studies. These studies may be classified into either computer simulated experimentation or field conducted experimentation. Field experimentation involving the detection process is usually performed to estimate or compare the effectiveness of materiel or methods of employment. Often empirical data from the field experimentation is then used as input to computer simulation models, or the analysis results of the empirical data are used to provide the basis of simulating detection in computer simulation models.

Because of the "no detections" (observers not detecting exposed targets) which occur in field experimentation involving detection processes, the analysis of empirical detection data presents unique problems. In the sections which follow, the analysis problems are discussed and a proposed analysis methodology is presented and illustrated.

II. PROBLEM DESCRIPTION AND BACKGROUND

A. Problem Description

A field experiment involving candidate land combat systems is designed and conducted. One of the many measures of effectiveness of the systems is detection time. During the conduct of the experiment, however, the systems do not always detect exposed enemy targets. Therefore, detection time data is not collected for all of the planned trials of the field experiment. Consequently, the original orthogonal design for the experiment is nonorthogonal with respect to the response variable,

detection time. The objective of this report is to present a method of analysis which uses both the detection times of detected targets and the exposure times of undetected targets.

B. Background

Land combat experimentation involving the detection of targets invariably results in targets not being detected for some of the experimental trials, e.g., Caviness et al. (1972) and McKinney et al. (1971) and (1972). Treating the "no detect" trials as missing values and applying one of the statistical techniques for estimating missing values does not have appeal because it does not utilize all the available information from the experimental data, namely, the duration of the time that line-of-sight existed between the observer and the target. Ignoring the no detect trials and analyzing only the data from trials for which a detection did occur does not have appeal for the same reason. Moreover, analyses based on all available experimental data addresses the unconditional detection probability of interest, whereas analyses based on only trials for which a detection did occur addresses the conditional probability of detection, given a detection has occurred.

A search for a proper method of analysis of the detection times which would utilize the target exposure times of the no detect trials led to the area of life testing. It was concluded that the detection phenomenon when all targets are not detected is similar to the censored sample situation in life testing.

III. LIFE TESTING

In life testing a number (N) of components are tested and the time to a component's failure is recorded. If components are withdrawn from the test before failure (in our case a target passes from an exposed state to a concealed state without being detected) the sample is termed censored. Censoring may be of two types:

1. Type I - in which at some predetermined fixed time, say t_0, testing is terminated, or
2. Type II - in which after some predetermined fixed number, say n, of sample items fail, testing is terminated.

With each type of censoring, the collected data consists of the n failure times $t_1, t_2, \ldots, t_n$, plus the information that the remaining (N-n) items survived beyond the time of termination, t_0 for Type I and t_n for Type II.

The above described censoring is termed singly censored samples. If, however, the initial censoring results in withdrawal of only a portion of the surviving items, with some remaining under test until ultimate failure or until a subsequent stage of censoring is performed, we have progressively (multiple) censored samples. In general then censoring occurring progressively in k stages at times T_i; i=1,2,...,k, and at each i^{th} stage of censoring r_i sample items are selected randomly from the survivals at time T_i and removed (that is, censored) from further observation. This is analogous to our detection phenomenon. We have a target coming from a concealed state to an exposed state just as a test item starting under observation during test. If, however, a target passes from an exposed state back to a concealed state without being

detected, it is removed from further observation at a time T_i (equal to the target's total exposure time). Further, in our case each of the k r_i equal one because in general the exposure times of any two or more undetected targets are not identical.

Past experience has shown a positive skewness in the empirical data distributions of time variables associated with the target detection process, Bates (1971) and McKinney et al. (1971) and (1972). Moreover, in McKinney et al. (1972) it was shown that the two-parameter Weibull distribution gave adequate approximations to detection time sample distributions. In the probability density function (pdf) of the two-parameter Weibull distribution,

$$f(x) = (\beta/\alpha^{\beta})x^{\beta-1} \exp[-(x/\alpha)^{\beta}]; \; x \geq 0, \; \alpha > 0, \; \beta > 0, \qquad (1)$$

α is the scale parameter and β is the shape parameter.

The Weibull distribution provides considerable flexibility for approximating a variety of distributions. When $\beta = 1$ we have the exponential distribution and when $\beta = 3.5$ we have a distribution very close to the normal distribution. In FIGURE 1 on the next page, the Weibull pdf is shown for three different shape parameters. The middle curve is a positively skewed distribution similar to that of our target detection times.

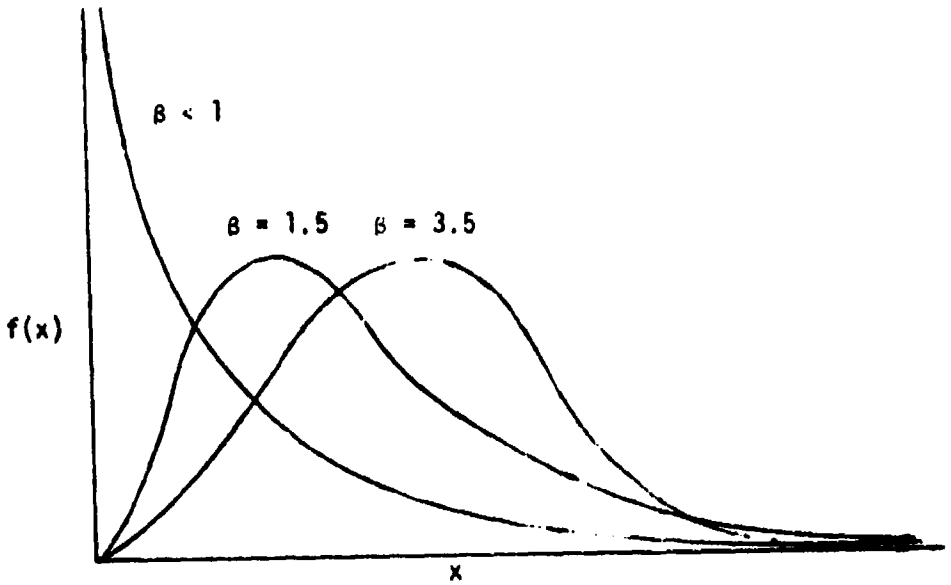


FIGURE 1, Weibull Probability Density Function

The flexibility of the Weibull distribution can be further illustrated in terms of the cumulative distribution function (cdf). In the context of our detection problem, the cdf $F(x_i)$ is the probability of detection by time x_i. FIGURE 2a is an S shaped cdf similar to that of a normal distribution. FIGURE 2b illustrates the cdf of a Weibull distribution having the same shape parameter as the distribution in FIGURE 2a, but a larger scale parameter. FIGURE 2c has the same scale parameter as FIGURE 2a, but a smaller shape parameter.

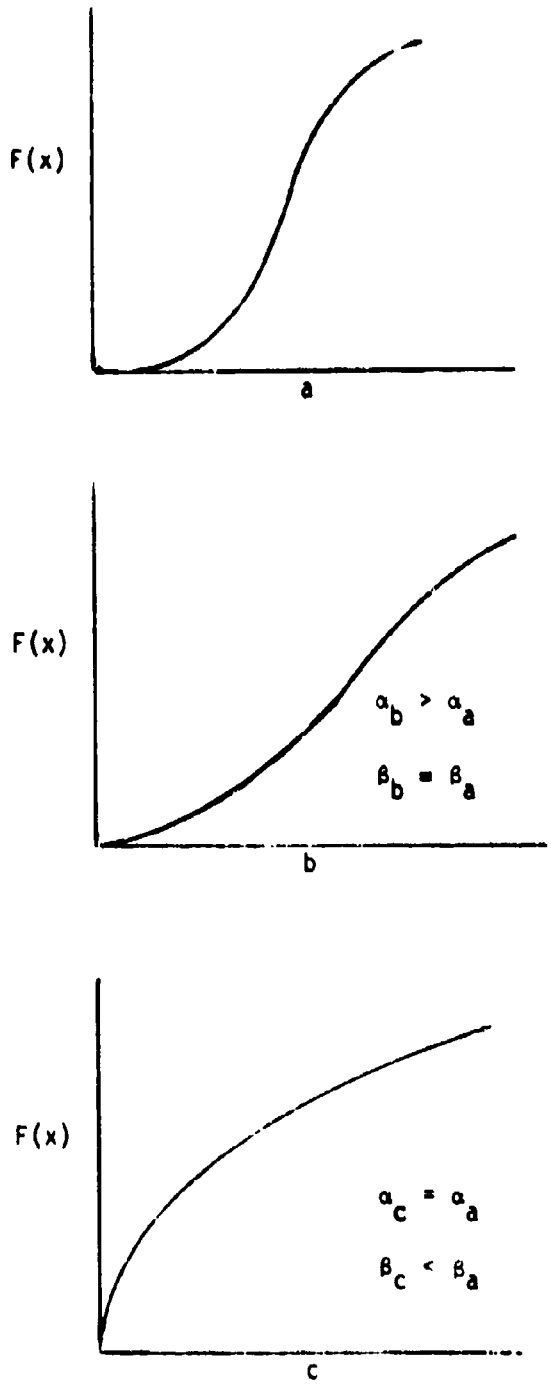


FIGURE 2, Weibull Cumulative Distribution Function

IV. ESTIMATION

The first step in the analysis process is the approximation of the distribution of target detection times. This involves estimating the two parameters, α and β, of equation (1). Substituting $\hat{\alpha}$ and $\hat{\beta}$ for α and β in f(x) gives the approximation distribution, $\hat{f}(x)$, of target detection times. The estimation technique which is employed evolved from life testing.

Cohen (1963) shows that although intermediate steps in the derivations differ, the maximum likelihood estimation equations for Type I and Type II progressively censored samples yield the same end result. The maximum likelihood estimation equations for the two-parameter Weibull distribution are given in Cohen (1965). The equations are nonlinear in the parameters and must, therefore, be solved by iterative procedures. He solves the expression,

$$[(\Sigma^{*} x_i^{\hat{\beta}} \ln x_i / \Sigma^{*} x_i^{\hat{\beta}}) - (1/\hat{\beta})] = (1/n)\Sigma \ln x_i \qquad (2)$$

for $\hat{\beta}$. The asterisk denotes that the summation is over the entire sample with the r_i observations censored at time T_i assigned the value $x_i = T_i$. Then, substituting $\hat{\beta}$ obtained from equation (2) into the other maximum likelihood estimation equation, $\partial \ln L/\partial\alpha$, and solving for $\hat{\alpha}$ he gets

$$\hat{\alpha} = (\Sigma^{*} x_i^{\hat{\beta}}/n)^{(1/\hat{\beta})}, \qquad (3)$$

where ln L is the logarithm of the likelihood function. Substitution of the two obtained parameter estimates, $\hat{\alpha}$ and $\hat{\beta}$, into equation (1) yields the desired approximation distribution $\hat{f}(x)$.

The mean of f(x) is

$$E(x) = \hat{\alpha}\,\Gamma(1 + 1/\hat{\beta}), \tag{4}$$

and the approximate variance is

$$V(x) = (\partial f/\partial\alpha)^2 V(\hat{\alpha}) + (\partial f/\partial\beta)^2 V(\hat{\beta}) + 2(\partial f/\partial\alpha)(\partial f/\partial\beta)\mathrm{Cov}(\hat{\alpha},\hat{\beta}). \tag{5}$$

V. HYPOTHESIS TESTING

Suppose that in a field experiment two candidate detection devices are under study. One of the primary objectives of the experiment is to compare the detection distributions of the two devices and make inferences concerning the equality of the two populations. After applying the estimation techniques in the previous section to the empirical detection data collected on the performance of the two devices to approximate the distribution for each device, we are now interested in comparing these two distributions. Specifically the null hypothesis,

$$H_o: \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix} = \begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix}, \tag{6}$$

is tested against the two sided alternative hypothesis,

$$H_a: \begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix} \neq \begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix}. \tag{7}$$

The test statistic for testing the null hypothesis against the alternative hypothesis is Q, where

$$Q = [\hat{\alpha}_2 - \hat{\alpha}_1 , \ \hat{\beta}_2 - \hat{\beta}_1] \begin{bmatrix} \sigma^2(\hat{\alpha}) & \sigma(\hat{\alpha},\hat{\beta}) \\ \sigma(\hat{\alpha},\hat{\beta}) & \sigma^2(\hat{\beta}) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\alpha}_2 - \hat{\alpha}_1 \\ \hat{\beta}_2 - \hat{\beta}_1 \end{bmatrix}, \tag{8}$$

and where the variance-covariance matrix is

$$\begin{bmatrix} \sigma^2(\hat{\alpha}) & \sigma(\hat{\alpha},\hat{\beta}) \\ \sigma(\hat{\alpha},\hat{\beta}) & \sigma^2(\hat{\beta}) \end{bmatrix} = \begin{bmatrix} V(\hat{\alpha}_1) + V(\hat{\alpha}_2) & Cov(\hat{\alpha}_1,\hat{\beta}_1) + Cov(\hat{\alpha}_2,\hat{\beta}_2) \\ Cov(\hat{\alpha}_1,\hat{\beta}_1) + Cov(\hat{\alpha}_2,\hat{\beta}_2) & V(\hat{\beta}_1) + V(\hat{\beta}_2) \end{bmatrix} \tag{9}$$

Equation (8) is a quadratic form and is approximately distributed as a Chi-square variate with two degrees of freedom, see for example, Mood (1950), Rao (1952), or Wilks (1962). That is,

$$Q \sim \chi^2(2). \tag{10}$$

An inspection of equation (8) shows that close agreement between the two distributions yields a small statistic, while a large difference between the two yields a large statistic. Therefore, the critical region of the test is the upper tail of the χ^2-distribution. Consequently, to test the null hypothesis of equation (6), compare Q with $\chi^2(1-\alpha,2)$. If $Q \geq \chi^2(1-\alpha,2)$, reject the null hypothesis at the α-level of significance; otherwise do not reject the null hypothesis. By rejecting the null hypothesis, we are saying that the two detection distributions are not equal.

VI. TEST DISCRIMINATION

A. General

In the previous section it was seen that the determination of a difference between distributions is dependent upon the scale parameter, α, and the shape parameter, β. For this study it was decided to set α equal to 25 and concentrate our efforts on the shape parameter, β. When $\beta = 1$, the Weibull distribution is equivalent to the exponential distribution and when $\beta = 3.5$, the distribution is approximately normal. Since the shape of the detection distribution is expected to be within this range, shape parameter values between 1.0 and 3.5 are studied.

B. Sample size of 500

Test performance in application can be no better than the asymptotic power of the test. Because no information is available on the power of the test, an initial sensitivity analysis is performed. Consequently, large samples having a moderate amount of censoring are first studied.

Weibull distributions of sample size 500 having three different percentages of censoring (0, 10, and 20) were generated by Monte Carlo simulation. The scale parameter was arbitrarily fixed at $\alpha = 25$. The range of the shape parameter values (1.0 to 3.5) was divided into five sub-ranges of length 0.5 each. Within each sub-range β was incremented in steps of 0.1 to give six β-values, e.g., (1.0, 1.1, 1.2, 1.3, 1.4, 1.5), (1.5, 1.6, 1.7, 1.8, 1.9, 2.0), ..., (3.0, 3.1, 3.2, 3.3, 3.4, 3.5). For each of the six β-values, a Weibull distribution was generated for each of the three percentages censored. This gave eighteen distributions

for each of the five β-value sub-ranges or a total of 153 pair-wise comparisons. For completeness and anticipated follow-on analyses, summary statistics are tabulated in APPENDIX A. TABLES A-1 through A-5 contain the five sets of summary statistics of the eighteen distributions.

Within each set of eighteen distributions, all possible (153) comparisons were made between pairs of distributions. That is, the null hypothesis of equality of the two distributions, equation (6), was tested. This gave 153 Q-statistics. The corresponding $\hat{\beta}$ differences ($\hat{\beta}_j - \hat{\beta}_i; i \neq j;$ $i=1,2,\ldots,17; j=2,3,\ldots,18$) were calculated and paired with the Q-statistics. Within each set of $\hat{\beta}$ differences and Q-statistics, six different combinations existed between the percentages censored in the two distributions being compared-(0,0), (0,10), (0,20), (10,10), (10,20), and (20,20). The distribution of the 153 cases over the six combinations is shown in TABLE 1 below.

TABLE 1

CENSORING DISTRIBUTION

Combination Number	Percentage Censored (Sample j, Sample i)	Number of Samples
1	(0,0)	15
2	(0,10) or (10,0)	36
3	(0,20) or (20,0)	36
4	(10,10)	15
5	(10,20) or (20,10)	36
6	(20,20)	15

The theoretical relationship between the $\hat{\beta}$ differences and Q is parabolic. Therefore, a quadratic in $\hat{\beta}$ differences was fitted for each of the six combinations in TABLE 1, using $\hat{\beta}$ differences as the independent variable and the Q-statistic as the dependent variable. Within each of the five β-value sub-ranges, the quadratic fit for each of the six censoring combinations was evaluated for Q = 5.991, the $\chi^2(2)$ critical value for the 0.05-level of significance. This gives the difference between the shape parameters of two distributions which would be declared significant at the 0.05-level of significance. The largest variation among each set of the six $\hat{\beta}$ differences was 0.04. This is well within the variability of the generated data. The six combinations of each β-value sub-range were then "pooled" and a quadratic fit was made to each of the five sub-ranges of the 153 $\hat{\beta}$ differences. All fits were "good"; the coefficients of determination ranged from 0.90 to 0.97. Each of the five sub-range quadratic regression equations was then evaluated for two levels of significance (0.05 and 0.01) or Q = 5.991 and Q = 9.210. The resulting relationship between β and the β differences detectable for the two significance levels is graphically illustrated in FIGURE 3 on the following page.

FIGURE 3 suggests a strong linear relationship between β and the β difference that is detectable. In fact, the ratio of the plotted β differences over their respective sub-range mid-points is nearly constant for each level of significance. At the 0.05-level of significance, the ratio is approximately 0.12: at the 0.01-level of significance, it is approximately 0.15.

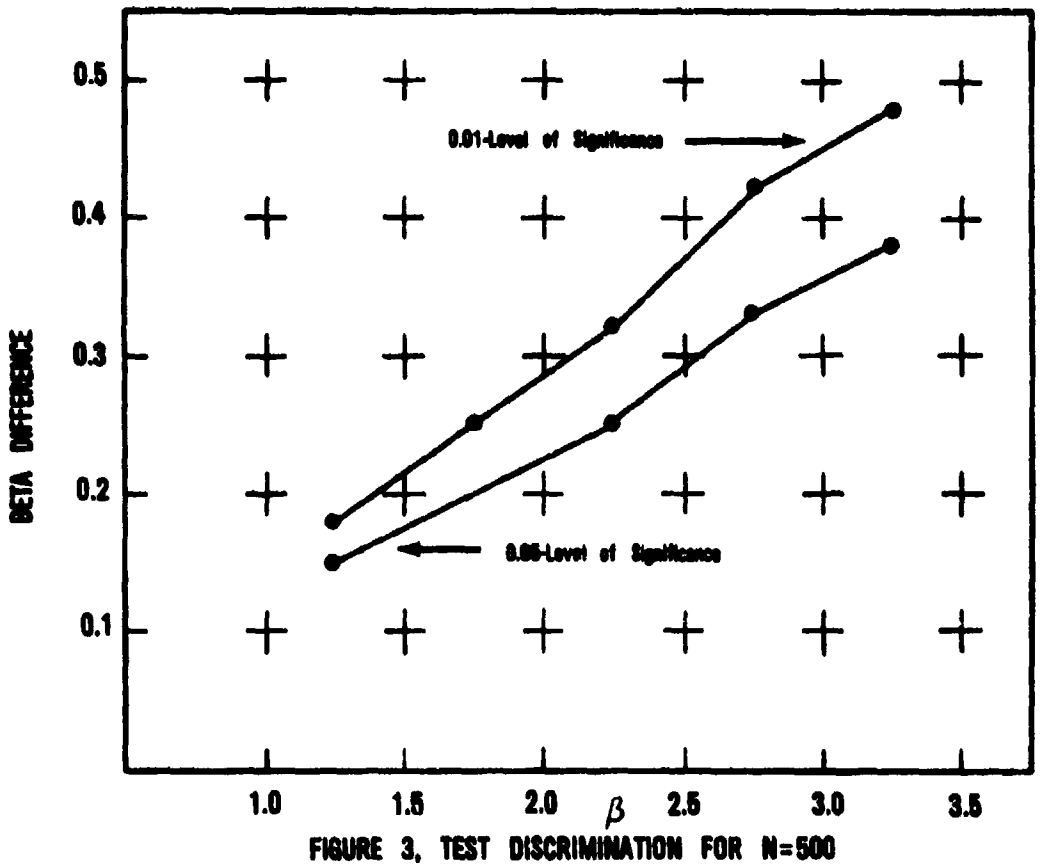


FIGURE 3. TEST DISCRIMINATION FOR N=500

FIGURES A-1 through A-5 of APPENDIX A pictorially illustrate typical distributions, within each of the five sub-ranges, which are statistically different when equality is tested at the 0.05-level of significance. Each of the five figures contains a plot of two distributions, taken from the samples shown in TABLES A-1 through A-5, respectively. The distribution having the smaller shape parameter is drawn with a solid line and its shape parameter estimate is denoted by $\hat{\beta}_1$; the distribution having the larger shape parameter is shown with a dashed line and its shape parameter is denoted by $\hat{\beta}_2$. For example in FIGURE A-1,

samples were selected from a distribution with $\beta = 1.1$ (with no censoring) and with $\beta = 1.2$ (with 10% censoring); and the two sample estimates of the shape parameter are $\hat{\beta}_1 = 1.062$ and $\hat{\beta}_2 = 1.227$. The Q-statistic for testing the null hypothesis of equation (6) is also given on each figure. In each case, the Q-statistic is between $\chi^2(0.95,2) = 5.991$ and $\chi^2(0.99,2) = 9.210$. That is, the level of significance at which the null hypothesis would be rejected is between 0.05 and 0.01. The five figures illustrate the test discrimination between distributions of different shapes over a range of shape parameter values from 1.0 to 3.5.

C. Sample Size of 250

In practice large samples are often not available. Therefore, test performance for two smaller samples ($N = 250$ and $N = 100$) are studied. The results for $N = 250$ are presented first.

Weibull distributions of sample size 250 were generated. The same scale and shape parameters and the same percentages of censoring were used as for $N = 500$. The procedure described in Section A above was repeated using $N = 250$. The summary statistics are given in TABLES B-1 through B-5 of APPENDIX B. This time the largest variation among each set of the six $\hat{\beta}$ differences was 0.07. Again, this variation is within the variability of the generated data. The Beta differences obtained from the evaluations of the five quadratic regression equations are given in TABLE 2 below. As before, there appears to be a linear relationship between β and the β difference that is detectable.

TABLE 2

BETA DIFFERENCES FOR N = 250

Significance Level	Beta				
	1.0-1.5	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5
0.05	0.20	0.29	0.37	0.45	0.54
0.01	0.24	0.36	0.47	0.55	0.67

The test discrimination for each of the five sub-ranges is illustrated in FIGURES B-1 through B-5 in APPENDIX B. The notation in the figures is the same as that described in the previous section. The distribution having the smaller shape parameter estimate is denoted by $\hat{\beta}_1$ and the larger is denoted by $\hat{\beta}_2$. The significance level of each pair of illustrated distributions is between 0.05 and 0.01. The Q-statistic is again given on each of the five figures.

D. Sample Size of 100

In the examination of the test performance for N = 100, the sub-ranges of the shape parameter values had to be reconstructed. This was because the Beta difference which is distinguishable is larger than 0.5 for shape parameters greater than 1.5. Therefore, the shape parameter range was divided into three sub-ranges rather than the five previously used. The three sub-ranges were 1.0-1.5, 1.5-2.5, and 2.5-3.5. Within the first sub-range, β was incremented in steps of 0.1 as before. But within the two larger sub-ranges, β was incremented in steps of 0.2. This gave six β-values for each of the three sub-ranges. The summary statistics of the three sets of eighteen distributions are given in TABLES C-1, C-2, and C-3.

The largest variation among each set of the six β differences was 0.08, again within the variability of the data. The Beta differences from the three quadratic regressions are given in TABLE 3. Test discrimination is pictorially illustrated in the three figures of Appendix C.

TABLE 3

BETA DIFFERENCES FOR N = 100

Significance Level	Beta 1.0-1.5	1.5-2.5	2.5-3.5
0.05	0.30	0.48	0.73
0.01	0.37	0.57	0.89

The test discrimination for all three sample sizes is shown in FIGURE 4. All three sample sizes exhibit a linear relationship between β and the β difference that is detectable. As expected, the β difference that is detectable is smaller for large sample sizes than the β difference that is detectable for small sample sizes. The dependence of the β difference that is detectable upon β is greater for small sample sizes than it is for large sample sizes. The trend of the lines for N = 100 has the steepest slope.

VII. CONCLUSIONS

The test statistic performed satisfactorily over the range of shape parameters and the percentages of censoring investigated. For the three sample sizes and the parameter values studied, test discrimination is not degraded when censoring does not exceed twenty percent of the sample size.

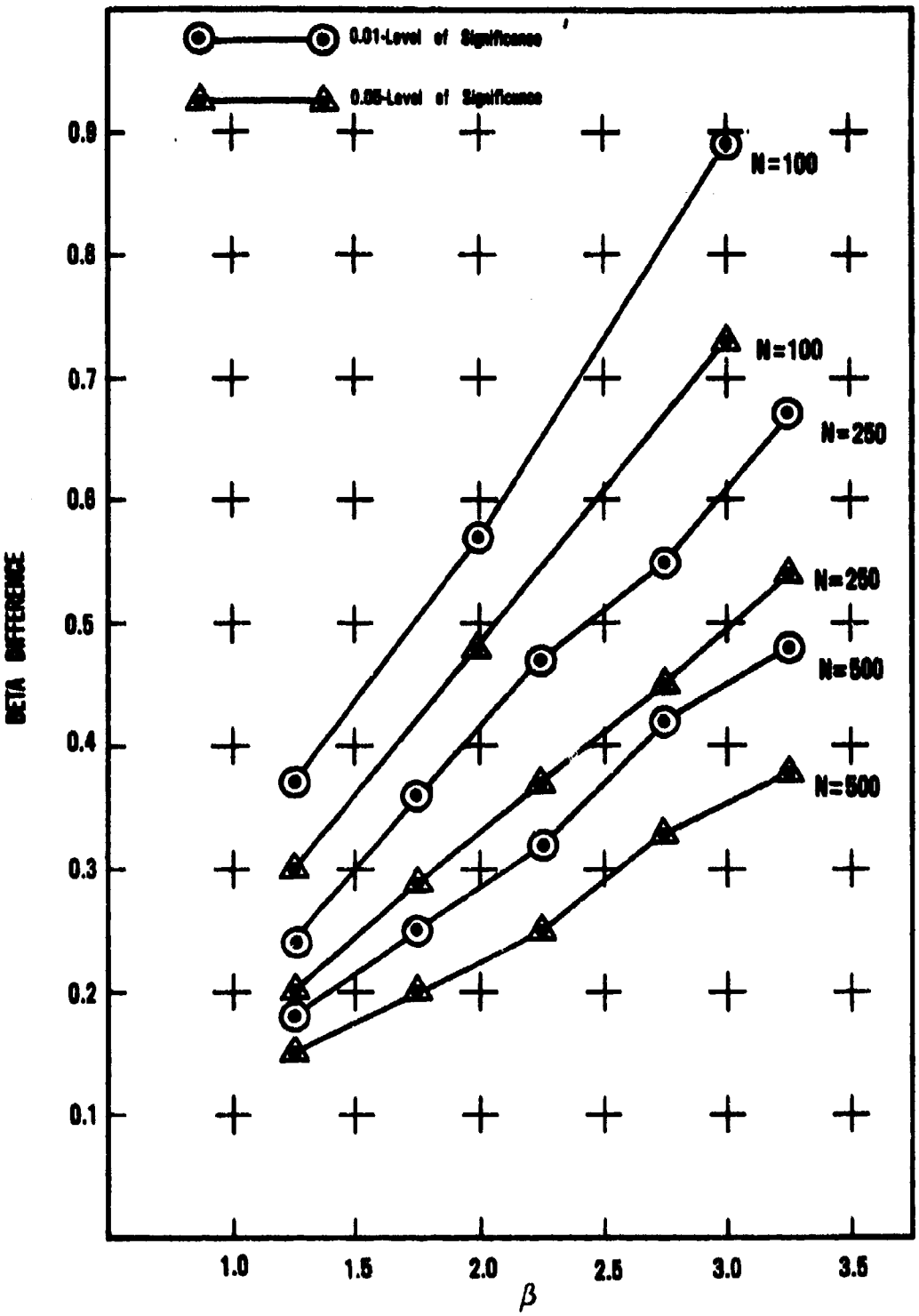


FIGURE 4, TEST DISCRIMINATION

Therefore, under moderate degrees of censoring, the Q-statistic provides a useful test statistic for testing the equality of two fitted Weibull distributions. The relationships shown in FIGURE 4 between β and the β differences that are distinguishable can serve as indicators of test discrimination. These indicators should be of value when designing target detection experimentation and when analyzing target detection data in which all exposed targets are not detected.

VIII. REFERENCES

(1) Bates, C. B. (1971), MFR, CDCSAG-SSD, 16 July 71, subject: Analysis of Detection Time Distributions, Fort Belvoir, Virginia.

(2) Caviness, J. A. et al (1972), Target Detection and Range Estimation, HumRRO-TR-72-34, Human Resources Research Organization, Fort Benning, Georgia.

(3) Cohen, A. C. (1963), Progressively Censored Samples in Life Testing, Technometrics, Vol. 5, No. 3, pp 327-339.

(4) Cohen, A. C. (1965), Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples, Technometrics, Vol. 7, No. 4, pp 579-588.

(5) McKinney, J. J. et al (1971), Attack Helicopter Survivability as a Function of Exposure Time, CDCSAG TR 11-71, Fort Belvoir.

(6) McKinney, J. J. et al (1972), A comparison of Attack Helicopter Survivability as a Function of Exposure Time at Two Geographic Locations, CDCSAG TR 11-72, Fort Belvoir, Virginia.

(7) Mood, A. M. (1950), Introduction to the Theory of Statistics, McGraw-Hill Book Co., Inc., New York.

(8) Rao, C. R. (1952), Advanced Statistical Methods in Biometric Research, John Wiley and Sons, Inc., New York.

(9) Wilks, S. S. (1962), Mathematical Statistics, John Wiley and Sons, Inc., New York.

APPENDIX A

SAMPLE SIZE OF 500

TABLE A-1, N = 500 and Shape Parameter Equal 1.0 - 1.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.750	1.0	1.053	24.251	530.968
2	10	25	27.050	1.0	1.028	26.747	677.018
3	20	25	30.912	1.0	1.023	30.625	896.403
4	0	25	25.219	1.1	1.062	24.629	538.161
5	10	25	26.285	1.1	1.098	25.379	535.669
6	20	25	27.943	1.1	1.219	26.181	465.978
7	0	25	25.637	1.2	1.188	24.179	417.413
8	10	25	26.702	1.2	1.227	24.979	418.800
9	20	25	27.661	1.2	1.123	26.515	559.533
10	0	25	25.975	1.3	1.279	24.071	359.644
11	10	25	25.717	1.3	1.312	23.708	332.413
12	20	25	27.895	1.3	1.346	25.592	369.171
13	0	25	24.131	1.4	1.461	21.857	231.181
14	10	25	25.234	1.4	1.518	22.747	233.317
15	20	25	25.710	1.4	1.388	23.466	293.222
16	0	25	25.669	1.5	1.502	23.169	246.870
17	10	25	25.341	1.5	1.427	23.029	268.047
18	20	25	28.123	1.5	1.473	25.445	308.610

TABLE A-2, N = 500 and Shape Parameter Equal 1.5 - 2.0

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.023	1.5	1.599	21.540	190.289
2	10	25	26.163	1.5	1.546	23.537	241.584
3	20	25	27.159	1.5	1.560	24.410	255.493
4	0	25	25.947	1.6	1.585	23.284	225.728
5	10	25	26.201	1.6	1.763	23.326	186.667
6	20	25	24.382	1.6	1.590	21.874	198.292
7	0	25	24.338	1.7	1.700	21.716	172.883
8	10	25	24.927	1.7	1.775	22.183	166.842
9	20	25	25.551	1.7	1.836	22.701	164.275
10	0	25	24.856	1.8	1.839	22.083	154.996
11	10	25	26.119	1.8	1.795	23.231	179.268
12	20	25	28.095	1.8	1.934	24.917	180.238
13	0	25	24.585	1.9	1.899	21.816	142.759
14	10	25	25.649	1.9	1.769	22.830	177.822
15	20	25	26.921	1.9	1.842	23.916	181.228
16	0	25	24.507	2.0	1.945	21.732	135.675
17	10	25	25.258	2.0	1.954	22.396	142.935
18	20	25	26.617	2.0	2.019	23.585	149.435

TABLE A-3, N = 500 and Shape Parameter Equal 2.0 - 2.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	25.833	2.0	2.050	22.885	136.899
2	10	25	25.150	2.0	2.039	22.282	131.058
3	20	25	25.268	2.0	1.995	23.280	148.812
4	0	25	25.608	2.1	2.069	22.683	132.273
5	10	25	25.918	2.1	2.003	22.968	143.820
6	20	25	25.970	2.1	2.096	23.002	132.933
7	0	25	25.965	2.2	2.372	23.012	106.574
8	10	25	25.956	2.2	2.253	22.990	116.586
9	20	25	25.952	2.2	2.317	22.993	110.950
10	0	25	25.530	2.3	2.281	22.615	110.349
11	10	25	25.119	2.3	2.270	22.251	107.766
12	20	25	26.443	2.3	2.387	23.439	109.267
13	0	25	24.427	2.4	2.329	21.643	97.346
14	10	25	25.088	2.4	2.399	22.240	97.532
15	20	25	26.236	2.4	2.577	23.298	94.108
16	0	25	24.550	2.5	2.614	21.809	80.384
17	10	25	24.814	2.5	2.478	22.012	90.150
18	20	25	26.159	2.5	2.585	23.231	93.100

TABLE A-4, N = 500 and Shape Parameter Equal 2.5 - 3.0

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.657	2.5	2.573	21.894	83.374
2	10	25	25.261	2.5	2.688	22.461	81.143
3	20	25	26.613	2.5	2.690	23.664	89.922
4	0	25	24.947	2.6	2.637	22.168	81.790
5	10	25	25.938	2.6	2.607	23.040	90.202
6	20	25	25.906	2.6	2.786	23.064	80.241
7	0	25	23.718	2.7	2.571	21.060	77.219
8	10	25	25.821	2.7	2.669	22.953	85.816
9	20	25	25.625	2.7	2.705	22.789	82.562
10	0	25	25.040	2.8	2.750	22.282	76.633
11	10	25	25.372	2.8	3.100	22.690	64.130
12	20	25	25.394	2.8	2.975	22.668	68.911
13	0	25	25.166	2.9	2.856	22.426	72.552
14	10	25	26.543	2.9	2.988	23.698	74.736
15	20	25	25.832	2.9	3.054	23.085	68.194
16	0	25	25.034	3.0	2.804	22.292	74.090
17	10	25	25.906	3.0	2.958	23.120	72.396
18	20	25	26.220	3.0	2.836	23.359	79.728

TARLF A-5, N = 500 and Shape Parameter Equal 3.0 - 3.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.828	3.0	3.100	22.204	61.431
2	10	25	25.701	3.0	3.144	23.000	64.230
3	20	25	26.197	3.0	3.067	23.417	69.606
4	0	25	23.978	3.1	3.007	21.414	60.328
5	10	25	24.927	3.1	3.175	22.317	59.436
6	20	25	25.758	3.1	3.136	23.048	64.801
7	0	25	25.062	3.2	3.288	22.477	56.625
8	10	25	25.815	3.2	3.176	23.113	63.718
9	20	25	24.788	3.2	3.197	22.200	58.088
10	0	25	25.489	3.3	3.251	22.847	59.696
11	10	25	25.632	3.3	3.181	22.950	62.642
12	20	25	25.384	3.3	3.410	22.808	54.600
13	0	25	25.141	3.4	3.369	22.675	54.652
14	10	25	25.163	3.4	3.673	22.699	47.311
15	20	25	25.401	3.4	3.457	22.840	53.421
16	0	25	24.879	3.5	3.437	22.363	51.742
17	10	25	24.816	3.5	3.525	22.337	49.316
18	20	25	25.355	3.5	3.674	22.873	48.005

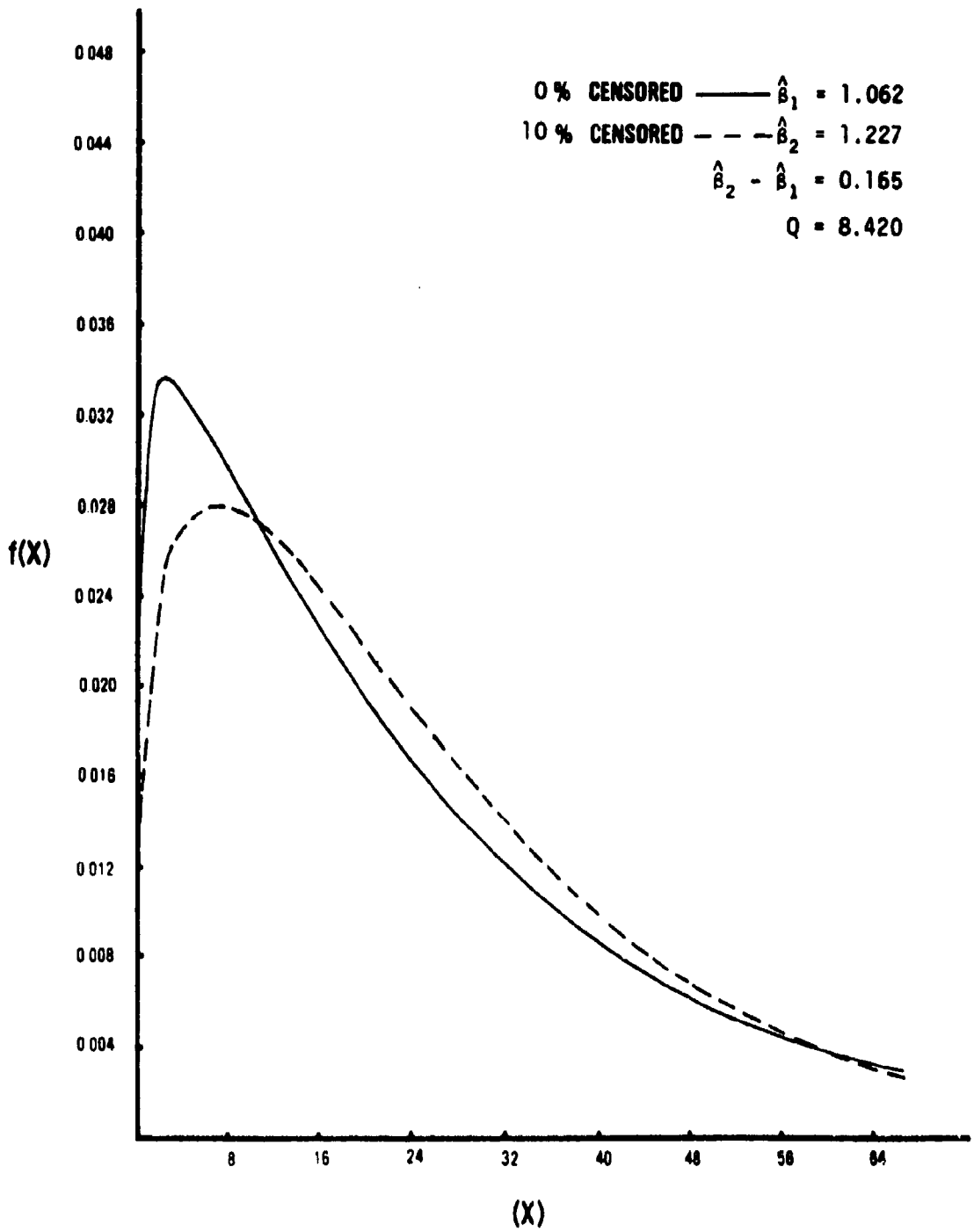


FIGURE A-1, N = 500 and Beta Between 1.0 and 1.5

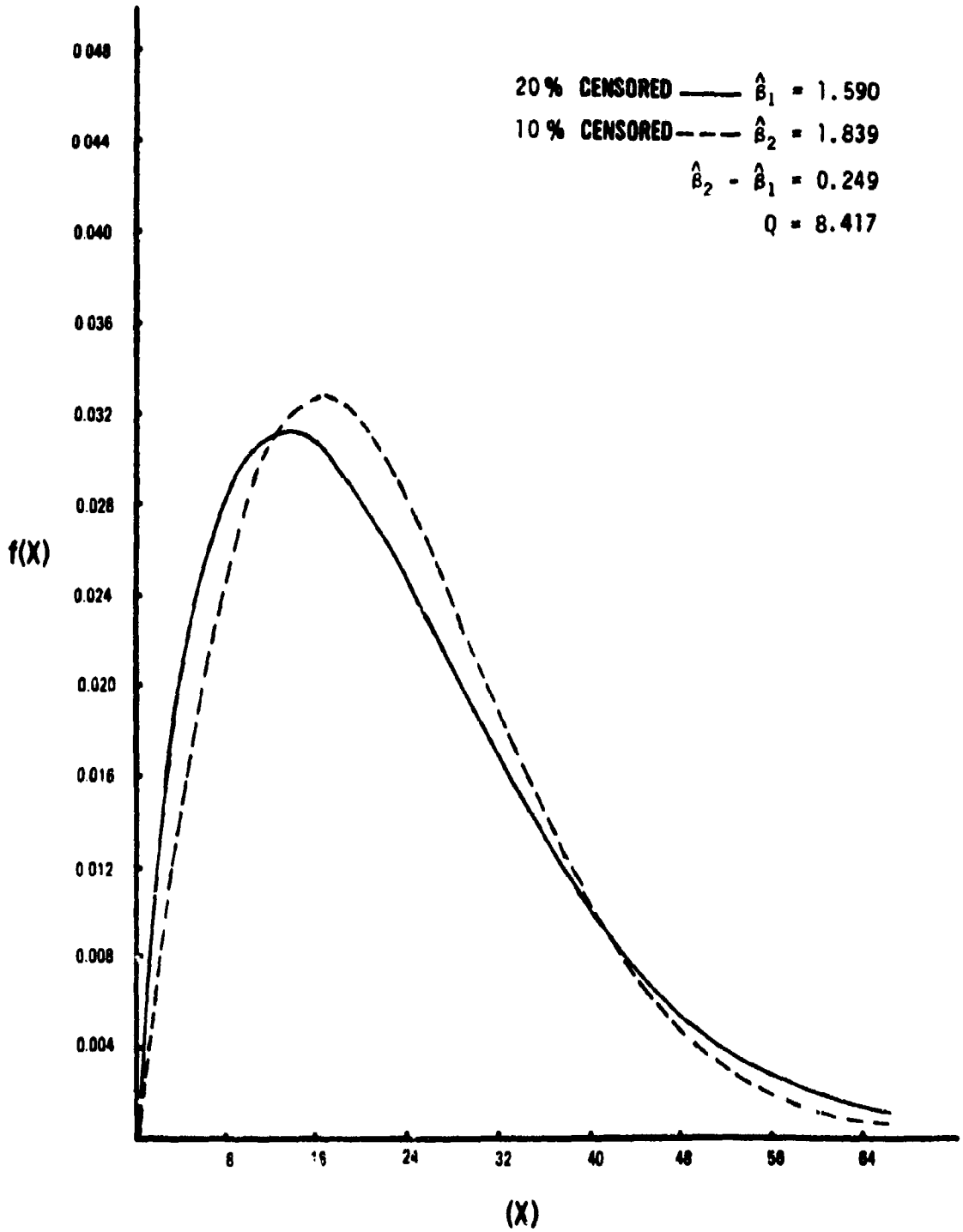


FIGURE A-2, N = 500 and Beta Between 1.5 and 2.0

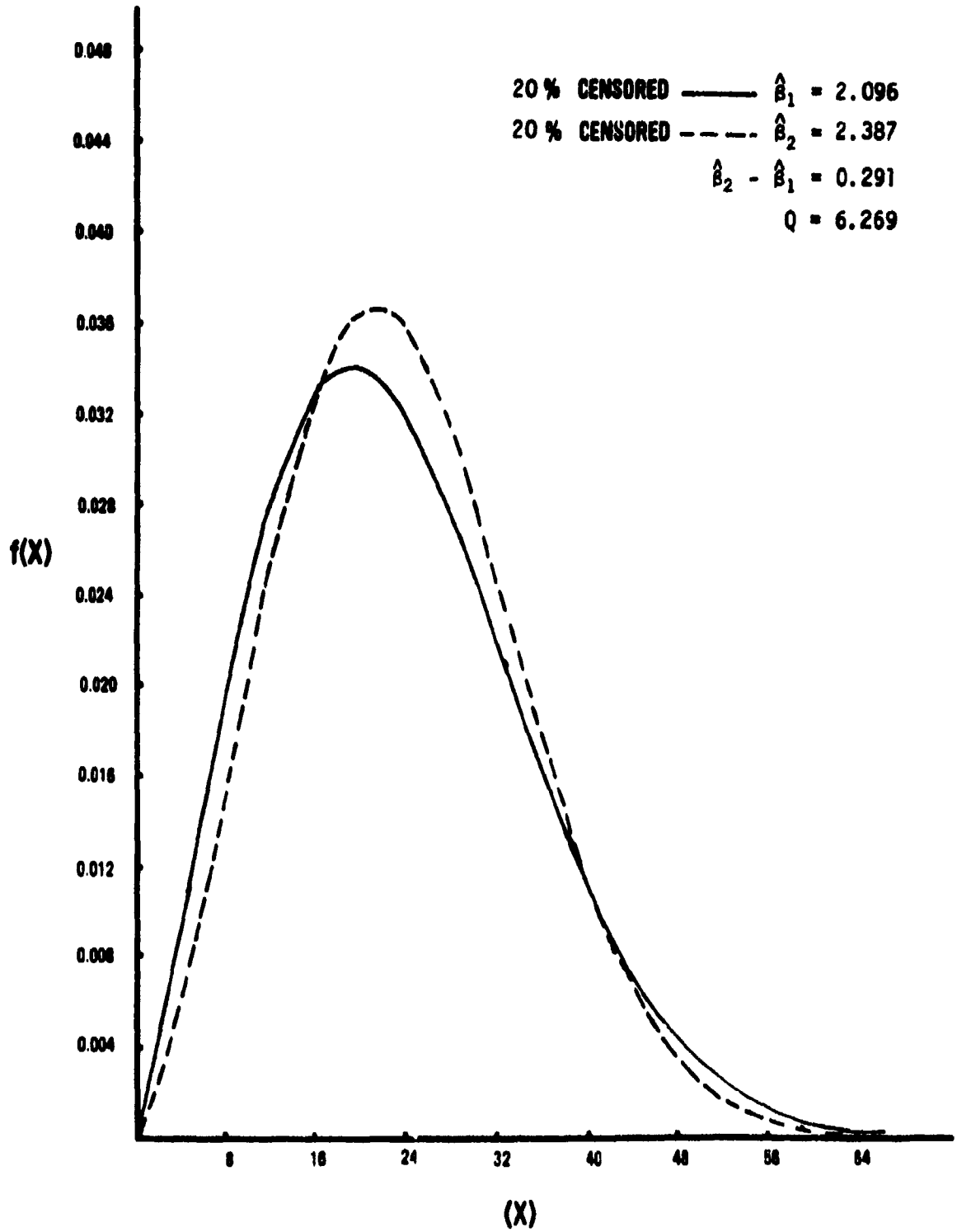


FIGURE A-3, N = 500 and Beta Between 2.0 and 2.5

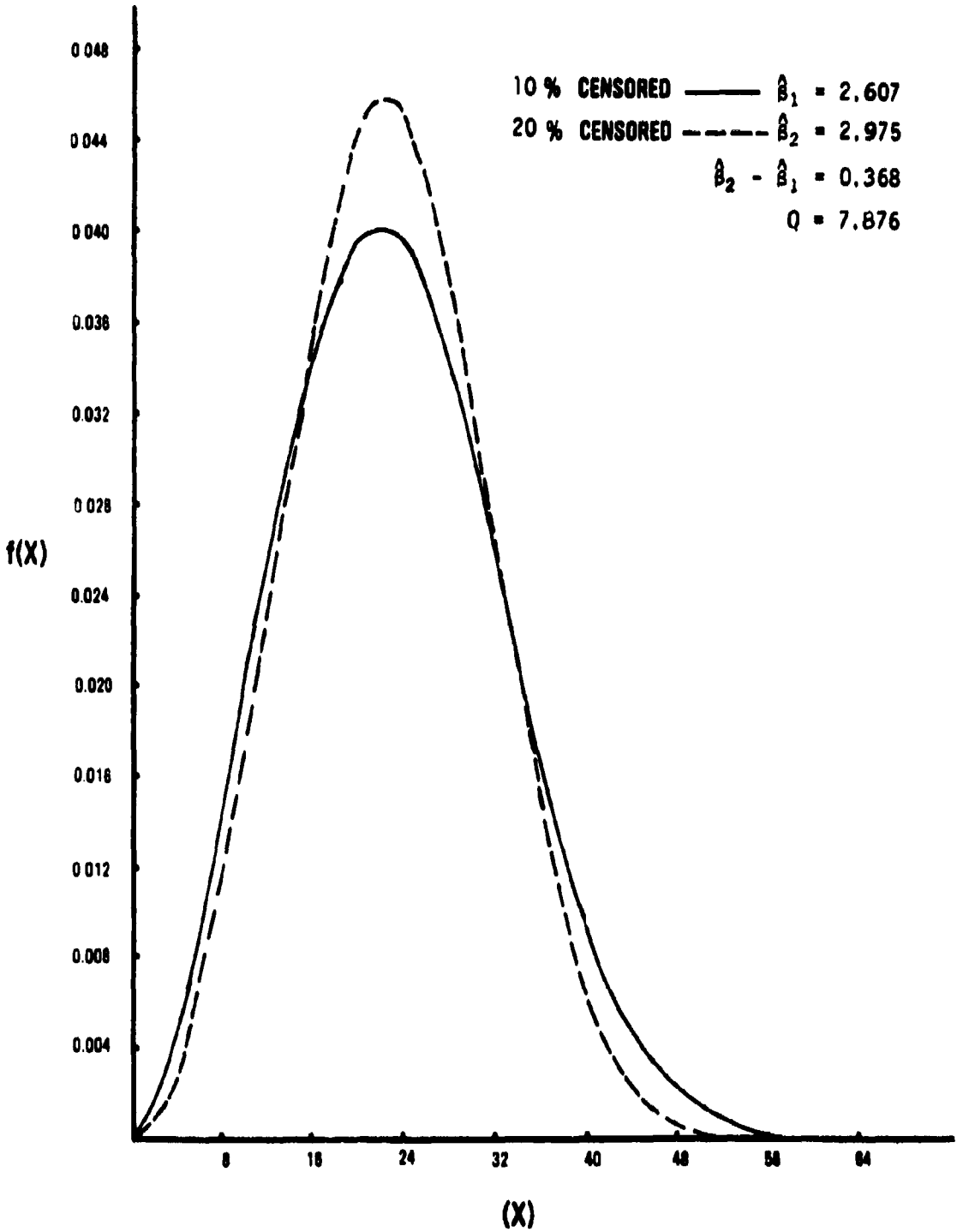


FIGURE A-4, N = 500 and Beta Between 2.5 and 3.0

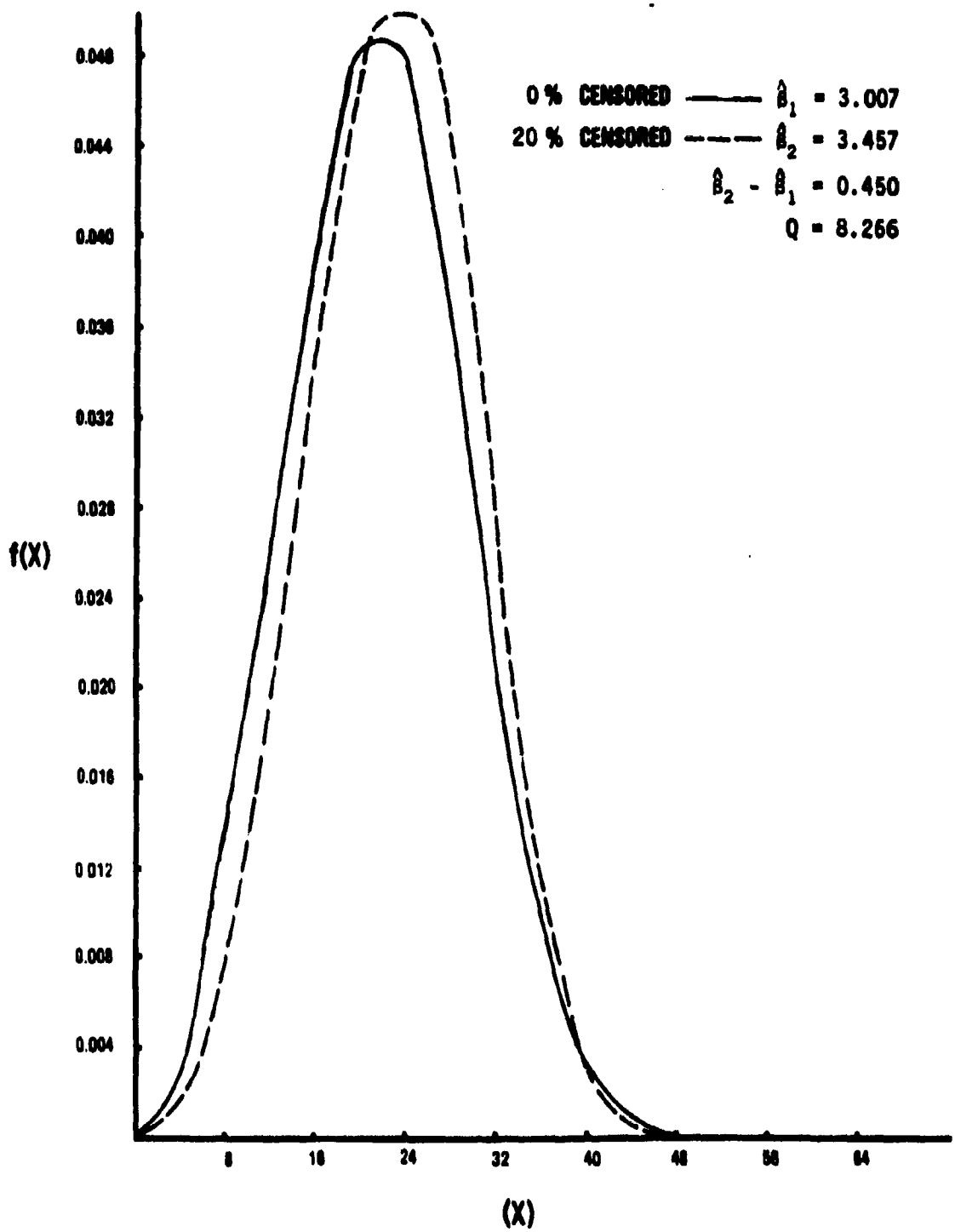


FIGURE A-5, N = 500 and Beta Between 3.0 and 3.5

APPENDIX B

SAMPLE SIZE OF 250

TABLE B-1, N = 250 and Shape Parameter Equal 1.0 - 1.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	23.762	1.0	1.020	23.573	534.632
2	10	25	26.378	1.0	1.127	25.258	504.394
3	20	25	29.632	1.0	1.086	28.718	700.942
4	0	25	25.476	1.1	1.134	24.347	462.878
5	10	25	29.179	1.1	1.170	27.631	560.984
6	20	25	26.487	1.1	1.158	25.158	474.698
7	0	25	24.300	1.2	1.286	22.494	310.885
8	10	25	27.358	1.2	1.339	25.124	359.439
9	20	25	28.148	1.2	1.256	26.186	440.023
10	0	25	25.737	1.3	1.393	23.476	291.445
11	10	25	26.655	1.3	1.390	24.322	314.016
12	20	25	23.056	1.3	1.258	21.443	294.347
13	0	25	27.990	1.4	1.413	25.474	334.136
14	10	25	24.890	1.4	1.392	22.706	272.956
15	20	25	27.801	1.4	1.384	25.384	344.597
16	0	25	22.362	1.5	1.482	20.218	192.794
17	10	25	26.176	1.5	1.490	23.651	261.168
18	20	25	25.832	1.5	1.522	23.279	243.071

TABLE B-2, N = 250 and Shape Parameter Equal 1.5 - 2.0

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.720	1.5	1.693	22.062	179.775
2	10	25	26.497	1.5	1.527	23.870	254.103
3	20	25	26.613	1.5	1.514	23.998	260.989
4	0	25	26.040	1.6	1.774	23.174	182.153
5	10	25	26.950	1.6	1.776	23.982	194.670
6	20	25	23.984	1.6	1.603	21.500	188.652
7	0	25	23.940	1.7	1.634	21.424	180.782
8	10	25	26.037	1.7	1.752	23.187	186.544
9	20	25	27.766	1.7	1.658	24.819	236.237
10	0	25	25.325	1.8	1.818	22.511	164.492
11	10	25	25.188	1.8	1.742	22.440	176.679
12	20	25	26.896	1.8	1.789	23.926	191.213
13	0	25	26.207	1.9	2.020	23.222	144.716
14	10	25	26.284	1.9	1.881	23.331	166.163
15	20	25	26.581	1.9	1.836	23.617	177.900
16	0	25	26.000	2.0	2.015	23.039	143.155
17	10	25	25.068	2.0	1.876	22.254	151.880
18	20	25	26.476	2.0	2.021	23.460	147.584

TABLE B-3, N = 250 and Shape Parameter Equal 2.0 - 2.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.476	2.0	2.104	21.678	117.223
2	10	25	26.505	2.0	2.147	23.473	132.560
3	20	25	27.163	2.0	2.020	24.068	155.524
4	0	25	25.263	2.1	2.184	22.373	116.729
5	10	25	26.646	2.1	2.071	23.602	142.957
6	20	25	26.016	2.1	2.276	23.045	115.090
7	0	25	22.595	2.2	2.079	20.014	102.081
8	10	25	25.165	2.2	2.082	22.290	126.272
9	20	25	27.227	2.2	2.252	24.116	128.451
10	0	25	25.664	2.3	2.216	22.730	117.367
11	10	25	25.207	2.3	2.556	22.378	88.159
12	20	25	26.006	2.3	2.419	23.058	103.245
13	0	25	24.950	2.4	2.292	22.103	104.538
14	10	25	26.426	2.4	2.447	23.435	104.468
15	20	25	25.957	2.4	2.374	23.006	106.331
16	0	25	25.662	2.5	2.490	22.767	95.563
17	10	25	26.282	2.5	2.416	23.302	105.697
18	20	25	26.208	2.5	2.417	23.236	105.035

TABLE B-4, N = 250 and Shape Parameter Equal 2.5 - 3.0

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.262	2.5	2.535	21.534	82.838
2	10	25	25.532	2.5	2.565	22.668	89.858
3	20	25	26.152	2.5	2.437	23.190	103.054
4	0	25	25.741	2.6	2.465	22.831	97.911
5	10	25	24.870	2.6	2.569	22.082	85.021
6	20	25	26.853	2.6	2.774	23.903	86.829
7	0	25	25.724	2.7	2.653	22.862	86.062
8	10	25	25.816	2.7	2.736	22.968	82.166
9	20	25	26.075	2.7	2.904	23.252	75.714
10	0	25	25.143	2.8	2.687	22.356	80.425
11	10	25	26.269	2.8	2.748	23.375	84.475
12	20	25	26.334	2.8	2.914	23.486	76.778
13	0	25	24.546	2.9	2.704	21.830	75.828
14	10	25	26.268	2.9	3.026	23.466	71.612
15	20	25	27.227	2.9	2.913	24.283	82.091
16	0	25	25.083	3.0	2.901	22.367	70.178
17	10	25	25.996	3.0	3.100	23.248	67.300
18	20	25	26.109	3.0	3.176	23.376	65.154

TABLE B-5, N = 250 and Shape Parameter Equal 3.0 - 3.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	24.158	3.0	3.049	21.588	59.806
2	10	25	25.033	3.0	3.258	22.441	57.350
3	20	25	25.607	3.0	3.189	22.931	62.237
4	0	25	25.825	3.1	3.228	23.140	62.027
5	10	25	24.663	3.1	3.143	22.070	59.200
6	20	25	25.610	3.1	3.389	23.004	56.148
7	0	25	25.665	3.2	3.032	22.929	68.138
8	10	25	26.050	3.2	3.543	23.454	53.880
9	20	25	25.577	3.2	3.443	22.993	54.520
10	0	25	23.590	3.3	3.026	21.074	57.762
11	10	25	25.560	3.3	3.237	22.906	60.457
12	20	25	25.075	3.3	3.614	22.600	48.293
13	0	25	24.413	3.4	3.395	21.931	50.869
14	10	25	25.768	3.4	3.462	23.171	54.834
15	20	25	24.639	3.4	3.685	22.231	45.098
16	0	25	25.200	3.5	3.552	22.692	50.221
17	10	25	25.484	3.5	3.589	22.960	50.455
18	20	25	25.830	3.5	3.481	23.234	54.600

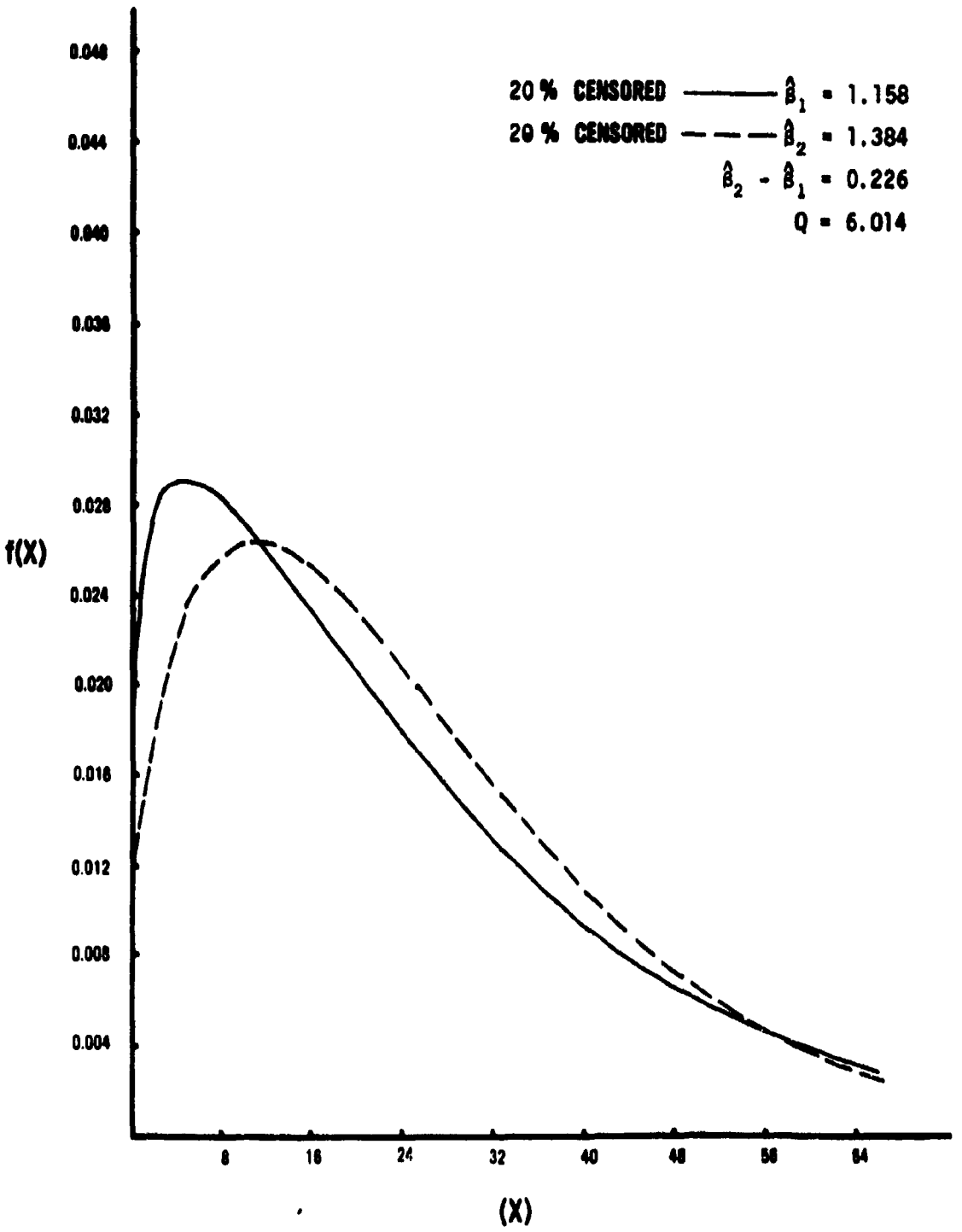


FIGURE B-1, N = 250 and Beta Between 1.0 and 1.5

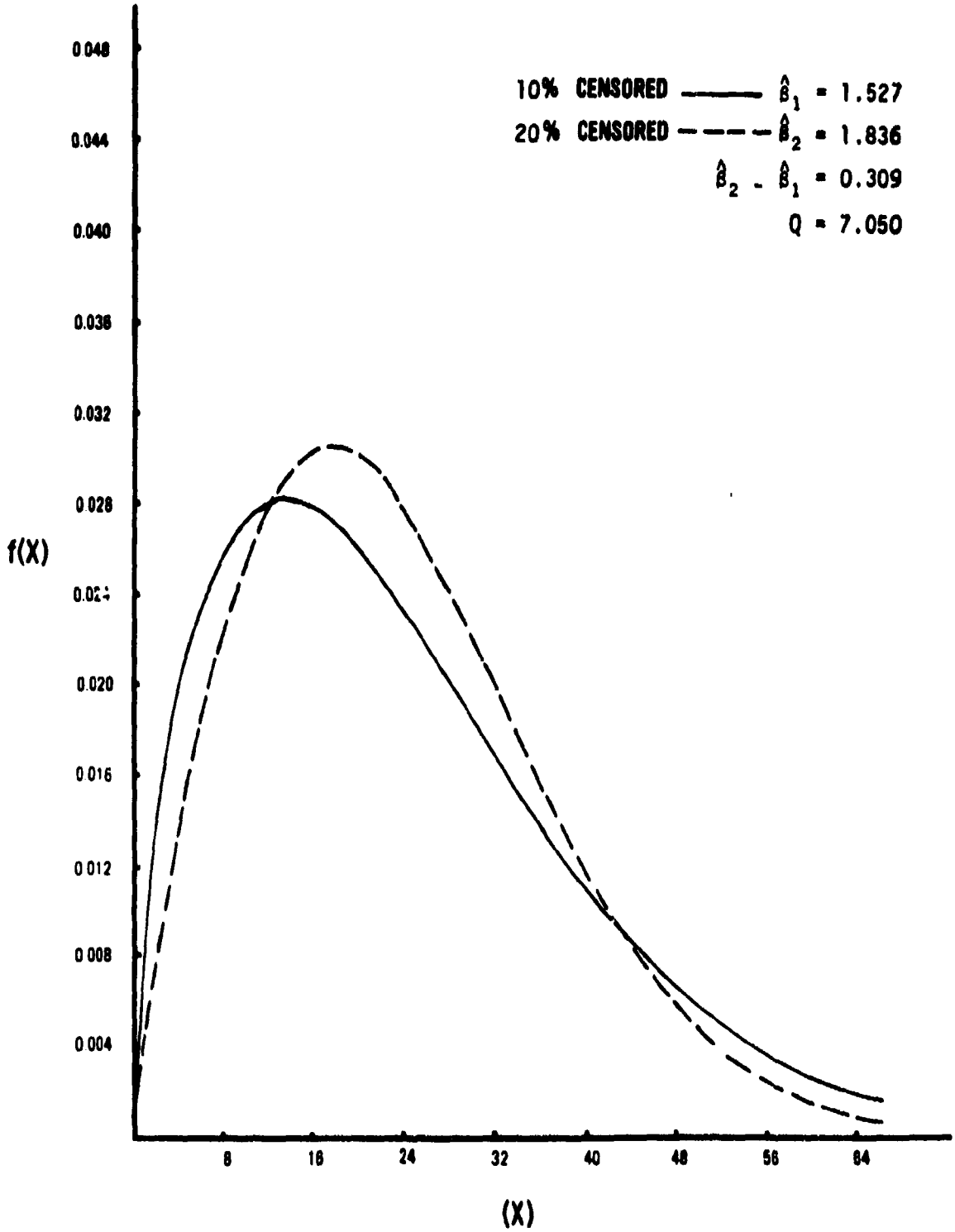


FIGURE B-2, N = 250 and Beta Between 1.5 and 2.0

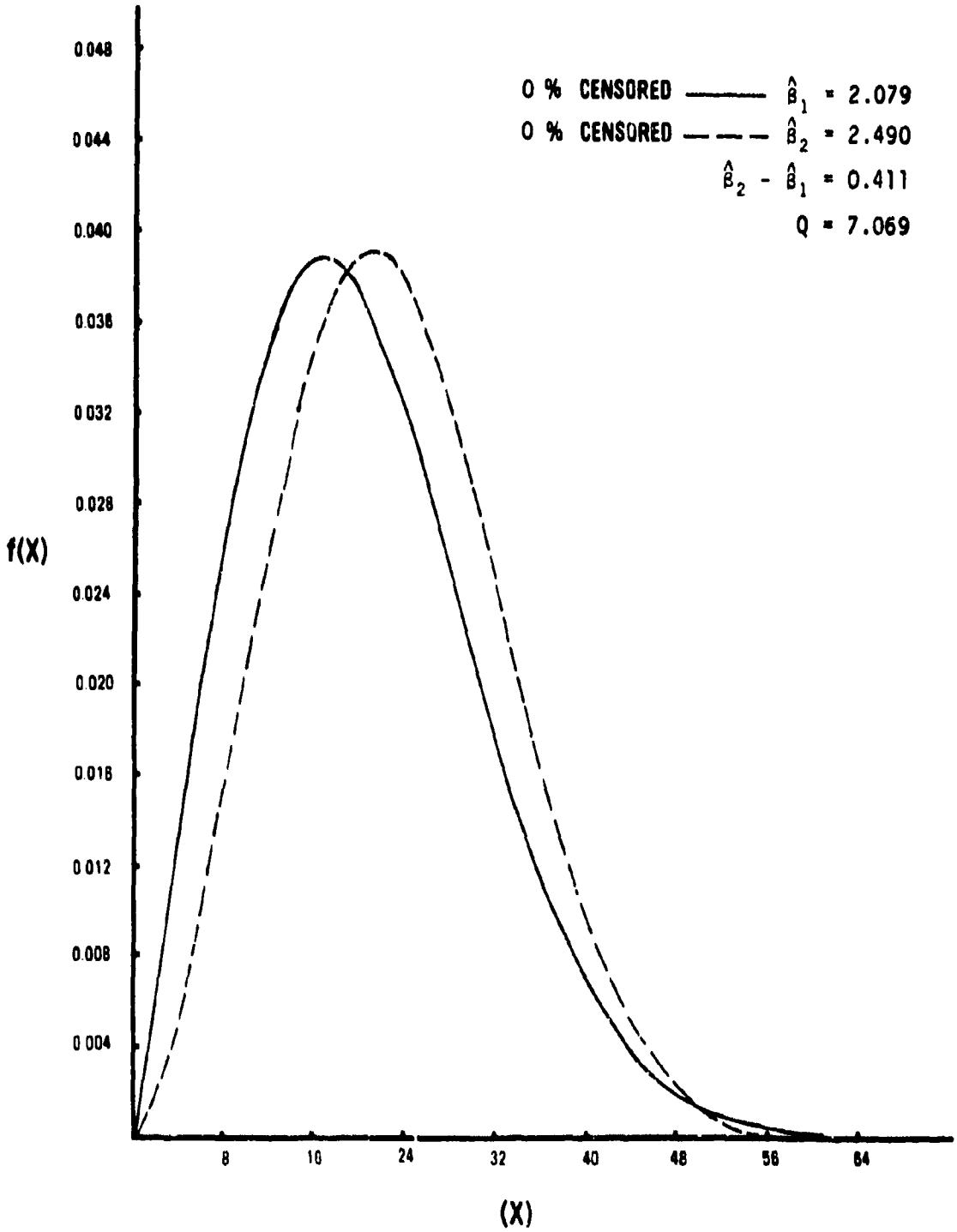


FIGURE B-3, N = 250 and Beta Between 2.0 and 2.5

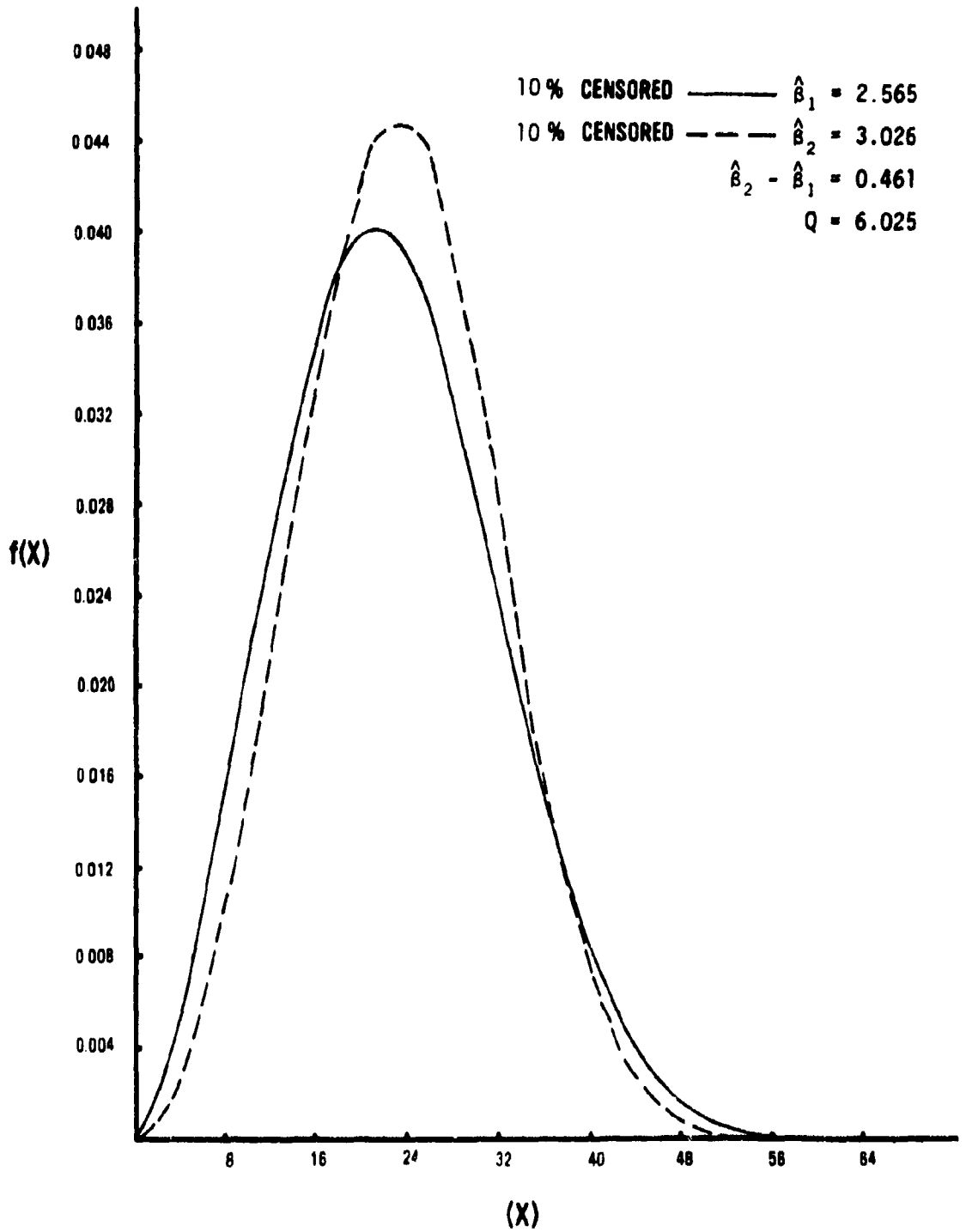


FIGURE B-4, N = 250 and Beta Between 2.5 and 3.0

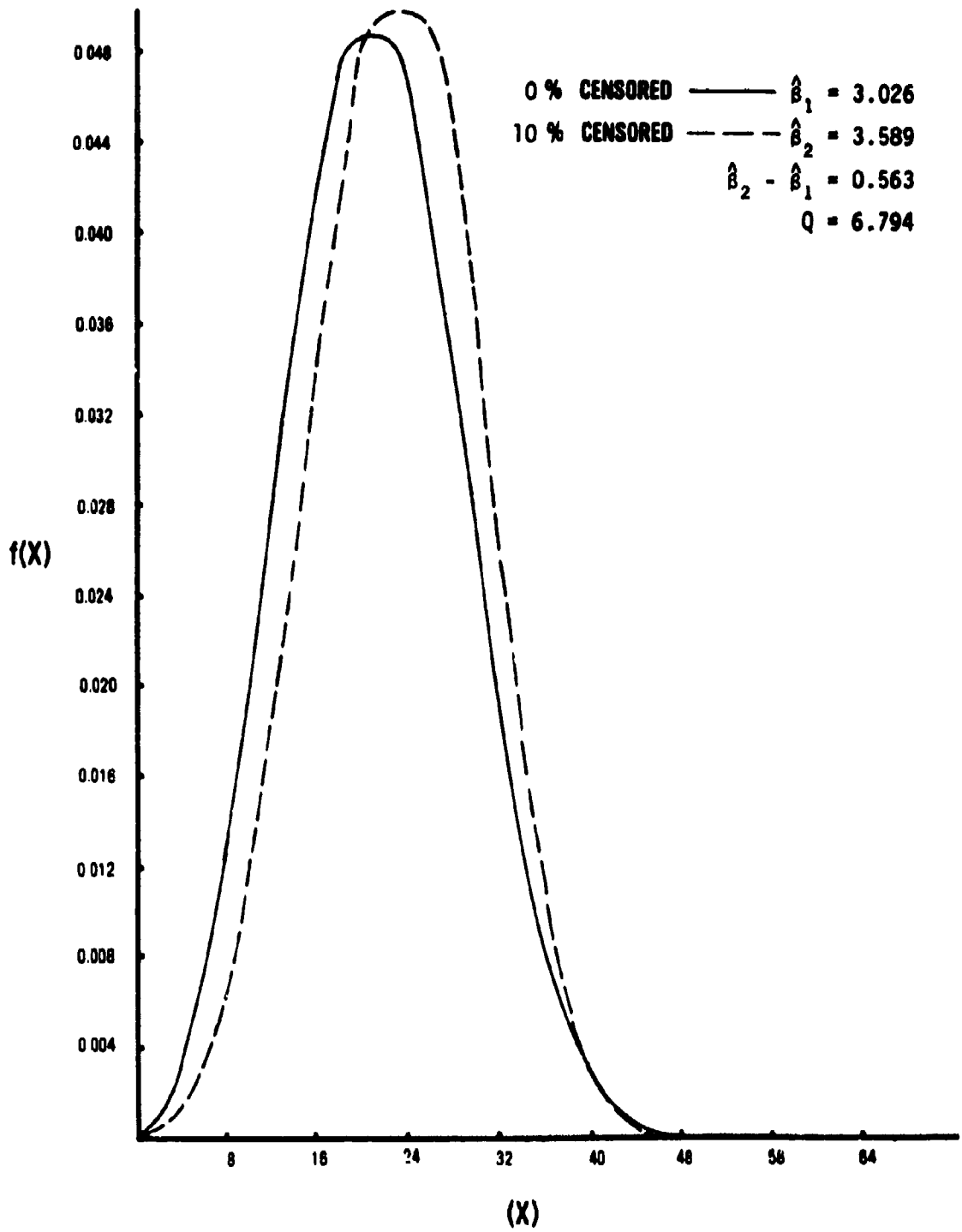


FIGURE B-5, N = 250 and Beta Between 3.0 and 3.5

APPENDIX C

SAMPLE SIZE OF 100

TABLE C-1, N = 100 and Shape Parameter Equal 1.0 - 1.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	23.178	1.0	1.099	22.373	415.600
2	10	25	28.484	1.0	0.997	28.516	817.443
3	20	25	30.436	1.0	1.082	29.537	747.180
4	0	25	23.569	1.1	1.030	23.286	511.188
5	10	25	27.753	1.1	1.171	26.276	506.755
6	20	25	26.699	1.1	1.137	25.498	505.449
7	0	25	29.079	1.2	1.164	27.580	565.090
8	10	25	24.477	1.2	1 348	22.451	283.365
9	20	25	23.936	1.2	1.265	22.235	313.330
10	0	25	21.880	1.3	1.306	20.189	243.187
11	10	25	25.352	1.3	1.295	23.432	332.685
12	20	25	30.068	1.3	1.391	27.433	399.079
13	0	25	25.828	1.4	1.352	23.679	313.640
14	10	25	25.178	1.4	1.468	22.791	249.206
15	20	25	25.076	1.4	1.356	22.978	293.746
16	0	25	24.874	1.5	1.642	22.251	193.331
17	10	25	25.526	1.5	1.381	23.315	291.978
18	20	25	27.881	1.5	1.481	25.209	299.958

TABLE C-2, N = 100 and Shape Parameter Equal 1.5 - 2.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	25.677	1.5	1.970	22.763	146.523
2	10	25	25.802	1.5	1.470	23.353	261.083
3	20	25	28.284	1.5	1.739	25.199	223.403
4	0	25	24.982	1.7	1.761	22.242	170.148
5	10	25	26.609	1.7	2.069	23.570	142.841
6	20	25	30.676	1.7	1.856	27.243	231.938
7	0	25	25.189	1.9	1.884	22.358	152.180
8	10	25	26.307	1.9	2.115	23.299	134.137
9	20	25	22.409	1.9	1.763	19.949	136.580
10	0	25	23.490	2.1	1.890	20.848	131.500
11	10	25	26.654	2.1	2.090	23.608	140.722
12	20	25	23.775	2.1	2.130	21.056	108.133
13	0	25	25.297	2.3	2.182	22.403	117.322
14	10	25	24.436	2.3	2.307	21.649	99.114
15	20	25	24.798	2.3	2.561	22.017	85.016
16	0	25	24.991	2.5	2.298	22.140	104.409
17	10	25	25.673	2.5	2.596	22.802	88.972
18	20	25	27.186	2.5	2.636	24.157	97.178

TABLE C-3, N = 100 and Shape Parameter Equal 2.5 - 3.5

Sample Number	Percent Censored	α	$\hat{\alpha}$	β	$\hat{\beta}$	E(x)	V(x)
1	0	25	25.672	2.5	2.806	22.773	77.236
2	10	25	27.798	2.5	2.673	24.712	99.212
3	20	25	25.351	2.5	2.774	22.566	77.412
4	0	25	25.412	2.7	3.021	22.699	67.207
5	10	25	23.864	2.7	2.823	21.256	66.566
6	20	25	25.472	2.7	2.836	22.693	75.244
7	0	25	22.876	2.9	2.716	20.347	65.350
8	10	25	24.891	2.9	3.496	22.394	50.329
9	20	25	25.791	2.9	2.498	22.883	96.030
10	0	25	24.755	3.1	2.938	22.086	66.896
11	10	25	25.776	3.1	3.119	23.059	65.494
12	20	25	26.556	3.1	3.371	23.847	60.940
13	0	25	27.079	3.3	3.637	24.414	55.696
14	10	25	26.633	3.3	3.356	23.910	61.756
15	20	25	25.906	3.3	3.346	23.254	58.731
16	0	25	25.802	3.5	3.421	23.188	56.100
17	10	25	25.686	3.5	3.830	23.225	45.922
18	20	25	25.569	3.5	3.434	22.983	54.747

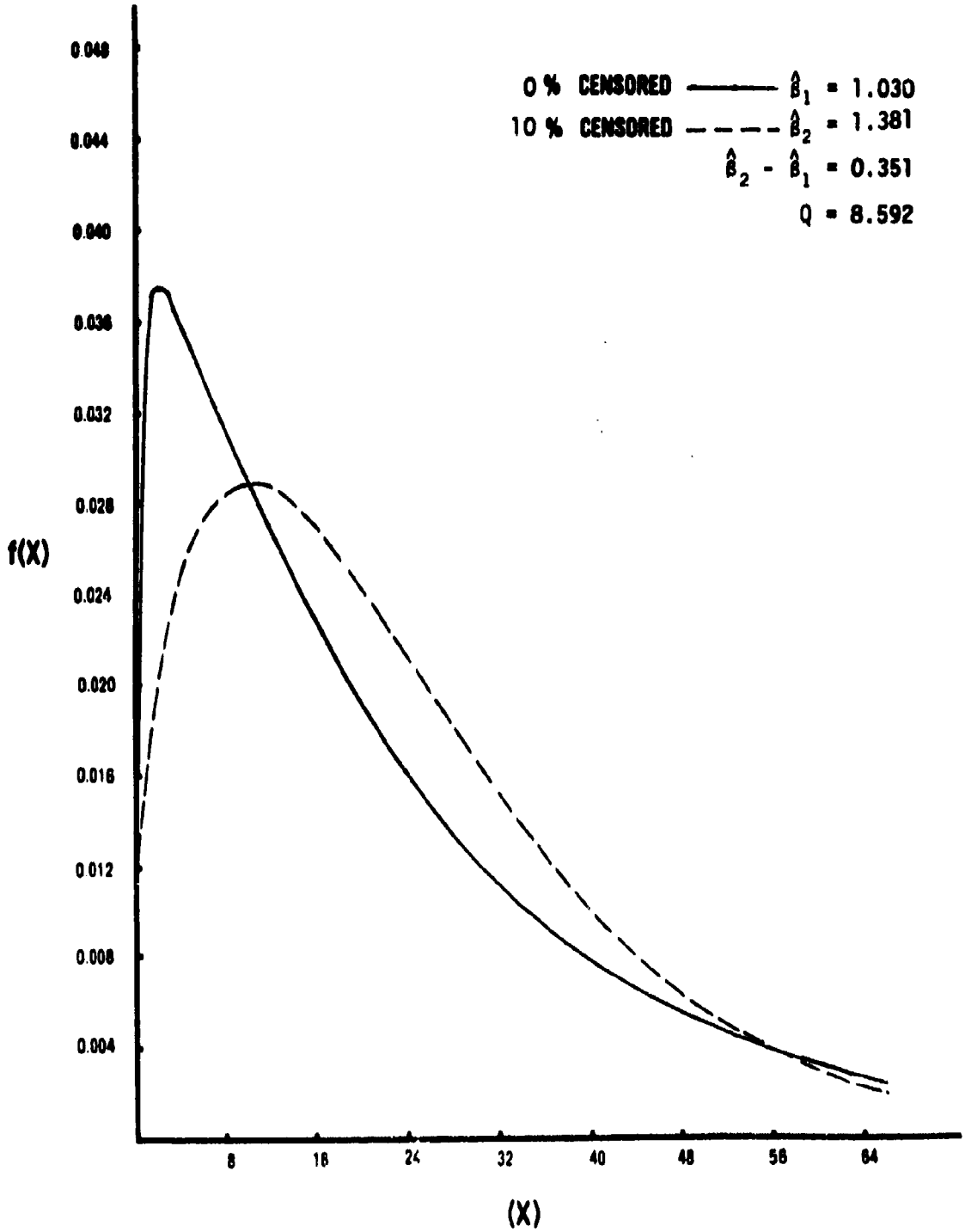


FIGURE C-1, N = 100 and Beta Between 1.0 and 1.5

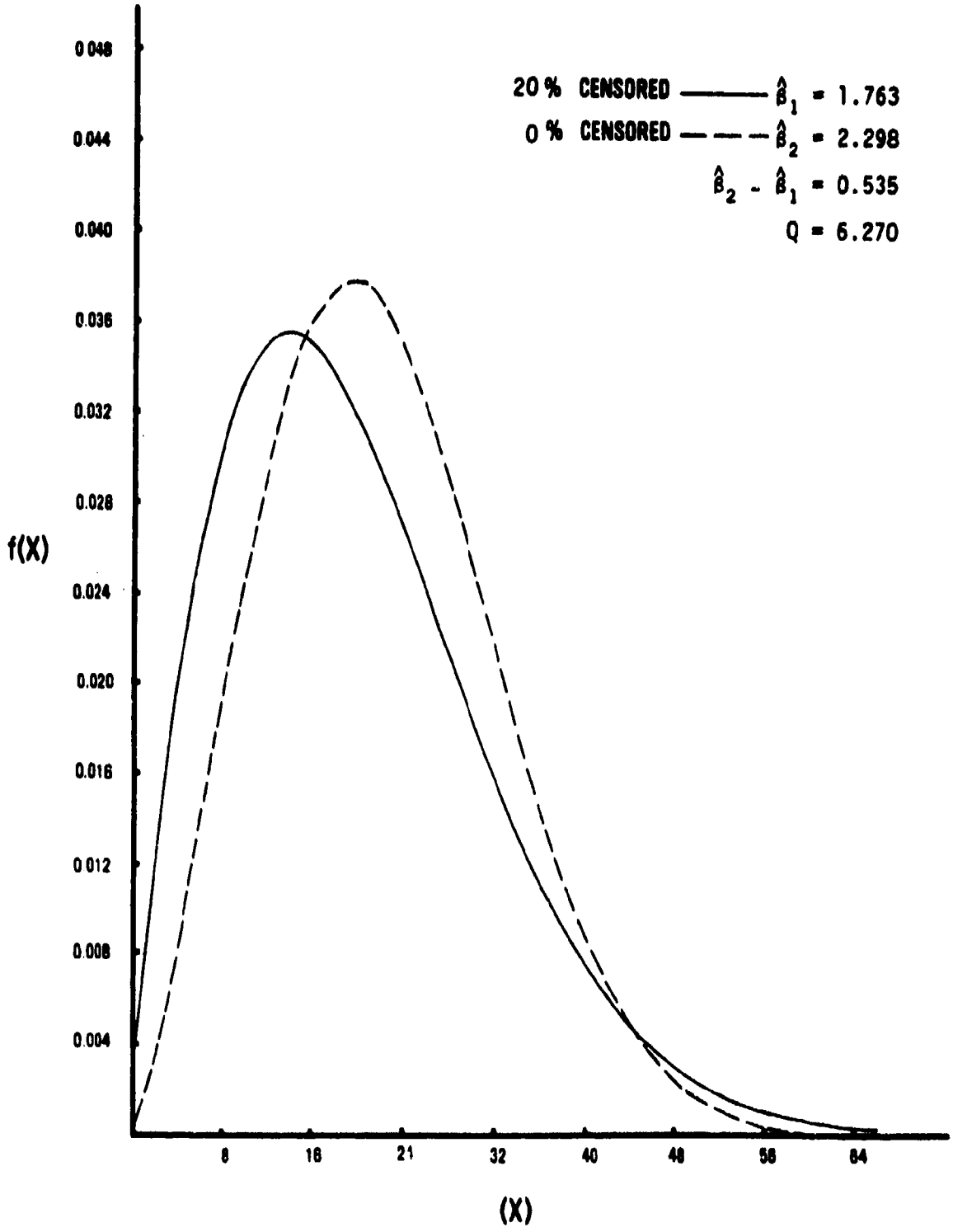


FIGURE C-2, N = 100 and Beta Between 1.5 and 2.5

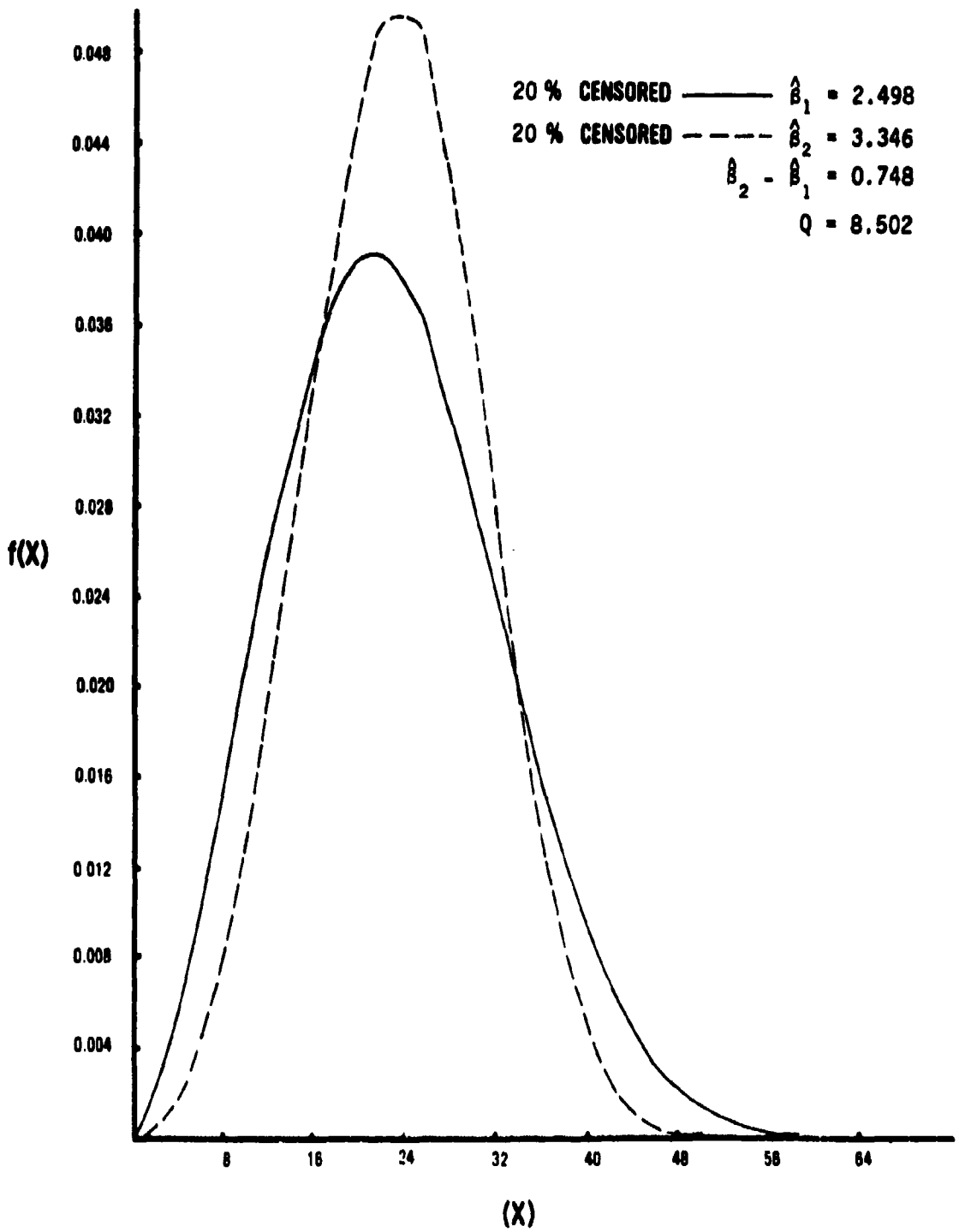


FIGURE C-3, N = 100 and Beta Between 2.5 and 3.5

ON THE ROBUSTNESS OF THE EXPONENTIAL DISTRIBUTION

George C. Canavos
Virginia Commonwealth University
Richmond, Virginia

ABSTRACT. This paper examines the robustness of the exponential time-to-failure distribution when this probability law is compared against some logical alternatives such as the Weibull and gamma distributions relative to estimation procedures involving the scale parameter.

1. INTRODUCTION. Since the pioneering work on life testing and reliability estimation during the early 1950's - see, for example, [1] and [2] - the exponential distribution has been the most widely assumed probability law in describing times to failure of many types of components and systems. There is little doubt that this distribution has played a key role in both theory and application over the past twenty or so years. Surely, therefore, it is of continued interest to query, "What if the assumption of the exponential probability law does not hold? To what extent then will such an occurrence affect subsequent inferences and estimation procedures derived as a result of and depending on this assumption?"

A substantive study on the robustness of the exponential distribution is hereby attempted. Where possible, the treatment is analytic. Particular attention is given to the estimation of the scale parameter and the ramifications regarding the mean-squared error (MSE) of its estimate if the exponential assumption does not hold. The effect on the MSE is determined as a function of a situation in which the true sampling distribution of lifetimes is not the assumed exponential but rather is either a Weibull or a gamma. By following such a procedure, the degree of robustness of the exponential distribution is measured and quantified.

2. THEORETICAL DEVELOPMENT OF ROBUSTNESS. Let $x_1, x_2, \ldots, x_n$ denote the times-to-failure of n like items. Assume that these lifetimes follow the exponential distribution with probability density function (pdf)

$$f(x;\theta) = \frac{1}{\theta} \exp(-\theta x) \ , \ x > 0 \qquad (1)$$

where interest is on the estimation of the parameter θ. By appealing to the likelihood function

$$\ell(x_1, x_2, \ldots, x_n; \theta) = \frac{1}{\theta^n} \exp\left(\frac{-\sum_{i=1}^{n} x_i}{\theta}\right)$$

one can easily determine the minimum variance unbiased estimate (MVUE) of θ to be

$$\hat{\theta} = \sum_{i=1}^{n} \frac{x_i}{n} . \tag{2}$$

Suppose, however, that in reality the lifetimes $x_1, x_2, \ldots, x_n$ are realizations of a Weibull random variable with pdf

$$h(x; \theta, \alpha) = \frac{\alpha}{\theta} x^{\alpha-1} \exp\left(-\frac{1}{\theta} x^{\alpha}\right) , \quad x > 0 \tag{3}$$

where α is a shape parameter. Again it is a rather straightforward procedure to determine that the MVUE of θ in this case is

$$\hat{\theta} = \sum_{i=1}^{n} \frac{x_i^{\alpha}}{n} . \tag{4}$$

Thus, if in reality the lifetimes follow the Weibull, the optimal efficiency (in the classical sense) for estimating θ is provided by the MSE of the MVUE estimator (4) which reduces to

$$MSE(\hat{\theta})_W = \frac{\theta^2}{n}. \tag{5}$$

Since the exponential distribution was assumed to accurately represent the lifetimes $x_1, x_2, \ldots, x_n$, however, the estimate of θ is determined by (2). Thus, what effect would the fact that the lifetimes follow the Weibull as opposed to the exponential have on the MSE of the estimator given by (2)? That is, if in reality $x_1, x_2, \ldots, x_n$ follow the Weibull with pdf given by (3), then for equation (2)

$$MSE(\hat{\theta}) = var(\hat{\theta}) + E\{\theta - E(\hat{\theta})\}^2$$

where

$$E(\hat{\theta}) = E\left(\sum_{i=1}^{n} \frac{x_i}{n}\right)$$
$$= \theta^{1/\alpha} \Gamma\left(1 + \frac{1}{\alpha}\right)$$

and

$$\text{var}(\hat{\theta}) = \text{var}\left(\sum_{i=1}^{n} \frac{x_i}{n}\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \text{var}\,(x_i)$$

$$= \theta^{2/\alpha} \{\Gamma(1 + \tfrac{2}{\alpha}) - \Gamma^2(1 + \tfrac{1}{\alpha})\}$$

Hence after some algebraic manipulation, the MSE with respect to the indicated perturbation is expressed by

$$\text{MSE}(\hat{\theta})_{E|W} = \frac{\theta^{2/\alpha}\{\Gamma(1 + \frac{2}{\alpha}) + (n - 1)\Gamma^2(1 + \frac{1}{\alpha})\} - 2n\theta^{1+\frac{1}{\alpha}}\Gamma(1 + \frac{1}{\alpha}) + n\theta^2}{n} \tag{6}$$

where the notation "E|W" indicates assumed exponential but in reality Weibull sampling. A comparison between equations (5) and (6) provides a measure of robustness relative to MSE in the assumption of the exponential distribution when estimating the scale parameter θ. Numerical results are given in the next section.

Analogous to the previous discussion, consider now the gamma distribution. As before, assume the lifetimes $x_1, x_2, \ldots, x_n$ follow the exponential with pdf given by (1). What are the consequences relative to the MSE of (2) if in fact the more appropriate probability law is the gamma with pdf

$$g(x;\theta,\alpha) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} \exp\left(-\frac{x}{\theta}\right), \quad x > 0 . \tag{7}$$

First, with respect to (7), it is easy to show that the MVUE of θ is

$$\hat{\theta} = \sum_{i=1}^{n} \frac{x_i}{\alpha n} \tag{8}$$

while

$$\text{MSE}(\hat{\theta})_G = \frac{\theta^2}{\alpha n} \tag{9}$$

Then to determine the MSE of (2), consider

$$E(\hat{\theta}) = E\left(\sum_{i=1}^{n} \frac{x_i}{n}\right)$$

$$= \alpha\theta$$

while

$$\text{var}(\hat{\theta}) = \text{var}\left(\sum_{i=1}^{n} \frac{x_i}{n}\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \text{var}\,(x_i)$$

$$= \frac{\alpha\theta^2}{n} .$$

Thus, the perturbed MSE of (2) reduces to

$$\text{MSE}(\hat{\theta})_{E|G} = \frac{\theta^2\{\alpha + n(1 - \alpha)^2\}}{n} \qquad (10)$$

As before, the comparison between equations (9) and (10) should reveal the degree of robustness of the exponential distribution as measured by the MSE of the scale parameter θ.

3. NUMERICAL RESULTS. To evaluate the robustness of the exponential with regard to the estimation of the scale parameter when the true sampling distribution is the Weibull, the ratio of equation (6) to equation (5) is formed. The notion here is that since in reality the lifetimes follow the Weibull time-to-failure probability law, then the best efficiency of the MVUE of θ is provided by (5). Thus the "perturbed" MSE given by (6) should be compared to (5). Table 1 contains this ratio computed for several values of θ, α and the sample size n.

By a similar argument, the ratio of equation (10) to equation (9) is formed to quantify the robustness of the exponential relative to the gamma distribution. However in this case, the ratio is the simple expression given by

$$R = \frac{\theta^2\{\alpha + n(1 - \alpha)^2\}/n}{\theta^2/n\alpha}$$

$$= \alpha\{\alpha + n(1 - \alpha)^2\}$$

which is seen to be independent of the value of θ. For various values of α and n, this ratio is given in Table 2.

4. CONCLUDING REMARKS. Based on the results contained herein, it is apparent that relative to the estimation of the scale parameter, the exponential distribution is extremely sensitive if in reality the Weibull is the sampling distribution and

Table 1

Ratio of $MSE(\hat{\theta})_{E|W}$ to $MSE(\hat{\theta})_{W}$

n = 5

θ \ α	0.8	0.9	1.10	1.20	1.50	2.00	2.50
5	6.97	2.29	0.71	0.75	1.24	1.86	2.21
10	11.60	2.93	0.74	0.93	1.77	2.61	3.03
15	15.46	3.39	0.77	1.05	2.07	2.99	3.41
20	18.87	3.76	0.80	1.15	2.27	3.23	3.64
25	21.96	4.08	0.82	1.22	2.43	3.39	3.80
30	24.81	4.35	0.84	1.29	2.56	3.52	3.92
35	27.48	4.60	0.86	1.34	2.66	3.62	4.01
40	30.00	4.83	0.87	1.39	2.74	3.70	4.08
45	32.39	5.03	0.89	1.43	2.81	3.77	4.14
50	34.68	5.22	0.90	1.47	2.88	3.83	4.19

n = 10

5	9.38	2.62	0.85	1.15	2.36	3.69	4.40
10	16.75	3.58	0.98	1.58	3.46	5.20	6.05
15	23.02	4.28	1.07	1.86	4.08	5.96	6.82
20	28.61	4.86	1.15	2.07	4.51	6.44	7.28
25	33.71	5.35	1.21	2.24	4.82	6.78	7.60
30	38.45	5.79	1.27	2.37	5.07	7.03	7.83
35	42.89	6.18	1.31	2.49	5.28	7.23	8.01
40	47.10	6.54	1.36	2.60	5.45	7.40	8.16
45	51.11	6.87	1.39	2.69	5.60	7.54	8.28
50	54.94	7.18	1.43	2.77	5.73	7.65	8.38

n = 20

5	14.20	3.29	1.13	1.94	4.58	7.33	8.79
10	27.05	4.86	1.45	2.87	6.83	10.38	12.09
15	38.14	6.06	1.68	3.47	8.10	11.91	13.63
20	48.10	7.04	1.85	3.91	8.96	12.87	14.55
25	57.23	7.90	2.00	4.26	9.60	13.55	15.19
30	65.73	8.65	2.12	4.55	10.10	14.06	15.66
35	73.72	9.34	2.22	4.80	10.52	14.46	16.02
40	81.30	9.96	2.32	5.01	10.87	14.79	16.31
45	88.53	10.54	2.40	5.20	11.16	15.07	16.55
50	95.45	11.09	2.48	5.37	11.43	15.31	16.75

the shape parameter is less than one. However, there is a modest range of the shape parameter - say (1.0,1.3) - for which there is substantial robustness on the part of the exponential distribution. Moreover, the robustness is more apparent for smaller smaple sizes and smaller values of θ.

For the case involving the gamma distribution, to some extent the opposite appears to hold. That is, for values of the shape parameter that are less than unity, considerable robustness is apparent especially for small sample sizes with only a modest amount present in the neighborhood but on the positive side of one.

Table 2

Ratio of $MSE(\hat{\theta})_{E|G}$ to $MSE(\hat{\theta})_{G}$

α \ n	5	10	20
0.50	0.875	1.50	2.75
0.60	0.840	1.32	2.28
0.70	0.805	1.12	1.75
0.80	0.800	0.96	1.28
0.90	0.855	0.90	0.99
0.95	0.914	0.93	0.95
1.00	1.000	1.00	1.00
1.10	1.265	1.32	1.43
1.20	1.680	1.92	2.40
1.40	3.080	4.20	6.44
1.60	5.440	8.32	14.08

REFERENCES.

1. Epstein, B., and M. Sobel, "Life Testing," Jour. Amer. Statist. Assoc., 48, 486-502 (1953).

2. Epstein, B., and M. Sobel, "Some Theorems Relevant to Life Testing from an Exponential Distribution," Ann. Math. Statist., 25, 373-381 (1954).

RANDOM INTERVAL RELIABILITY

Gerald R. Andersen
Headquarters, U.S. Army Materiel
Development and Readiness Cmd.
5001 Eisenhower Ave.
Alexandria, Virginia

Abstract. Simple expressions are derived for interval reliability when, in addition to random life and repair times, the time of request for system availability and the duration of the mission occasioned by that request are random variables, rather than numerical constants. The results constitute a simple generalization of the interval reliability results noted in Barlow and Proschan [1].

The investigation was motivated by the desire to discourage the extensive misapplication of the result of [1] p. 82 in setting reliability values for large scale Army systems in pre-development requirements documents.

1. Introduction. Let Γ be a stochastic process whose value, $\Gamma(t)$, at a particular time $t \geq 0$, describes the operating state of some system at time, t. We will only consider systems with two states, up (operable/operating) or down (in repair). Specifically, we will say that the system is up at time t if $\Gamma(t)=1$ and down at time t if $\Gamma(t)=0$. We assume that $\Gamma(0)=1$ with probability one.

Starting at time t=0, let $X_1, Y_1, X_2, Y_2, \ldots$ denote the successive lengths of time that the process, Γ, spends in the up or down state, respectively.

Let

$$T_\nu = X_\nu + Y_\nu \quad , \quad \nu \geq 1, \tag{1.1}$$

$S_0=0$ and define S_n by setting

$$S_n = \sum_{\nu=0}^{n} T_\nu \tag{1.2}$$

Throughout most of this note each of the sequences $\{X_i\}$ and $\{Y_i\}$ will consist of independent and identically distributed (IID) r.v.'s. In this case, $\{S_n\}$ is the usual type of renewal process

used to study systems where the X_i's are the times to failure and the Y_j's are the times to replacement or to repair to-original-condition.

Associated with this renewal process, $\{S_n\}$, is the counting process N(t), where

$$N(t) = k \quad \text{and} \quad S_{N(t)} = S_k \tag{1.3}$$

if, and only if,

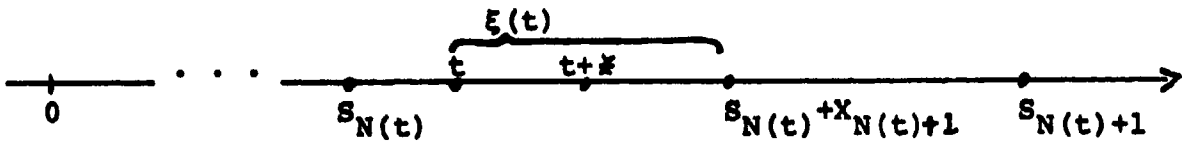
$$S_k \leq t < S_{k+1} \tag{1.4}$$

The "residual life" process, $\xi(t)$, defined by setting

$$\xi(t) = S_{N(t)} + X_{N(t)+1} - t \tag{1.5}$$

$(t \geq 0)$ is useful in investigating the probability that $\Gamma(t)=1$ during various intervals of time.

Since N(t) represents the number of times the process $\Gamma(t)$ returns to the up state during the interval (0,t), the event that $\xi(t) > x$ coincides with the event that the system is in the up state at time t and remains in that state for at least x units of time



In section 2 we will obtain exact and asymptotic expressions for the probability that $\xi(\tau)$ exceeds the quantity M when <u>both</u> τ and M are random variables. This probability, that the system is up throughout the interval $[\tau, \tau+M]$, is called interval reliability by Barlow and Proschan [1] p. 82, in the case where τ and M are non-random. It is interesting to note that many Army documents, including a guide on reliability techniques [10], apply the result in [1] but with the claims that either τ or M are random.

The mathematics required to make this extension from the well-known results in Barlow and Proschan, or Gnedenko [4], or Feller [11] is very simple, but in some ways the results are reasonably interesting. In spite of this, it is doubtful that one would announce the results of such a simple task if it were

not for the insane realism that some practitioners of reliability inject into reliability "requirements" as deduced from mathematical facts about residual life. This topic is expanded on in Example A of section 2.

In section 3, we note the well-known fact that an aysmptotic result of section 2 is the limit of a statistic which gives the percentage of time, during n renewals, that the system is up and remains up for a sufficient amount of time to support a mission of duration x. A result is then stated concerning the asymptotic normality of a similar statistic (one representing the percentage of up-time that the system is available for a mission of duration x).

Section 4 is an attempt to consider the interval reliability problem when successive system life and repair times are not identically distributed.

2.0 Residual life; independent and identically distributed case. Let the sequences $\{X_i\}$ and $\{Y_j\}$ of section 1 be sequences of independent and identically distributed positive random variables (r.v.'s) and assume also that $\{X_i\}$ and $\{Y_j\}$ are independent of each other. Thus, in this section, the X_i's have the usual interpretation of time to system failure and the Y_j's the time to replace or repair the system to a state which is as good as new.

We will denote the common distribution function (d.f.) of the X_i's by G, of the Y_j's by H and, where appropriate, use X to refer to one of the X_i's and Y to one of the Y_j's. Set F equal to the d.f. of $T = X+Y$. Let the positive r.v.'s τ and M of section 1 be independent of each other and of the sequences $\{X_i\}$ and $\{Y_j\}$. Denote the d.f.'s of τ and M by K and L, respectively. Although termed a positive r.v., M will be allowed to take the value zero with positive probability; especially, the case M=0 with probability one (a.s.). This allows "availability" as well as interval reliability statements to be included in the same expression. When M=0 (a.s.), the L=ε, where ε will denote the unit d.f.:

$$\varepsilon(y) = \begin{cases} 0, & \text{if } y < 0 \\ 1, & \text{if } y \geq 0 \end{cases} \qquad (2.1)$$

To avoid needless complications, we suppose that K(0)=0 and G(0)=0 (the latter guarantees that passage of the system from one down state to the next is never instantaneous). It follows that F(0)=0. Let

$$U(t) = \sum_{k=1}^{\infty} F^{*k}(t), \qquad (2.2)$$

where F^{*k} denotes the k- fold convolution of F with itself. It is well-known that the renewal function $U(t) < +\infty$ for each t ($0 \leq t < +\infty$) and $U(t) = EN(t)$ (cf. section 1). Consult Feller [11] for facts about U, but note that his U counts $S_0=0$ as the first renewal of the process $\{S_n\}$ and so equals 1+U, U being given as in (2.2). The definition (2.2) follows most "applied" probability and reliability texts (e.g. [1], [11], [4]).

The physical meaning of these four sets of r.v.'s is as stated in section 1. Mathematically, since we have assumed that all r.v.'s are independent, we can, without loss of generality, take them to be defined on the same probability space, Ω.

Set Q equal to the d.f. of the r.v. Z = X-M so that

$$Q(z) = \int_{0-}^{\infty} G(z+y)dL(y) \tag{2.3}$$

for all z in $(-\infty,\infty)$.

Results: We shall now state and discuss the results of this section; if a proof is cumbersome it is placed at the end of the section.

Theorem 1. If EX, EY and $E\tau$ are finite, then

$$P(\xi_\tau > M) = \int_0^{\infty}\{K(z)+\int_0^{\infty}(K(z+s)-K(s))dU(s)\}dQ(z) \tag{2.4}$$

Thus, (2.4) gives the probability that the system is up at some randomly selected moment in time, τ, and remains up for a random duration, M, of the mission occassioned by the request at time τ. By specifying only the d.f. of τ in Theorem 1, we have the following

Corollary 1. If the request time τ is exponentially distributed with mean $1/\lambda$, $(\lambda>0)$, then

$$P(\xi_\tau > M) = (1-\hat{F}(\lambda))^{-1} \int_0^{\infty}(1-e^{-\lambda z})dQ(z) \tag{2.5}$$

where $\hat{F}$ is the Laplace-Stieltjes transform of the d.f. F.

To verify the corollary from (2.4) just note that

$$K(s+z) - K(s) = e^{-\lambda s}(1-e^{-\lambda z}),$$

so that $P(\xi_\tau > M) = \int_0^{\infty}\{(1-e^{-\lambda z}) + (1-e^{-\lambda z})\int_0^{\infty} e^{-\lambda s}dU(s)\}dQ(z)$

$$= (1+\hat{U}(\lambda))\int_0^{\infty}(1-e^{-\lambda z})dQ(z),$$

where $\hat{U}$ is the Laplace-Stieltjes transform of U. Equation (2.5) follows since $\hat{U}(\lambda) = \hat{F}(\lambda)/(1-\hat{F}(\lambda))$ for all $\lambda>0$ (recall that $F(0)=0$).

Remark 1: It is both intuitively and analytically obvious that (2.5) may be written in the form

$$P(\xi_\tau > M) = P(\tau \leq X-M \mid \tau \leq X+Y). \tag{2.6}$$

Intuitively, because the exponential distribution has no memory and analytically, because

$$\int_0^{\infty}(1-e^{-\lambda z})dQ(z) = P(\tau \leq X-M) = P(\tau \leq X-M,\ \tau \leq X+Y)$$

and

$$1 - \hat{F}(\lambda) = P(\tau \leq X+Y).$$

Remark 2. The artificiality of the exponential assumption on τ can be attenuated somewhat by noting that if K is taken to be a mixture of exponentials:

$$K(z) = \Sigma a_\nu (1-e^{-\lambda_\nu z}), \tag{2.7}$$

$a_\nu \geq 0$ for all ν, $\Sigma a_\nu = 1$, $\lambda_\nu \geq 0$ (that is, the tail of K can be expressed as a Dirichlet series). Then (2.5) preserves in the form

$$P(\xi_\tau > M) = \sum_\nu a_\nu (1-\hat{F}(\lambda_\nu))^{-1} \int_0^\infty (1-e^{-\lambda_\nu z}) dQ(z).$$

Remark 3. Set Q_+ equal to the d.f. of $(X-M)^+$ where s^+ denotes the function which equals s if $s>0$ and 0 if $s \leq 0$. Then since $1-e^{-\lambda z}$ vanishes at 0, the only point on $[0,\infty)$ where Q_+ and Q differ, we can replace Q by Q_+ and write (2.5) as

$$P(\xi_{\tau(\lambda)} > M) = (1-\hat{Q}_+(\lambda))/(1-\hat{F}(\lambda)) \tag{2.8}$$

This form not only suggests easy computation (simulation is easily carried out from (2.6)), but it motivates the following observation: if $\lambda \to 0+$ (so that $E\tau \to +\infty$), then, writing $\tau = \tau(\lambda)$,

$$P(\xi_{\tau(\lambda)} > M) \to E(X-M)^+/(\mu_1+\mu_2) \tag{2.9}$$

where $\mu_1 = EX$, $\mu_2 = EY < +\infty$ and $E(X-M)^+ \leq EX < +\infty$.

Just recognize the RHS of (2.8) as the ratio of the difference quotients of $\hat{Q}_+$ and $\hat{F}$; passing to the limit as $\lambda \to 0+$ gives the ratio of the means of $(X-M)^+$ and X+Y (which both exist since $EX < +\infty$ and $EY < +\infty$).

As one would expect, the limit in (2.9) is preserved if the exponentiality of request time is dropped and $\tau(\lambda)$ is replaced by any sequence $\{\tau_n\}$ which converges in probability to $+\infty$.

Theorem 2. Let $\mu_1 = EX$ and $\mu_2 = EY$ be finite and T non-lattice. If τ_n $(n \geq 1)$ is a sequence of positive r.v.'s and $\tau_n \to +\infty$ in probability, then

$$P(\xi_{\tau_n} > M) \to E(X-M)^+/(\mu_1+\mu_2) \tag{2.10}$$

as $n \to +\infty$.
(The proof is at the end of this section.)

Remark 4. A simple calculation shows that

$$E(X-M)^+ = E(X-M|X>M)\ P(X>M) \tag{2.11}$$

Also, if we let the minimum of two real numbers a and b be denoted by $a \wedge b$ and observe the identity

$$(a-b)^+ = a-a\wedge b,$$

then we can express (2.10) in the following two equivalent forms

$$P(\xi_{\tau_n} > M) \to \frac{E((X-M)|X>M)}{\mu_1+\mu_2} \cdot P(X>M) = \frac{\mu_1 - E(X\wedge M)}{\mu_1+\mu_2} \tag{2.12}$$

as $n \to +\infty$.

When M=0 a.s., the RHS of both (2.10) and (2.12) reduce to the so-called "availability" of the system; $\mu_1/(\mu_1+\mu_2)$. The last relation in (2.12) is therefore especially intuitive since it shows directly the amount by which the availability should be decreased if one wants to account for the system being up throughout a mission of (random) duration, M.

In view of the above, it would seem to be appropriate to call

$$A(M) = \frac{E(X-M)^+}{\mu_1+\mu_2} \tag{2.13}$$

system availability for missions of length M.

Remark 5. When $t>0, \mathcal{M}>0$ are (nonrandom) real numbers and $\tau=t$, $M=\mathcal{M}$ (a.s.) then the classical limit of $P(\xi_t>\mathcal{M})$, as $t\to\infty$, (e.g. [1], [4], [11]) agree with all the above-mentioned forms; just note that $\int_{\mathcal{M}}^{\infty} \bar{G}(y)dy = \int_{\mathcal{M}}^{\infty}(y-\mathcal{M})dG(y) = \int_0^{\infty}(y-\mathcal{M})^+dG(y)$, $\bar{G} = 1 - G$.

Examples:

A. Let τ be exponential as in Corollary 1 and, in this first example, let X also be exponential with parameter θ_1 ($EX = \mu_1 = \theta_1^{-1}$). Whenever X has this distribution it follows from (2.3) that

$$Q(y) = e^{-\theta_1 y}\ \hat{L}(\theta_1),$$

where $\hat{L}$ is the Laplace-Stieltjes transform of L. Direct calculation then gives

$$\int_0^{\infty}(1-e^{-\lambda y})dQ(y) = \lambda\hat{L}(\theta_1)/(\lambda+\theta_1) \quad .$$

Since

$$\hat{F}(\lambda) = \theta_1 \hat{H}(\lambda)/(\lambda+\theta_1)$$

we find (using Corollary 1) that

$$P(\xi_\tau > M) = \frac{\mu_1 \hat{L}(\theta_1)}{\mu_1 + (1-\hat{H}(\lambda))\lambda^{-1}} \qquad (A.1)$$

where the distribution of M and Y remain to be specified. To note the resemblance to (2.12) just observe that $\hat{L}(\theta_1) = P(X>M)$, in this example. Of course, if we let $\lambda \to 0+$ in (A.1), we would obtain a special case of (2.12).

Now, if we further specify the distribution of Y to be exponential with parameter θ_2, $EY = \theta_2^{-1} = \mu_2$, we obtain (from (A.1))

$$P(\xi_\tau > M) = \frac{\mu_1}{\mu_1 + \frac{1}{\theta_2+\lambda}} \hat{L}(\theta_1) \qquad (A.2)$$

Finally, taking M to be exponential also, (A.2) becomes

$$P(\xi_\tau > M) = \frac{\mu_1}{\mu_1 + (\theta_2+\lambda)^{-1}} \cdot \frac{\mu_1}{\mu_1+\mu_M} \qquad (A.3)$$

where $\mu_M = EM$. So, in this case, $P(\xi_\tau > M)$ has the appearance of the product of two "availability" terms.

If, instead of being exponential, M is taken to be degenerate at $\aleph$, i.e., $L(s) = \varepsilon(s-\aleph)$, where ε is defined in (2.1), it follows from (A.1) that

$$P(\xi_\tau > M) = \frac{\mu_1}{\mu_1 + (1-\hat{H}(\lambda))\lambda^{-1}} e^{-\aleph/\mu_1} \qquad (A.4)$$

with the distribution of Y unspecified.

Notice that if $\lambda \to 0$ in (A.4) (or just use (2.12)) the RHS of (A.4) is

$$\frac{\mu_1}{\mu_1+\mu_2} \cdot e^{-\aleph/\mu_1} \qquad (A.5)$$

It is the almost exclusive use/misuse of this formula that causes one to produce the variations on a theme found in this note.

For example, one objectionable use of (A.5) is to specify λ and μ_2, and then to set the expression in (A.5) equal to some high number, such as .97, and solve for μ_1. (This, of course, is done with no knowledge that the life distribution is exponential.) This μ_1, call it μ_1^0, is then claimed, in advance development documents, to be the "required" mean-time-to-failure of the system; usually this is a complex military system which has either never been produced before or one for which we lack a substantial base-line of experience under a realistic mission profile. To make matters worse, this value of μ_1^0 and a similarly derived value, μ_1^1, obtained by setting (A.5) equal to some slightly smaller number such as .94, are used as the null and alternate hypotheses, respectively, in a statistical acceptance plan. Note that when this so-called acceptance plan is applied, it will be to a total population of perhaps one or two systems. Moreover, the system will be constantly undergoing design changes and differing conditions of stress. Needless to say, such practices often produce a reject signal from the testing community. If, on the basis of experience and common sense, the systems under test are judged to do their job reliably, at reasonable cost and more effectively than any system in the arsenal, these reject signals are properly ignored, but often not without the significant costs of re-tests, check-tests, needless re-design and a near infinity of meetings, briefings and "analyses".

The purpose then of the present note is to furnish Army statisticians with two more "degrees-of-freedom" (mission and request time distributions) in numerous formulas that will aid him in convincing the occasional naive practitioner of reliability that applications of (A.5) as described above are a totally unrealistic way of setting reliability requirements. This can be done by producing a wide variety of answers with judicious choice of distributions for mission and request times. The variability obtained through distribution which cannot be predicted might be enough to convince the R&D community to state reliability figures-of-merit as goals-to-point-toward and not hard requirements to be "demonstrated" in some psuedo-statistical test. The only possible danger is that the results stated in this paper will be misused in the same way as (A.5). It should be emphasized that designing reliable military systems is of the utmost importance and it is <u>not</u> the purpose of these remarks to argue otherwise. On the contrary, it is hoped that by discouraging an absurd approach to setting reliability requirements emphasis will be placed on engineering reliability into new systems.

Before concluding example A, consider two additional distributions for M. First, when M is uniformly distributed over (0,T):

$$P(\xi_T > M) = \frac{1-e^{-\theta_1 T}}{\theta_1 T} \quad \frac{\mu_1}{\mu_1+(\lambda+\theta_2)^{-2}} \qquad \text{(A.6)}$$

with τ, X and Y exponential as above.

The second is when M is normally distributed as $N(\gamma,\sigma)$ conditioned to be positive. Then

$$P(\xi_\tau > M) = \frac{\mu_1}{\mu_1+(\theta_2+\lambda)^{-1}} \exp\left(-(\theta_1\gamma - \frac{(\theta_1\sigma)^2}{2}\right) \frac{\Phi(\frac{\gamma}{\sigma}-\sigma\theta_1)}{\Phi(\frac{\gamma}{\sigma})} \qquad (A.7)$$

where Φ is the d.f. of the standard, N(0,1), normal r.v.

B. Because of the ease of calculation we consider the case when $\overline{X}$ is Rayleigh (α),

$$P(X>s) = e^{-s^2/2\alpha^2},$$

and M is Rayleigh (σ). Then

$$EX\wedge M = \int_0^\infty e^{-s^2/2\alpha^2} \cdot e^{-s^2/2\sigma^2} ds$$

$$= \int_0^\infty e^{-s^2/2\beta^2} ds = \sqrt{\frac{\pi}{2}}\,\beta$$

where $\beta = \alpha\sigma/\sqrt{\alpha^2+\sigma^2}$ and so

$$E(X-M)^+ = EX - EX\wedge M = \sqrt{\frac{\pi}{2}}\,\alpha\,(1-\frac{\sigma}{\sqrt{\alpha^2+\sigma^2}})$$

Since $EX=\sqrt{\frac{\pi}{2}}\,\alpha$, we therefore can write (2.13) in the form

$$A(M) = \frac{EX}{EX+EY} \cdot (1-\frac{\sigma}{\sqrt{\alpha^2+\sigma^2}}) \qquad (B.1)$$

$$= A(0)\,(1-\frac{\sigma}{\sqrt{\alpha^2+\sigma^2}})$$

Another simple application of Theorem 2 is obtained when X is Gamma $(2,\beta)$:

$$P(X>s) = e^{-\beta s}\,(1+\beta s)$$

and M is the square of a $N(0,\sigma)$ r.v.. Then (after some tedious calculations)

$$A(M) = A(0)\,\frac{(1+(1.5)\sigma^2\beta)}{(1+2\sigma^2\beta)^{3/2}} \qquad (B.2)$$

C. We now return to Theorem 1 and show its relationship to some known results on availability, without using the exponential assumption of Corollary 1, for the request time distribution.

For this purpose, let the request time be a fixed constant, i.e.,

$$\tau(\omega) = T > 0$$

for all $\omega\varepsilon\Omega$. Then $K(x) = \varepsilon(x-T)$, where ε is defined in (2.1).

Then the function

$$\Psi(s,z) = k(s+z) - k(s)$$

is equal to one in the unbounded region defined by $0 < s < T$, and $s + z > T$ and zero otherwise.

It follows that the RHS of (2.4) is given by

$$\int_0^\infty K(y)dQ(y)+\int_0^\infty\int_0^\infty (K(s+y)-K(s))dQ(y)dU(s) = \qquad (C.1)$$

$$= \int_T^\infty dQ(y) + \int_0^T \bar{Q}(T-s)dU(s) \quad , \ \bar{Q} = 1 - Q.$$

For ease of computation, let X, Y and M be exponential with $\mu_1=\theta_1^{-1}$, $\mu_2=\theta_2^{-1}$ and $EM=\mu_M=\alpha^{-1}$, respectively. Then it is easy to show that

$$U(t) = \frac{1}{\mu}(t- \frac{1}{a} (1-e^{-at})),$$

where $\mu=\mu_1+\mu_2$ and $a=\theta_1+\theta_2$.

Since, $\bar{Q}(y) = \alpha \exp(-\theta_1 y)/(\theta_1+\alpha)$, for $y>0$, we obtain ((2.4) and (C.1))

$$P(\xi_T > M) = \frac{\mu_1}{\mu_1+\mu_M} \cdot \frac{1}{\mu} \ [\mu_1+ \mu_2 e^{-aT}] \qquad (C.2)$$

Notice that if $T\to+\infty$, (C.2) becomes (A.3) with $\lambda=0$ as it should. Also, if $\mu_M=0$ in(C.2)then (23) of [12] is obtained.

Proof of Theorem 1. Proceeding either with a standard Renewal theory argument or directly from the equation for the d.f. of ξ_t, t nonrandom, in [1] p.60 or [II] p.354 we obtain

$$P(\xi_t > M) = \overline{Q}(t) + \overline{Q}*U(t), \tag{2.14}$$

where Q is given in (2.3). Alternately, this is a special case of a more general (non-identically distributed) case derived in section 4. Since K(0)=0 (K=d.f. of τ)

$$\int_0^\infty \overline{Q}(t)\,dK(t) = \int_0^\infty K(t)\,dQ(t)$$

Consider

$$\int_0^\infty \overline{Q}*U(t)\,dK(t) = \int_0^\infty \int_0^t \overline{Q}(t-y)\,dU(y)\,dK(t)$$

$$= \int_0^\infty \int_0^\infty \overline{Q}(s)\,dK(s+y)\,dU(y)$$ and observe that this last integral, call it I, is finite. This follows from $E\tau<+\infty$, and the well-known fact that $U(y)\sim y/\mu$ as $y\to\infty$, since then

$$I \leq \int_0^\infty (1-K(y))\,dU(y) = \int_0^\infty U(y)\,dK(y) = 0(E\tau) < +\infty$$

(we have made use of the fact that $U(y)(1-K(y)) \sim \frac{1}{\mu}\, y(1-K(y)) \to 0$ if $E\tau < +\infty$). Now

$$\int_0^\infty \overline{Q}(s)\,d\overline{K}(s+y) = -\overline{Q}(0)\overline{K}(y) + \int_0^\infty \overline{K}(s+y)\,dQ(s)$$

so that

$$\int_0^\infty \overline{Q}*U(t)\,dK(t) = -\int_0^\infty \int_0^\infty \overline{Q}(s)\,d\overline{K}(s+y)\,dU(y)$$

$$= \overline{Q}(0)\int_0^\infty \overline{K}(y)\,dU(y) - \int_0^\infty \int_0^\infty \overline{K}(s+y)\,dQ(s)\,dU(y)$$

$$= \int_0^\infty \int_0^\infty \overline{K}(y)\,dU(y)\,dQ(s) - \int_0^\infty \int_0^\infty \overline{K}(s+y)\,dQ(s)\,dU(s)$$

$$= \int_0^\infty \int_0^\infty (K(s+y)-K(y))\,dU(y)\,dQ(s)$$

Proof of Theorem 2. The sequence $\tau_n \to +\infty$ in probability if given $\varepsilon>0$, $A>0$ there exists an integer $n_o=n_o(\varepsilon,A)$ such that

$$P(\tau_n > A) \geq 1-\varepsilon \quad \text{if } n>n_o.$$

Letting K_n be the d.f. of τ_n, $n \geq 1$, we can write

$$P_n = P(\xi_{\tau_n} > M) = \int_0^\infty \bar{Q}(t)dK_n(t) + \int_0^\infty \bar{Q}*U(t)dK_n(t)$$

$$= I(n) + J(n)$$

Now, let $\varepsilon>0$ be arbitrary and choose $A>0$ such that $\bar{Q}(A)=P(X-M>A)<\varepsilon$, then,

$$0 \leq I(n) = \int_0^A \bar{Q}(t)dK_n(t) + \int_A^\infty \bar{Q}(t)dK_n(t)$$

$$\leq K_n(A) + \varepsilon(1-K_n(A))$$

$$\leq K_n(A) + \varepsilon$$

so that

$$0 \leq \limsup_{n\to\infty} I(n) \leq \varepsilon .$$

Therefore, since $\varepsilon>0$ is arbitrary,

$$\lim_{n\to\infty} I(n) = 0.$$

For the term $J(n)$, we of course follow the usual proof and use the Key-Renewal Theorem. This places an integrability requirement on Q which is equivalent (in our case) to showing that $\int_0^\infty \bar{Q}(t)dt < +\infty$. This follows from $EX < +\infty$. The assumption that T is non-lattice is trivial in our application and can be guaranteed by requiring, for example, either X or Y to have absolutely continuous d.f.'s.

The Key-Renewal Theorem then states that

$$\bar{Q}*U(t) \to \frac{1}{\mu}\int_0^\infty \bar{Q}(y)dy$$

as $t \to +\infty$ where $\mu=\mu_1+\mu_2$. In what follows, call this limit B.

The argument for J_n is similar to the one applied to I_n. That is, by the previous limit, there is some C such that $Q*U(t)$ is within a preselected distance $\delta>0$ of B for all $t>C=C(\delta)$. Given some $\varepsilon>0$, the convergence to $+\infty$ of τ_n is then used to find an $n_o=n_o(\varepsilon,C)$ so that $P(\tau_n>C) \geq 1-\varepsilon$ if $n>n_o$. All this allows us to conclude that both

$$\int_C^\infty \bar{Q}*U(t)dK_n(t) \geq (B-\delta)P(\tau_n>C) \geq (B-\delta)(1-\epsilon) = B-\epsilon'$$

and

$$\int_C^\infty \bar{Q}*U(t)dK_n(t) \leq (B+\delta)$$

if $n>n_o$.

Since, also,

$$0 \leq \int_0^C \bar{Q}*U(t)dK_n(t) = O(P(\tau_n \leq C) = o(1)$$

as $n\to\infty$, we have

$$\lim_{n\to\infty} J(n) = B = \frac{1}{\mu}\int_0^\infty \bar{Q}(t)dt \ .$$

It remains to evaluate the integral of $\bar{Q}$: (Recall $EX<+\infty$)

$$\int_0^\infty \bar{Q}(y)dy = \int_0^\infty\int_0^\infty P(X>y+m)dL(m)dy$$

$$= \int_0^\infty\int_0^\infty P(X>y+m)dy\ dL(m)$$

$$= \int_0^\infty\int_m^\infty P(X>s)ds\ dL(m)$$

$$= \int_0^\infty P(X>s)\ \int_0^s dL(m)ds$$

$$= \int_0^\infty P(X>s)\ P(M\leq s)ds$$

$$= EX - \int_0^\infty P(X>s,\ M>s)ds$$

$$= EX - \int_0^\infty P(X\wedge M>s)ds$$

$$= EX - E(X\wedge M)$$

where $X\wedge M$ = minimum of X and M. Clearly, $0 \leq EX-E(X\wedge M) \leq EX < +\infty$.

3. Additional Comments on the IID Case. Using the stochastic model of section 2, the percentage of time, during n renewals of the system, that the system is up and remains up for a sufficient amount of time to support a mission of length $\ast$ is given by

$$p_n = p_n(\ast) = \sum_{i=1}^{n} (X_i - \ast)^+ / S_n \ , \tag{3.1}$$

($n \geq 1$). (Throughout this section, $\ast$ will be a strictly positive real number.) Assuming that $ET = EX + EY < +\infty$, it follows from the law of large numbers and Slutsky's Theorem (cf. Cramér [3] p. 255) that

$$p_n(\ast) \to \frac{1}{\mu} E(X-\ast)^+, \tag{3.2}$$

in probability as $n \to \infty$, $\mu = \mu_1 + \mu_2 = EX + EY$.

Thus, the statistic $p_n(\ast)$ is a consistent estimator of the quantity $E(X-\ast)^+/\mu$, the ubiquitous limiting interval reliability of [1] and a special case of Corollary 1. The simple, practical nature of $p_n(\ast)$ probably explains the interest in describing systems by means of interval reliability.

A related statistic with similar intuitive appeal is

$$\psi_n(\ast) = \sum^{n} (X_i - \ast)^+ / \sum^{n} X_i$$

Clearly, this statistic gives the percentage of up-time that the system is available for a mission of length $\ast$ and is a consistent estimator of the quantity $\psi(\ast) = E(X-\ast)^+/\mu_1$. From Corollary 1 of section 2, this quantity is also easily seen to be the limit of the probability that $\xi_{\tau_n} > \ast$ given that $\xi_{\tau_n} > 0$ as $n \to \infty$, when $\tau_n \to +\infty$, in probability.

Using the work of Skorohod [9] Chapter 1, Sec. 6, Pyke [7], Pyke and Shorack [8], and arguments similar to those in recent work of Barlow and Proschan [2], it can be shown (under additional assumptions) that $\sqrt{n}\,(\psi_n(\ast) - \psi(\ast))$ converges in probability to a normally distributed r.v., $N(0, \sigma(\ast))$, where the variance can be calculated explicitly, in terms of $\psi(\ast)$, μ_1, Var X, Var $(X-\ast)^+$, and the d.f.G. The proof is outside the scope of this note and will be reported elsewhere.

The usefulness of such a result is that it places emphasis on $\psi_n(\ast)$, a directly measurable quantity, rather than on $\psi(\ast)$, which requires a distributional assumption.

(4.0). Residual Life: non-identically distributing case.

In this section we suppose only that the sequences $\{X_i\}$ and $\{Y_j\}$ of positive random variables are each sequences of independent r.v.'s and, further, that the sequence $\{X_i\}$ is independent of the sequence $\{Y_j\}$.

Let G_i be the distribution function (d.f.) of X_i, $i \geq 1$, H_j the d.f. of Y_j, $j \geq 1$, and set F_i equal to the d.f. of $T_i = X_i + Y_i$, $i \geq 1$. As before, let M be a positive r.v. with d.f. L and assume that M and the $\{X_i\}$ and $\{Y_j\}$ sequences are independent.

Set Q_i = d.f. of the r.v. $Z_i = X_i - M$, so that

$$Q_i(z) = \int_{-\infty}^{+\infty} G_i(z+y)\,dL(y) \tag{4.1}$$

for all z in $R_1 = (-\infty, \infty)$.

Finally, observe that since we have not assumed that the T_j, $j \geq 1$, are identically distributed r.v.'s, it is possible for the partial sums S_n of section 1 to converge to some proper r.v. in distribution (and hence with probability one (a.s.) on Ω). For simplicity, we want to avoid this possibility and retain the property of IID r.v.'s which states that $S_n \to +\infty$ (a.s.). Thus, when we consider an instance where $\sum^{\infty} X_j$ converges, the divergence of S_n to $+\infty$ will be guaranteed (even though the Y_i's are not identically distributed) by assuming that $\sum^{n} Y_i \to +\infty$ (a.s.).

Now, recall the definition of ξ_t, for non-random $t \geq 0$, given in section 1 and partition the interval $[0,\infty)$ by the sequence of partial sums S_n, $n \geq 0$. Then

$$P(\xi_t > M) = \sum_{k=0}^{\infty} P(\xi_t > M,\ S_k \leq t < S_{k+1}) \tag{4.2}$$

$$= \sum_{k=0}^{\infty} \int_0^t P(\xi_t > M,\ t < S_{k+1} \mid S_k = x)\, dP(S_k \leq x)$$

$$= \sum_{k=0}^{\infty} \int_0^t P(X_{k+1} - M > t - x,\ t - x < T_{k+1})\, dP(S_k \leq x)$$

$$= \sum_{k=1}^{\infty} \int_0^t P(Z_{k+1} > t - x)\, d\Big[\prod_{j=1}^{k} * F_j\Big](x) + P(Z_1 > t)$$

where, after conditioning on S_k we have used a familiar property of conditional probabilities (cf. Krickeberg [6] p. 170 problems 3 and 4), the independence of T_{k+1} and S_k and, for the last equality, the fact that the occurrence of the event $[Z_{k+1} > t-x]$ implies the occurrence of the event $[T_{k+1} > t-x]$. Therefore, using the d.f.'s introduced above and the usual notation for a convolution product:

$$\prod_{j=1}^{k} * F_j(t) = (F_1 * F_2 * \cdots * F_k)(t) = P(S_k \leq t),$$

we can write (4.2) in the form

$$P(\xi_t > M) = \bar{Q}_1(t) + \sum_{k=1}^{\infty} \bar{Q}_{k+1} * \prod_{j=1}^{k} * F_j(t) \qquad (4.3)$$

where $\bar{Q}_j = 1 - Q_j$ and $t \geq 0$.

It is easy to see that under the assumption of the last section (that is, where the sequences are identically as well as independently distributed), the last equation reduces to equation (2.14) of section 2.

Let the r.v. $\eta(M)$ be the amount of time that the random function $t \to \xi_t$ is greater than M. If I is used to denote the indicator function of the set of positive real numbers; that is,

$$I(y) = \begin{cases} 1, & y > 0 \\ 0, & y \leq 0 \end{cases} \qquad (4.4)$$

then $\eta(M)$ can be written as

$$\eta(M) = \int_0^{\infty} I(\xi_t - M)\,dt \qquad (4.5)$$

Of course, $\eta(M)$ may be a defective r.v. in the sense that it may take the value $+\infty$ with positive probability. Taking expectations of both sides of (2.5) it is easy to see that

$$E\eta(M) = \int_0^{\infty} P(\xi_t > M)\,dt \qquad (4.6)$$

whether the RHS is finite or not.

We note in passing that the case when the underlying stochastic structure consists of sequences of IID r.v.'s, the RHS of (4.6) is infinite. This fact might motivate one to ask whether or not this integral is Abel summable to a finite value. That is, does

$$A(\lambda) = \int_0^\infty \lambda e^{-\lambda t} P(\xi_t > M)\,dt$$

converge as $\lambda \to 0+$? It is amusing to recognize this integral as $P(\xi_{\tau(\lambda)} > M)$, where $\tau(\lambda)$ is an exponentially distributed r.v. and apply Remark 4 or Theorem 2 of section 2 to obtain $A(\lambda) \to \mu^{-1} E(X-M)^+ < +\infty$ as $\lambda \to 0+$, if $\mu < +\infty$.

Alternately, use only the classical case with M random; then an application of the Dominated Convergence Theorem gives

$$A(\lambda) = \int_0^\infty e^{-y} P(\xi_{y\lambda^{-1}} > M)\,dy \to \mu^{-1} E(X-M)^+$$

as $\lambda \to 0+$.

Returning to the non-IID case we can state the following

<u>Theorem 3</u>: <u>If the series</u> $\sum_1^\infty E(X_\nu - M)^+ < +\infty$ <u>then</u>

$$E\eta(M) = \sum_{\nu=1}^\infty E(X_\nu - M)^+ \tag{4.7}$$

This follows easily. Just let $V_n(t)$ denote the general term in the series (4.3) and note that

$$\int_0^\infty V_n(t)\,dt = E(X_{n+1} - M)^+$$

Then since the V_n are non-negative and integrable over $[0,\infty)$, and the series of integrals of the V_n converge, equation (4.7) follows from a well-known result about interchanging summation and integration (e.g. page 114, (2) [5]). This proves (4.7). This equation then shows that $\eta(M)$ is a proper r.v. and

$$\eta(M) = \sum_{\nu=1}^\infty (X_\nu - M)^+$$

with probability one.

A simple example of (4.7) is obtained by assuming that both X_ν and M are exponential with mean values θ_ν and C, respectively. Then $E(X_\nu - M)^+ = EX_\nu - EX_\nu \wedge M = \theta_\nu^2/(\theta_\nu + C)$. Then if we take $\theta_\nu = C/\nu$, $En(M) = C$.

BIBLIOGRAPHY

1. Barlow, R.E. and Proschan, F., Mathematical Theory of Reliability, John Wiley & Sons, New York, 1967.

2. Barlow, R. E. and Proschan, F., "Asymptotic Theory of Total Time on Test Processes, with Applications to Life Testing" Operations Research Center, ORC 75-18, University of Calif., Berkely, Sept. 1975.

3. Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princ. N.J., 1958.

4. Gnedenko, B.V., Belyayev, Yu. K. and Solovyev, A., Mathematical Methods of Reliability, Academic Press, N.Y. 1969.

5. Halmos, P., Measure Theory, Van Nostrand Co., Princeton, N.J. 1959.

6. Krickeberg, K., Probability Theory, Addison- Wesley Publishing Col, 1965.

7. Pyke, R., "Applications of Almost Surely Convergent Constructions of Weakly Convergent Process" Mathematical Note No. 50, Mathematics Research Laboratory, Boeing Scientific Res. Lab., August 1968.

8. Pyke, R. and Schorack, G., "Weak Convergence of a two sample empirical process and a new approach to the Chernoff-Savage Theoran" Ann. Math. Statist. 39, No. 3, June 1968, p. 755.

9. Skorokhod, A. V., Studies in The Theory of Random Processes, Addison-Wesley Pub Co., Reading, Mass. 1965.

10. Introduction to Reliability (R), Availability (A) and Maintainability (M), ALMC report: ALM-35-3606-LC.

11. Feller, W., An Introduction to Probability Theory and Its Applications, Vol II., John Wiley and Sons, N.Y. 1966.

12. Kurkjian, B., and Inselmann, E., "Availability, Revisited" Transactions of the Eighteenth Conference of Army Mathematicians, ARO-D Report 73-1.

CONFIDENCE INTERVALS FOR A SUM OF RENEWAL PROCESSES WITH APPLICATION IN RELIABILITY

Ronald L. Racicot
Research Directorate
Benet Weapons Laboratory
Watervliet Arsenal
Watervliet, New York 12189

ABSTRACT. In reliability theory, the time flow of failures of a non-constant failure rate component which is replaced or renewed upon failure forms a renewal process. The inter-arrival times of failures in this case are independent identically distributed positive random variables. If a system which is composed of a number of such components is considered to have failed if one of its components fails, then the total number of system failures is a sum of the individual renewal processes. The problem considered in this paper is the computation of confidence intervals for the total number of system failures over a given period of time from total system tests and/or individual component tests. Although the application considered is one from reliability theory, the results are applicable to general sums of renewal processes.

In solving this particular problem, the reliability engineer often assumes that the sum of renewal processes asymptotically approaches a non-homogeneous Poisson process or, after a long period of time, a homogeneous Poisson process with exponentially distributed inter-arrival failure times. For these processes, a chi-square distribution can be used to determine confidence intervals for total number of failures from which confidenced reliability or MTBF can be determined. It can be shown, however, that the Poisson process is strictly a local property for sums of renewal processes and that confidence intervals derived from these assumptions are generally incorrect. This is shown by comparing the true variance of the number of system failures with the variance derived assuming the Poisson process.

A scheme for computing confidence intervals is presented in which the first 3 moments of failure times of the component processes are used to compute the mean and variance of total system failures. For a large number of components, the normal distribution adequately describes the distribution of system failures from which confidence intervals can be estimated.

NOTATION.

$f(t)$	pdf of inter-arrival times of failures;
$F(t)$	cdf corresponding to $f(t)$;
$\bar{F}(t)$	$1 - F(t)$;
$h(t)$	renewal rate; the unconditional pdf of component failure and subsequent renewal;
$h_j(t)$	renewal rate for component j;
$H(t)$	expected value of the number of system failures over the interval $(0,t)$;
$H_j(t)$	renewal function for component j; the integral of $h_j(t)$ over the interval $(0,t)$;
$\hat{H}(t)$	point estimate of $H(t)$;
$H_{true}(t)$	true value of $H(t)$;
$N(t)$	number of system failures over the interval $(0,t)$;
$N_j(t)$	number of failures of component j over the interval $(0,t)$;
n_c	number of components;
n_f	number of component failures;
n_m	number of missions over system life;
$P_N(t)$	probability of N failures in time t;
$R(t,\tau)$	reliability at time t for an interval τ;
$R_j(t,\tau)$	reliability of the jth component;
$R_a(\tau,n_m)$	average interval-reliability over system life for interval τ and number of intervals n_m;
$R_s(t,\tau)$	system reliability at time t for an interval τ;
$R_{sa}(\tau,n_m)$	average system reliability;

t	system time;
β	Weibull shape parameter;
η	Weibull scale parameter;
μ_j	mean inter-arrival failure time for component j;
μ_{3j}	third central moment of inter-arrival failure times for component j;
σ_j^2	variance of inter-arrival failure times for component j;
	and
τ	interval or mission length for which reliability is required.

1. INTRODUCTION. The general problem is to determine confidence intervals for reliability of a series system of components from test data. Previous solutions to this problem have been limited to constant failure rate components, binomial mission reliability which is constant in time and/or reliability for only the first system failure [1,2]. The case considered in this paper which is often of more interest to the reliability test engineer involves a system comprised of mechanical components which follow non-constant failure rate distributions. The system is operated continuously until failure of any of its components occurs at which time the component is replaced or renewed and system operation continued.

For the single component which is replaced or renewed upon failure, the renewal rate h(t) describes the unconditional failure rate of the component and is derived from the underlying distribution of inter-arrival failure times [3,4]:

$$H(t) = f(t) + \int_0^t f(t-x)h(x)dx. \tag{1}$$

The renewal rate is distinguished here from the hazard or conditional failure rate which describes failure of a non-repairable item.

Interval or mission reliability can be determined from the renewal rate [5-7]:

$$R(t,\tau) = 1 - \int_t^{t+\tau} \bar{F}(t+\tau-x)h(x)dx \tag{2a}$$

$$\approx 1 - \int_t^{t+\tau} h(x)dx \quad \text{for small } \tau. \tag{2b}$$

For practical applications, the transient interval-reliability can be average over system life to yield a single time independent reliability index that characterizes a given component:

$$R_a(\tau,n_m) = \frac{1}{n_m} \sum_{i=1}^{n_m} R(t_i,\tau) \tag{3}$$

For a series system of components

$$R_s(t,\tau) = \prod_{j=1}^{n_c} R_j(t,\tau) \tag{4}$$

and

$$R_{sa}(\tau,n_m) = \frac{1}{n_m} \sum_{i=1}^{n_m} \prod_{j=1}^{n_c} R_j(t_i,\tau) \tag{5}$$

The time flow of failures of a non-constant failure rate component which is replaced or renewed upon failure forms a renewal process [3]. The inter-arrival times of failures in this case are independent identically distributed positive random variables. If a system which is composed of a number of such components is considered to have failed if one of its components fails (series system assumption), then the total number of system failures is a sum of the individual renewal processes. The problem considered here is the computation of confidence intervals for the total number of system failures over a given period of time from total system tests and/or individual component tests. Although the application considered is one from reliability theory, the results are applicable to general sums of renewal processes.

Many properties of renewal processes and sums of renewal processes are covered in the literature; so only the final results are summarized here [3-7]. If $N_j(t)$ represents the total number of failures of component j over time interval (0,t) then for the system

$$N(t) = \sum_{j=1}^{n_c} N_j(t). \tag{6}$$

For components which fail independently of one another, the mean and variance of N(t) is equal to the sum of the mean and variance of the

component processes:

$$H(t) = E\{N(t)\} = \sum_{j=1}^{n_c} H_j(t) \tag{7}$$

$$h(t) = \frac{dH(t)}{dt} = \sum_{j=1}^{n_c} h_j(t) \tag{8}$$

$$Var\{N(t)\} = \sum_{j=1}^{n_c} Var\{N_j(t)\}. \tag{9}$$

For small mission time interval τ and a large number of components, the average reliability (5) can be shown to asymptotically approach the following value [5]:

$$R_{sa}(\tau, n_m) \approx 1 - \frac{1}{n_m} H(n_m \tau). \tag{10}$$

In reliability applications then, where the above assumptions hold, it suffices to deal with H(t) for the system with reliability being determined from (10).

2. NON-HOMOGENEOUS POISSON PROCESS AS AN APPROXIMATION TO N(t).
In considering the problem of non-constant failure rate components, the reliability engineer often assumes that the sum of renewal processes asymptotically approaches a non-homogeneous Poisson process (NHPP) with increasing number of components or, after a long period of time, a homogeneous Poisson process (HPP) with exponentially distributed inter-arrival failure times [5]. For these processes, the chi-square distribution can be used to determine confidence intervals for total number of failures from which confidenced reliability or MTBF (mean-time-between-failures) can be determined. In what follows, however, it is readily shown that the Poisson process is strictly a local property for sums of renewal processes and that the global confidence intervals derived from these assumptions are generally incorrect.

The distribution of number of failures for the NHPP is given as

$$P_N(t) = \frac{H(t)^N}{N!} e^{-H(t)} \tag{11}$$

$$E\{N(t)\} = H(t) \tag{12}$$

$$\mathrm{Var}\{N(t)\} = H(t) \tag{13}$$

It suffices to show that the true variance of the sum of renewal processes does not generally equal H(t) as shown by (13). Consider, for example, the asymptotic renewal process for large t in which the mean and variance for component j are given by [3]

$$\lim_{t\to\infty} \frac{H_j(t)}{t} = \frac{1}{\mu_j}; \quad H_j(t) \simeq \frac{t}{\mu_j} \tag{14}$$

$$\mathrm{Var}\{N_j(t)\} \simeq \frac{\sigma_j^2}{\mu_j^3}\, t \tag{15}$$

in which μ_j and σ_j^2 are the mean and variance of the inter-arrival failure times. Using (7) and (9) gives

$$H(t) \simeq \sum_{j=1}^{n_c} \frac{1}{\mu_j} \tag{16}$$

$$\mathrm{Var}\{N(t)\} \simeq \sum_{j=1}^{n_c} \frac{\sigma_j^2}{\mu_j^3} \tag{17}$$

In general, $H(t) \neq \mathrm{Var}\{N(t)\}$ and the sum of renewal processes for this example does not approach a NHPP or HPP in a global sense no matter how large n_c becomes. For equal components, for example, $1/\mu \neq \sigma^2/\mu^3$ unless $\sigma^2=\mu^2$. This is the case for the exponential distribution but is only a special case for other distributions. Although the asymptotic process for large t was considered, the same can be shown for the sum of ordinary renewal processes.

3. CONFIDENCE INTERVALS USING COMPONENT MOMENTS. Since the sum of renewal processes (6) is a sum of discrete, lattice type random variables, it asymptotically approaches the normal distribution as an envelope with increasing number of components [8]. Confidence intervals then can be estimated for H(t) using normal tables for large number of components with $\hat{H}(t)$ and its variance being determined from test data.

As will be shown later, an extra failure should be added to $\hat{H}(t)$ in determining upper confidence limits to remove bias.

The renewal function for component j can be estimated from the moments of the inter-arrival times of events for large t [3].

$$H_{jo}(t) \approx \frac{t}{\mu_j} + \frac{\sigma_j^2-\mu_j^2}{2\mu_j^2} + O(1/t) \tag{18}$$

$$Var\{N_{jo}(t)\} \approx \frac{\sigma_j^2 t}{\mu_j^3} + \left(\frac{1}{12} + \frac{5}{4}\frac{\sigma_j^4}{\mu_j^4} - \frac{2}{3}\frac{\mu_{3j}}{\mu_j^3} \right) + O(1/t) \tag{19}$$

for the ordinary renewal process and

$$H_{je}(t) \approx \frac{t}{\mu_j} \tag{20}$$

$$Var\{N_{je}(t)\} \approx \frac{\sigma_j^2 t}{\mu_j^3} + \left(\frac{1}{6} + \frac{\sigma_j^4}{2\mu_j^4} - \frac{\mu_{3j}}{3\mu_j^3} \right) + O(1/t) \tag{21}$$

for the equilibrium renewal process. In the ordinary renewal process all components are new at t=0. The equilibrium process, on the other hand, is one which has been running for a long time before it is first observed (see Cox [3], Chapter 2 for more detailed description of these processes).

Case 1: Complete Samples with large t

For this case the moments can be estimated without making any assumption about the underlying distribution:

$$\hat{\mu}_j = \sum_{i=1}^{n_{fj}} x_{ji}/n_{fj} \tag{22a}$$

$$\hat{\sigma}_j^2 = \sum_{i=1}^{n_{fj}} (x_{ji}-\hat{\mu}_j)^2/(n_{fj}-1) \tag{22b}$$

$$\hat{\mu}_{3j} = n_{fj} \sum_{i=1}^{n_{fj}} (x_{ji}-\hat{\mu}_j)^3/(n_{fj}-1)(n_{fj}-2) \qquad (22c)$$

$$\mathrm{Var}\{\hat{H}_j(t)\} = \mathrm{Var}\{N_j(t)\}/n_{fj} \qquad (23)$$

in which x_{ji}, $i=1,\ldots, n_{fj}$ are n_{fj} failure times for component j. Substituting (22) into (18), (19) and (23) or (20), (21) and (23) yields component estimates for $\hat{H}_j(t)$ and $\mathrm{Var}\{\hat{H}_j(t)\}$. System $\hat{H}(t)$ and its variance can then be determined from (7) and (9) from which confidence limits on the true value of $H(t)$ can be estimated using normal tables.

Case 2: Censored Samples

For this case, a theoretical distribution for inter-arrival failure times must be assumed, such as the Weibull or gamma, with the moments being estimated, for example, using maximum likelihood. Confidence limits can then be determined assuming the normal distribution for total number of pooled failures.

4. SOME NUMERICAL RESULTS FOR CASE 1

A particular example has been considered to study the frequency exactness of the confidence limits described above. For this study Monte Carlo simulation is used to artificially generate sample outcomes for a system with given component parameters. The system is assumed to be composed of n_c identical Weibull components with parameters η and β. Using these parameters, failure times for a given number of failures are generated for each component using random numbers with the quantities $\hat{\mu}_j$, $\hat{\sigma}_j^2$ and $\hat{\mu}_{3j}$ being computed from (22). From these $\hat{H}_j(t)$ and $\hat{\mathrm{Var}}\{N_j(t)\}$ are computed using (18) and (19) where large t is assumed. Estimates for the system $\hat{H}(t)$ and $\mathrm{Var}(\hat{H}(t))$ are then determined from (7), (9) and (23).

Assuming the normal distribution for $\hat{H}(t)$, confidence limits on $H(t)$ can be determined from the given set of sample outcomes. This is repeated 1000 times for a fixed set of parameters. The normal cdf, gauf $(H(t))$, is evaluated at the true and known value of $H(t)$ for each of these sample outcomes. For exact frequency confidence intervals,

the function gauf ($H_{true}(t)$) should be uniform on (0,1.0). Results indicate that although the confidence limits are not exact, they are close enough for practical purposes.

Table I lists some of the results of these trials for the upper 90% confidence limit on H(t) (lower 90% confidence limit on average reliability). An extra failure had to be added to the total number of system failures to remove bias. For exactness, the percent of trials in which H_{true} is greater than the upper 90% confidence limit, $\hat{H}_{90}$, should be 10%. As can be seen from the results in Table I, the confidence limits are close to this requirement. The confidence limit $\hat{H}_{90}$, therefore, is judged to be exact for this case as long as one extra failure is added to total number of test failures.

The main limitations of the above approach are the requirement for long system times and large number of components and/or failures for exactness. Also, in computing reliability from H(t), small mission times (high reliability) are required for the approximation (10). The computational methods involved, however, are relatively straightforward and the approach appears to be a sound one.

TABLE I

RESULTS OF MONTE CARLO TRIALS TO STUDY UPPER 90% CONFIDENCE LIMIT FOR SUM OF RENEWAL PROCESSES

NUMBER OF COMPONENTS	NUMBER OF FAILURES PER COMPONENT	$H_{true}(t=5)$	% OF TRIALS $H_{true} > \hat{H}_{90}$ *
10	10	51.7	9.8
10	5	51.7	10.6
5	5	25.8	9.6
2	5	10.3	7.1

* 90 PERCENTILE OF DISTRIBUTION GAUF ($\hat{H}$+1, $\sigma^2_{\hat{H}}$)

WEIBULL COMPONENT PARAMETERS: $\eta = 1.0$, $\beta = 3.0$

REFERENCES.

[1] N. R. Mann and F. E. Grubbs, "Approximately Optimum Confidence Bounds on Series System Reliability for Exponential Time to Failure Data," Biometrika, Vol. 59, 1972, pp. 191-204.

[2] A. Winterbottom, "Lower Confidence Limits for Series System Reliability from Binomial Subsystem Data," J. of the American Statistical Association, Vol. 69, No. 347, September 1974. pp. 782-788.

[3] D. R. Cox, Renewal Theory, New York: Wiley, 1963.

[4] E. Parzen, Stochastic Processes, San Francisco: Holden-Day, 1962, Chap. 5.

[5] B. V. Gnedenko, Yu. K. Belyayev and A. D. Solovyev, Mathematical Methods of Reliability Theory, New York: Academic Press, 1969, Chap. 2.

[6] E. Pieruschka, Principles of Reliability, Englewood Cliffs, New Jersey: Prentice-Hall, 1963, Chap. 7.

[7] R. E. Barlow and F. Proschan, Mathematical Theory of Reliability, New York: Wiley, 1965, Chap. 3.

[8] A. Papoulis, Probability, Random Variables, and Stochastic Processes," New York: McGraw-Hill, 1965.

DETECTING AN UNKNOWN SIGNAL IN A MULTIPLE OBJECT, TELEMETRY SITUATION

John Bart Wilburn, Jr.
Instrumentation and Methodology Branch
US Army Electronic Proving Ground
Ft. Huachuca, AZ 85613

ABSTRACT. The problem is that of detecting anomalie patterns in environmental grid data approximately coincident with a point stimulus in the region including all data sources.

The particular case involved is to replace the current, rather awkward, technique with a more concise and efficient alogorithm for detecting anomalous growth patterns of tree-ring chronologies approximately coincident with volcanic eruptions.

STATEMENT OF THE PROBLEM: The problem I am presenting here is a problem arising in my climatology research on estimating climatic anomalies following volcanic eruptions. People have long suspected that such anomalies would occur. (Franklin, 1783 Diary) It seems as no surprise to most people that something as majestic as a volcano should perturb climate and yet compelling evidence has not been found, probably due to the short length of meteorological data records available and/or improper methods of analysis.

I am estimating these climatic anomalies by computing a regression model for climatic variables such as seasonal temperature and precipitation averages based on tree-ring chronologies. In this way I am hoping to attach to a much longer record of data. The regression model is a principal component regression calculation which I discussed at this conference last year; and uses continuous tree-ring chronologies and a concurrent meteorological record taken at, or near, the tree site for which the model is computed. That is, for each tree site there is one model for each climatic variable for each season.

With these models, or transfer functions, I estimate the climatic anomalies following volcanic eruptions by applying anomalous sequences of annual tree growth rings following those eruptions as input to the transfer function.

The problem I am presenting here is how to improve the accuracy of the detection of anomalous tree growth due - probably - to volcanic activity and to perform the detection more economically.

This may not seem related to telemetry in the usual sense; however, I contend that it is, or has within it, a problem in multiple object telemetry. In this case, telemetry is interpreted as the receipt of a signal transmitted by a sensor operating in an environment wherein the

signal is supposed to contain information about it's environment.

In my case, the sensor is the tree. The signal is the chronology of it's annual growth rings. These growth rings differ in width in response to climatic conditions present at the site. Figure I illustrates a section of a chronology and a graph of the ring widths. As one can see, this signal looks very much like many other kinds of signals one may encounter in a telemetry operation.

The signal is supposed to contain information about the climatic conditions at the tree site during the time that the growth ring was influenced. A considerable amount of work done, and currently underway, at the Laboratory of Tree Ring Research at the University of Arizona supports this supposition. The problem is that not all tree ring chronologies are indicative of climate. Only sensitive trees have chronologies which reflect their past climate and then only when properly interpreted.

There are many factors which influence a tree's response to a particular climatic variable. Topography is the primary class of these factors which include: water runoff, exposure (north or shady side versus south or sunny side), altitude (growth season), subsurface conditions influencing root structures, availability of ground water and density of tree growth. However, these factors are, for the most part, reasonably constant over the time period considered; that is, a few hundred years. Thus, the sensitivity of a tree to climatic change can be considered to be reasonably constant except when it is obviously not true as in cases such as fire, earthquake, etc. Figure II illustrates these opposite conditions, complacent and sensitive trees, as a function of topography.

A sample illustration of this sensitivity is shown when we consider a tree which is living in an abundant environment (as seen by the tree) with a surplus of water. This tree would have a "complacent" ring series because such a tree will not suffer much, if at all, during a relatively dry growing season with less, but still adequate, precipitation. However, a farmer in the same area with a crop tuned to the normal precipitation (abundant from the tree's point of view) might consider that dry spell a near disaster. This complacency is compounded when one notes that most trees tend to integrate over several years with the emphasis placed on the climate of the year preceding the current growing season.

The point is that one may see that a given species of tree may have many different responses to highly similar climates, depending on the specific locations of the trees and the conditions preceding the current growing season of up to three years.

Now it is possible to see the nature of the problem I am addressing. As shown in Fig. III, I have selected, as sensors, ten tree sites; all Douglas Fir and all with fairly high variance in the chronology as an indication of sensitivity. These ten tree sites, indicated by the dots,

constitute a grid of climatic sensors, each of which has a response function defined only for it's own location, but, which has been assumed to be reasonably time invarient.

Now the problem becomes somewhat more complicated. This is because I am looking for the result of an unknown, but probably different response function to the output from another response function, which is the atmosphere, also unknown and responding to a point stimulus (the volcanic eruption). It is the nature of this atmospheric response function that I would like to eventually learn something about from the regression-based estimates of the climatic anomalies mentioned earlier.

The response of the atmosphere to this stimulus at some location on the earth is, most likely, some function of: the type of stimulus; that is large, small, duration, etc; the location of the tree site (sensor); the time lag from the eruption; the time of the year and the initial conditions at the time of the year.

The response function of the trees to the atmospheric (climatic) conditions is some function of: the season; it's own serial correlation; it's initial condition and it's location (topography). The response function of the trees omits the physiological variables as I am considering them as explicit since I am not modeling the tree growth.

The first part of the project, which is the subject of this paper, was to detect the anomalous, indirect response, if any exists, of the trees to volcanic eruptions. To date, the method of detecting these possible anomalous sequences of growth rings, or anomalous signals, has been as follows: First, I considered only one site at a time; thereby permitting me to ignore all parameters relating to location. Second, the tree integrates over all seasons; so, for the purposes of signal detection, I must ignore season. Now then, it must be remarked that the amount of change in the tree's variance due to volcanic activity may be only a very small portion of the total variance in the tree ring chronology.

Assuming that the chronology is a weakly, stationery, random series, a kind of signal averaging was accomplished to detect a possible average, or typical, response signal of the tree to specific "types" of volcanic eruptions.

The tree ring data were formed into a lagged array, as shown in Fig. IV, wherein the lag is fourteen years. The lag is more than sufficient to accommodate the serial correlation of about three years and is guessed to be sufficient time to cover any lag of the propagation of the atmospheric phenomena. This lag also side-steps two favorite cycles: lunar and solar.

The data in an array such as shown in Fig. IV contains all of the data and as such is referred to as: D^{t}_{nm}, the total ring array. A

similar array is formed from the columns of D^t such that the date of the growth ring index (percent of normal growth) in the first row of each column is the date of a volcanic eruption of a specified class of eruptions parameterized by size of eruption and the region of the earth containing the volcano. This data array is referred to as the signal array and is denoted by: D^s_{nq}.

A third array is the background array, D^b; and is the direct subtraction of D^s from D^t: $D^b = D^t \ominus D^s$.

Now then, the row averages of each of these arrays were computed. These constitute average growth curves of the tree for a fourteen-year period under: normal conditions, conditions coincident with volcanic activity of the class specified, and under conditions excluding those concurrent with that specific class of volcanic activity.

A CHI-square comparison was made with the following hypotheses:

1. That the average growth curve of the signal array, D^s, was indistinguishable from the average growth curve of the total array, D^t.

2. That the average growth curve of the signal array, D^s, was indistinguishable from that of the background array, D^b.

3. That the average growth curve of the background array, D^b, was distinguishable from that of the total array, D^t.

4. That the average growth curve of the total array, D^t, was distinguishable from the flat curve of the average of the total chronology.

If all of these hypotheses are rejected, then the average growth curve of that signal array is considered a probable, valid response to a volcanic eruption of the class specified. From about 300 cases, 35 passed this test at the 99% confidence level.

A second test was devised involving the comparison of the first eigenvectors of the variance/co-variance matrix of the ring signal array, D^s, computed two ways. The variance/co-variance matrices of the signal array were computed: (1) using the row averages of the total ring array, $\bar{d}_i^t$ as the mean; and (2) using the row averages of the ring signal array, D^s, in the usual fashion. Thus we have:

$$C^s_{nn}(\bar{d}^t) = \frac{1}{n-1}\left[(d^s_{ij} - \bar{d}^t_i)\ (d^s_{ij} - d^t_i)'\right]$$

and

$$C^s_{nn}(\bar{d}^s) = \frac{1}{n-1}\left[(d^s_{ij} - \bar{d}^s_i)\ (d^s_{ij} - \bar{d}^s_i)'\right]$$

Then extract the eigenvectors:

$$C^{s}_{nn}(\bar{d}^{t})\ E_{nn}(\bar{d}^{t}) = E_{nn}(\bar{d}^{t})\ \Lambda_{nn}(\bar{d}^{t})$$

and

$$C^{s}_{nn}(\bar{d}^{s})\ E_{nn}(\bar{d}^{s}) = E_{nn}(\bar{d}^{s})\ \Lambda_{nn}(\bar{d}^{s})$$

Next, compare $E_{in}(\bar{d}^{t})$ and $E_{in}(\bar{d}^{s})$. If they are significantly different, then the Array D^{s} is usable as an array of tree ring data comprised of significant responses. This was a very stringent test and out of the 35 candidates, only six passed.

The computer time required to perform all of these tests, for all ten sites and thirty classes of volcanic eruptions, was about ten hours on a CDC 6500. This did not include the comparison of the eigenvectors, but only their computation. Thus, the need for a new method.

Another, related, reason for initiating this work is to begin the development of a statistical description of tree growth which will contain information about both the spatial relationships of the tree sites; and, simultaneously, the temporal behavior of the individual tree sites and the interrelationship between the two descriptions of the tree growth.

One of the approaches to this problem I have started is to devise an entropy function for each column of the total array.

$$H^{l}_{j} = \sum_{i} P^{l}_{ii} \text{ lag } P^{l}_{ij}$$

l = tree site location

i = row

j = column

where P_{ij} is computed using the statistics of the chronology.

The intent was to detect a departure from normal growth during the fourteen year period following any year. The data array, D^{t}, would then be collapsed into a one dimensional sequence of entropy values for each tree site. These data streams could then be considered as variables indexed by location and analyzed by multivariate techniques for the time invarient relationship of the time lagged behavior between each site. Furthermore,

by computing a conditional entropy, the serial correlation of the trees could be accounted for.

In this way, it is hoped that those tree sites with large and/or correlated variance of abnormal behavior will be selected by eigenvector analysis.

Another variation of this method would be to form a lagged array from one of the principal components of a spatial array of tree ring chronologies sampling an entire region. Then, to perform the entropy calculation of that lagged array. This would highlight abnormal growth occurring simultaneously throughout the region.

Now, I would like to hear any comments and suggestions the panel might wish to make.

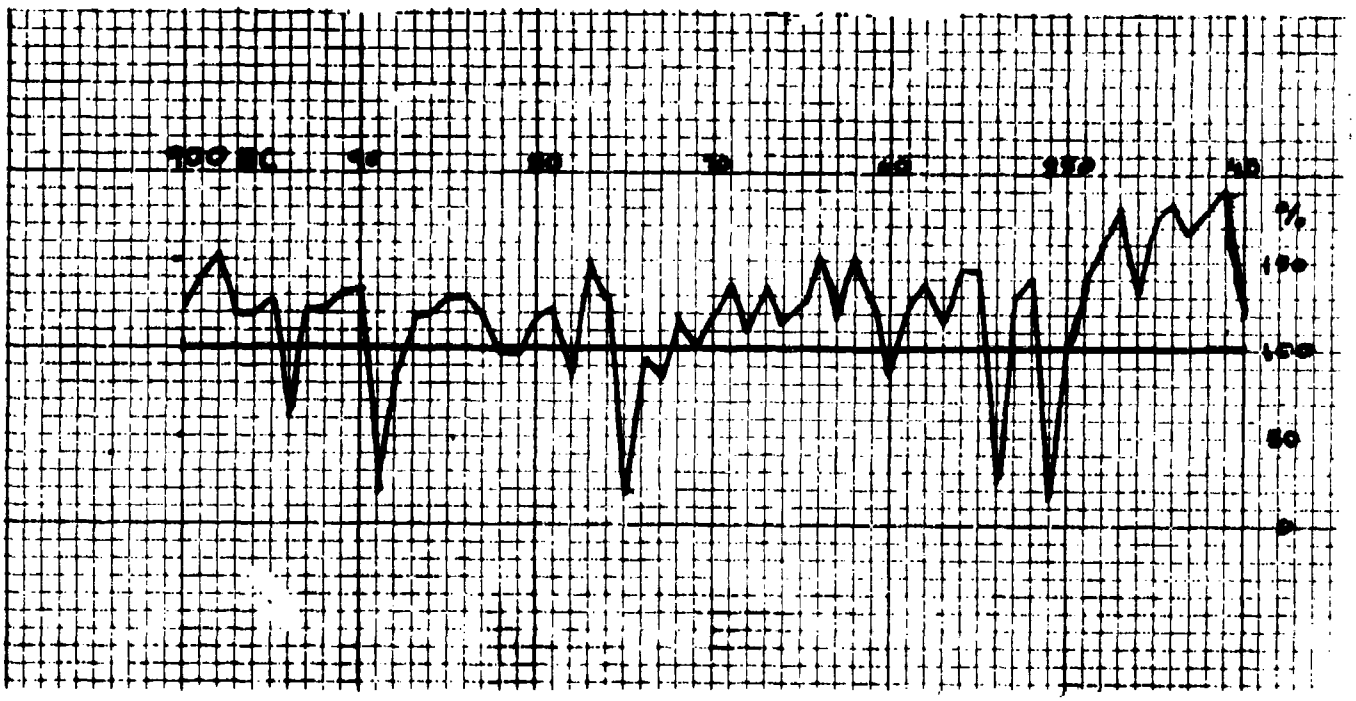


Figure 1. A segment of the bristlecone pine master chronology, representing three trees from 900 to 840 B.C.

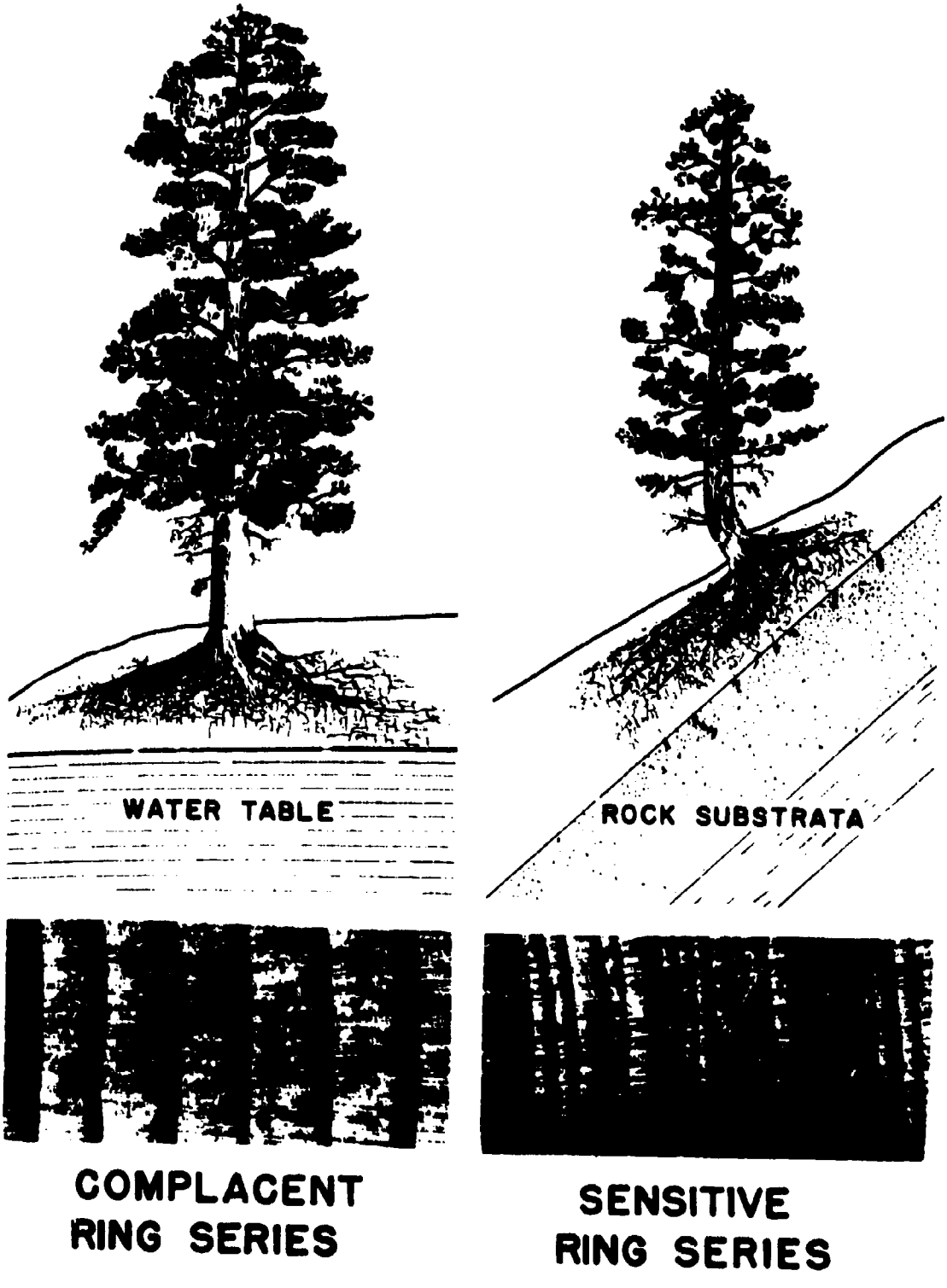

WATER TABLE
ROCK SUBSTRATA
COMPLACENT
RING SERIES
SENSITIVE
RING SERIES

Fig. II

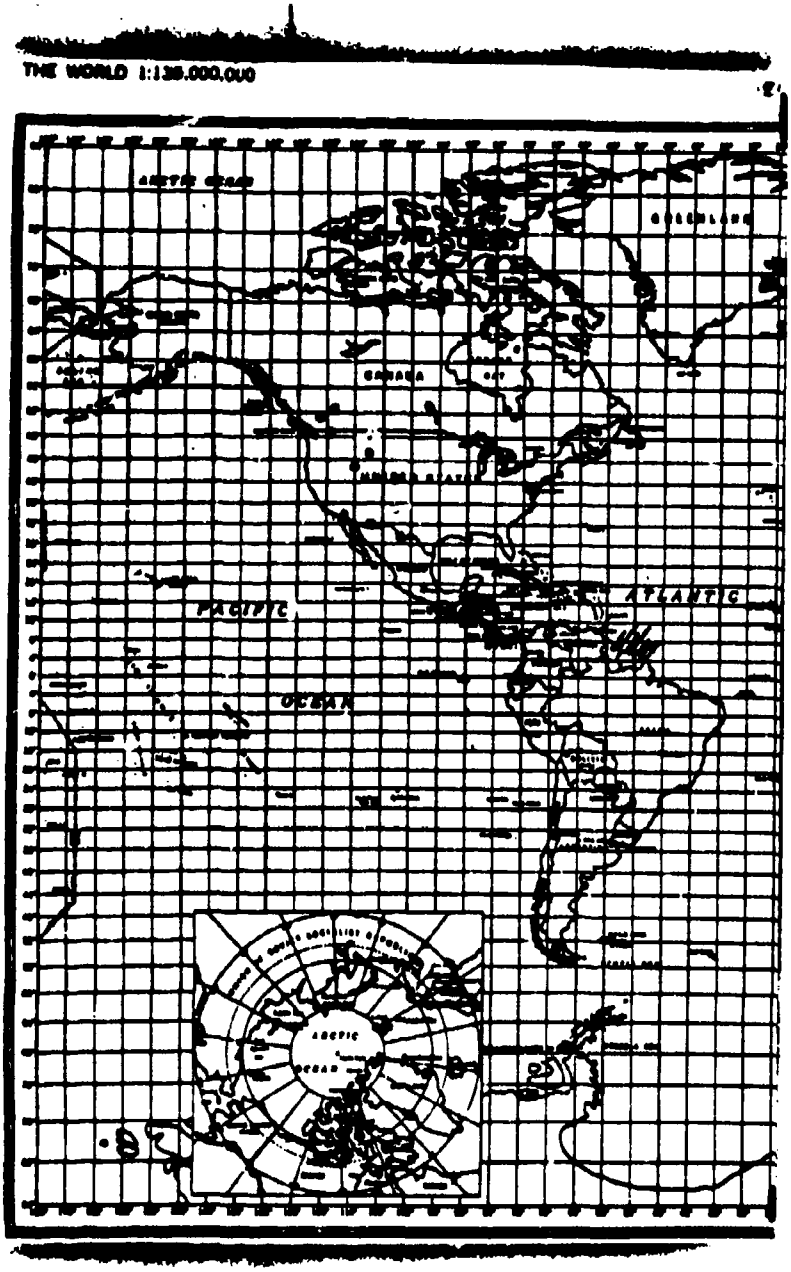


Fig. 111

Distribution of Ten Tree Sites
(Sensors) In North America

14-YEAR LAGGED ARRAY

D_{nq} formed from data stream m long

Columns

Rows	q = 1	2	3	. . .	m-14	m-13
n = 1	d_1	d_2	d_3	. . .	d_{m-14}	d_{m-13}
2	d_2	d_3	d_4	. . .	d_{m-13}	d_{m-12}
3	d_3	d_4	d_5	. . .	d_{m-12}	d_{m-11}
. . .						
13	d_{13}	d_{14}	d_{15}	. . .	d_{m-2}	d_{m-1}
14	d_{14}	d_{15}	d_{16}	. . .	d_{m-1}	d_m

Subscripts are the sequence number of the data in the original sequence m long.

The d_{ij} of D_{nq} are the tree ring indicies, as computed from a given chronology by the Laboratory of Tree Ring Research at the University of Arizona, corresponding to specific years of growth.

FIG IV

OUTLIER DETECTION PROCEDURES IN TRAJECTORY DATA REDUCTION

William S. Agee and Robert H. Turner
Analysis and Computation Division
National Range Operations Directorate
US Army White Sands Missile Range
White Sands Missile Range, New Mexico

ABSTRACT. Outlier detection procedures are used extensively in trajectory data reduction at White Sands Missile Range (WSMR). There are three distinct circumstances in which outlier detection procedures are used in trajectory data reduction. These are recursive filtering, weighted least squares batch processing of trajectory measurements, and unweighted least squares processing. Each of these processes use a different outlier detection procedure. This paper describes the use of outlier detection procedures at WSMR, the specific procedures used in the various data reduction processes, and the limits within which each of the procedures performs satisfactorily. Of prime concern are the situations in which the outlier detection procedures fail to detect some obvious outliers. These undetected outliers destroy automated data reduction procedures causing a significant number of reruns with human detection of these outliers. The performance of various outlier detection procedures, those currently used at WSMR and some others is shown on typical data sets for which the procedures fail. It is hoped that, in addition to obtaining some suggestions on improving outlier detection used in WSMR data reduction, this presentation will stimulate further investigation into outlier detection methods by Army researchers.

1. INTRODUCTION. Some outlier detection techniques for batch and recursive processors which produce trajectory estimates from instrumentation measurements are described.

Although there are some outlier detectors in the batch processor, a pre-processor is necessary to eliminate those outliers which could ruin the batch process beyond recovery. This pre-processor removes the trend using an unweighted least squares process and detects outliers using two tests. A better way of removing the trend is necessary when some types of outliers are present. Also, since some types of outliers produce a masking effect which makes sequential procedures insensitive, other tests are needed. The outlier detectors are good in the batch processor and very good in the recursive processor.

2. PRE-PROCESSOR

a. Process. Small samples (one to four seconds) of 10 to 50 measurements of each observation are fit to a second degree polynomial in time using unweighted least squares.

The observation model is

$$z_i = a_0 + a_1 t_i + a_2 t_i^2 + \epsilon_i \qquad i = 1, n$$

or

$$Z = TA + \epsilon$$

where ϵ is random noise with zero mean and σ^2 variance.

Minimizing $\epsilon^T \epsilon$ with respect to A we have

$$\hat{A} = (T^T T)^{-1} T^T Z$$

and the set of residuals

$$r = Z - TA$$

b. Outlier Detection. Sample skewness and kurtosis coefficients are computed from the residuals

$$\sqrt{b_1} = \sqrt{n-3} \sum_{i=1}^{n} r_i^3 / \left(\sum_{i=1}^{n} r_i^2 \right)^{3/2}$$

$$b_2 = (n-3) \sum_{i=1}^{n} r_1^4 / \left(\sum_{i=1}^{n} r_1^2 \right)^2$$

If either $\sqrt{b_1}$ or b_2 exceed their respective 5% significance level critical values, the observation corresponding to the largest residual is deleted and the entire process is repeated with the remaining observations.

We hope that this initial process will detect most of the outliers automatically with as little human intervention as possible and a minimum of false alarms. When there are too many outliers or a few large ones it is almost impossible to detect them. In these cases, if the presence of an outlier is detected, the good observations adjacent to the outliers are the ones rejected.

c. Examples. These two samples show that the presence of outliers can sometimes distort a curve fit so much that outliers cannot be detected. Furthermore, if the presence of outliers were detected, sometimes

the good observations are rejected while the outliers remain. Each sample has three obvious outliers which were not detected from the first set of residuals.

(1) Example 1. Assume some other test could detect the presence of outliers and that the observation with the largest residual was rejected. One of the outliers would be rejected. The two previously described tests and rejection criteria would now sequentially detect and reject the two remaining outliers.

(2) Data for Example 1.

Obs	Res(1)	Res(2)	Res(3)	Res(4)
.21709	-.33222	-.29484	-.20135	-.00001
.21824	-.31419	-.26636	-.17482	.00001
.95519	.44164	.49745	.58588	
.94511	.45245	.51376		
.93499	.46522			
.22288	-.22199	-.15714	-.08487	.00001
.22405	-.19391	-.13101	-.06642	-.00002
.22530	-.16375	-.10528	-.04951	.00002
.22652	-.13161	-.08006	-.03424	.00002
.22770	-.09751	-.05535	-.02063	-.00004
.22900	-.06128	-.03100	-.00852	.00000
.23028	-.02307	-.00715	.00195	.00001
.23155	-.01714	.01622	.01079	-.00001
.23286	.05940	.03915	.01805	.00000
.23418	.10367	.06162	.02370	.00001

(3) Example 2. Again assume that some other test could detect the presence of outliers and that the observation with the largest residual was rejected. The first point rejected would be the good observation in-between the outliers. Two outliers would be the next to go. Further application would reject good observations and never get the one remaining outlier. The outlier detectors previously described don't indicate the presence of outliers in any set of residuals.

(4) Data for Example 2.

Obs	Res(1)	Res(2)	Res(3)	Res(4)
-1.70987	-.15777	-.28786	-.36369	-.37731
-1.70942	-.00020	-.03242	-.08045	-.10634
-1.70893	.10548	.14636	-.12669	.09700
-1.70845	.15923	.24843	.25767	.23267
-1.70793	.16109	.27383	.31254	.30071
-1.70741	.11102	.22252	.29127	.30108
-1.70682	.00910	.09458	.19393	.23385
-1.70626	-.14478	-.11009	.02041	.09892
-1.70571	-.35060	-.39148	-.22927	-.10368
-1.70510	-.60828	-.74951	-.55502	-.37389
-1.70449	-.91788	-1.18425	-.95693	-.71177
1.43777	1.86223	1.44596		
1.44602	1.45641	.86545	1.16012	
-1.70257	-2.15818			
1.44667	.47314	-.54153	-.17727	.40876

d. Conclusion. More work needs to be done in

(1) Removing trends in the presence of outliers.

(2) Determining whether the testing and rejection of small subsets of observations as a one time process is more effective than the sequential application of testing and rejecting of one observation at a time.

3. BATCH PROCESSOR

a. Process. This is a weighted least squares process which uses observation variances as weights. It produces all position vector estimates simultaneously. It is a nonlinear process which linearized about a guess trajectory and is iterated to convergence before editing. The measurement model for the α^{th} observation at the i^{th} time point is

$$Z_{i\alpha} = h_\alpha(x_i) + \varepsilon_{i\alpha}$$

where $\varepsilon_{i\alpha}$ is random noise with zero mean and $\sigma^2_{i\alpha}$ variance.

Solve for $\hat{x}$ by minimizing the weighted sum of squares

$$\sum_{i=1}^{m} \sum_{\alpha \varepsilon I_i} \left(\frac{Z_{i\alpha} - h_\alpha(x_i)}{\sigma_{i\alpha}} \right)^2$$

with respect to x_i.

b. Outlier Detection

(1) At each time point i, for each observation α in the solution a normalized residual is computed

$$r^*_{i\alpha} = \frac{Z_{i\alpha} - h_\alpha(\hat{x}_i)}{\sigma(r_{i\alpha})}$$

where $\sigma^2(r_{i\alpha})$ is the estimated residual variance approximated by

$$\sigma^2(r_{i\alpha}) \approx \sigma^2_{i\alpha} + H_\alpha (H^T W H)^{-1} H^T_\alpha$$

$$H_\alpha = \frac{\partial h_\alpha(\hat{x}_i)}{\partial \hat{x}_i}$$

$$H^T W H = \sum_{\alpha \varepsilon I_i} \frac{H^T_\alpha H_\alpha}{\sigma^2_{i\alpha}}$$

If $3 < |r^*_{i\alpha}| < 5$, the respective observation is deleted temporarily.

If $|r^*_{i\alpha}| > 5$, the respective observation is deleted permanently.

If either of these tests reject any observations the solution is iterated to convergence with the remaining observations and tested again.

This test indicates those observations whose residuals are not consistent with their variance and geometry.

(2) When no more observations are rejected with the previous test, a sum of weighted residuals for each observation, over all the time points it was processed is computed.

$$R_\alpha = \sum_{i|\alpha\epsilon I_i} r^*_{i\alpha}$$

If max $|R_\alpha|>3$, all of the α^{th} observations are deleted from all further processing, all temporarily deleted observations are enabled and the whole process is reiterated. This test indicates a consistent bias in an instrument's set of observations.

4. RECURSIVE PROCESSOR

a. Process. This is an extended Kalman filter which produces state vector (position, velocity, acceleration) estimates sequentially. Observation variance estimates are also produced sequentially. The predicted state estimate is

$$\hat{x}(k+1|k) = F(k)\hat{x}(k)$$

the corrected state estimate is

$$\hat{x}(k+1) = \hat{x}(k+1|k) + K(k)r(k+1|k)$$

where K(k) is the Kalman filter optimal gain matrix and

$$r(k+1|k) = Z(k+1) - h(k+1)\hat{x}(k+1|k)$$

is the vector of observation residuals.

The variance estimate

$$\sigma_i^2(k+1) = \frac{1-w^2}{2w} Q_i(k+1)$$

is a steady state function of the exponentially weighted sum of squared residuals

$$Q_i(k+1) = w[Q_i(k)+r_i^2(k+1|k)]$$

$$0<w\leq 1$$

b. Outlier Detection. For each observation i at time k+1, a two-level outlier detection scheme is used on the normalized residual

$$r_i^*(k+1|k) = \left|\frac{r_i(k+1|k)}{\sigma(r_i)}\right|$$

$$\sigma^2(r_i) = \sigma_i^2(k) + H_i P H_i^T$$

$$H_i = \frac{\partial h_i(x)}{\partial x}$$

P is the state covariance matrix.

(1) If $r_i^*(k+1|k)>12$ reject the i^{th} observation for time k+1.

(2) If $4<r_i^*(k+1|k)<12$ update $Q_i(k+1)$.

(3) If $0<r_i^*(k+1|k)<4$ update $Q_i(k+1)$, $\sigma_i^2(k+1)$ and $\hat{x}(k+1|k)$.

5. REFERENCES

a. Agee, W.S. and Turner, R.H. (1975). "Dovap Best Estimate of Trajectory", Technical Report No. 55, WSMR.

b. Agee, W.S. and Turner, R.H. (1972). "Optimal Estimation of Measurement Bias", Technical Report No. 41, WSMR.

c. Ferguson, T.S. (1961). "On the Rejection of Outliers", Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, ed J. Neyman, pp 253-287 Berkeley: University of California Press.

d. Grubbs, F.E. (1969). "Procedures for Detecting Outlying Observations in Samples", Technometrics 11, 1-21.

APPLYING SIMULATION OF PHYSIOLOGICAL SYSTEMS TO THE DESIGN OF EXPERIMENTS: EXAMPLES OF ENDOCRINE AND RESPIRATORY FUNCTION

Stanley M Finkelstein
Division of Bioengineering and
Department of Operations Research and System Analysis
Polytechnic Institute of New York
Brooklyn, New York

and

Stanley S. Reisman
Department of Electrical Engineering
New Jersey Institute of Technology
Newark, New Jersey

ABSTRACT. The development of simulations of physiological systems has been used as a guide in the design of animal experimentation used to study such endocrine functions as glucose-insulin interaction and testosterone dynamics. Models of pulmonary respiratory function have been studied in an effort to redesign several pulmonary function tests so that particular system parameters could be evaluated directly from test results.

Model development is thus a useful procedure in studying physiological systems, for it focuses attention on the cause-effect relationship at each stage of the homeostatic process, and thus integrates in a systematic way all that is known about a particular system. In addition, the requirements and constraints of the model development clearly point out gaps in our knowledge of overall system function, and in an effort to obtain this missing data one can utilize the model structure in designing the necessary experimental protocols. The results of these experiments will help complete the model in a physiological meaningful way, and once complete, the model can be used to study the effects of parameter variation on system response under both normal and pathological situations.

The simulation can be used in conjunction with, and as a supplement to, animal experimentation. For example, the large number of extraneous, and possibly even unknown, factors which often obscure or invalidate the results of live animal experiments are not present in the model. The model user must be able to take advantage of the resulting simplified approach to the physiological system, but must, at the same time, be careful not to oversimplify the complex physical interrelationships to the point at which the results are physiologically meaningless.

This presentation will utilize several case studies to demonstrate the use of model development in designing experiments to study overall system function, subsystem operation and compartment analysis, and parameter evaluation.

1. INTRODUCTION. Model development is a useful procedure in studying physiological control systems, for it focuses attention on the cause-effect relationship at each step of the control process, and integrates in a systematic way all that is presently known about the particular system. Models can be presented in many different modes, some of which might be scaled versions of the actual system, physical analogs consisting of hardware elements or alternative living systems, and both analog or digital computer simulations. The emphasis in this presentation, however, will be on the mathematical descriptions of system function and the computer simulations of these relationships. In particular, the application of models in research, teaching, and the design of experiments will be discussed in terms of specific examples of endocrine and respiratory function.

Early application of the control engineer's approach to physiological system studies appeared in the work of Grodins[a] and Stark[b] in their studies of respiratory function and pupillary motion, respectively (1,2). Grodins' first model of respiratory function divided the body into two compartments, the lungs and the remaining tissue. In addition, he assumed that control of respiration was purely a function of carbon dioxide concentration at particular sites within the circulation. Circulation time was also assumed to be negligible. Validation studies were then performed on the model, at which time model results were compared with known experimental results from a living system. Deviations between the model and the living system suggested several additions to the model, which Grodins incorporated in subsequent more complex representations. A second model included circulation time as a non-negligible parameter, and added the effect of alveolar dead space to the two-compartment study. This more advanced model was able to be used to study both normal respiratory function and the abnormal behavior associated with Cheyne-Stokes breathing[c]. A third model added the brain compartment to the original structure, and also included the effect of oxygen concentration on respiratory control. The Grodins models illustrate one approach of model building, which begins with a simple, but non-trivial, model and adds additional complexity to make the model results agree with the results of physiological experimentation.

Stark, on the other hand, used the modeling approach in designing his experimental protocol to study pupillary diameter as a function of light incident to the eye. He used a qualitative description of the system to develop a block diagram representing the functional portions of the pupillary control mechanism. Available data could then be used to describe quantitatively the overall closed loop system, but it could not be used to develop the mathematical relationships between the subsystem variables within the closed loop. Stark then designed an experiment which would produce the necessary information on open loop response in an in vivo, physiologically undisturbed human subject. Incident light was focused at the plane of the iris so that the cross section of light entering the eye was less than the smallest pupil diameter. Incident light intensity and pupil response were

a. first published in 1954
b. first published in 1959
c. Cheyne-Stokes breathing: periodic increase and decrease in depth of breathing (tidal volume)

then recorded with an infrared electro-optical arrangement, from which frequency response curves could be developed. Transfer functions for the open loop system were then constructed and a mathematical description of the overall system was thus determined. Stark thus used a modeling approach to describe the information flow through the system, and to see how available data could be used to quantitatively describe system function. When such descriptions could not be developed, the structure and suggested cause-effect pathways within the model could be used to aid in the design of an experiment which would produce the specific information necessary for system quantification. Although this procedure was satisfactory in the case of pupillary dynamics, it is not always possible to satisfy model requirements within physiological constraints. However, the modeling approach does, as a minimum, suggest guidelines for experimental design which would result in the necessary input-output analytical relationships between system variables.

2. APPLICATION OF MODELS. Models of physiological systems have been used in research, teaching, and the design of experiments. There are two distinct steps involved in applying the modeling approach to experimental design. In developing the model, areas where the available data are not adequate to explain the operation of the system will become clarified, and a study of the flow of information necessary to completely implement the model will suggest tests and experimental procedures for generation of additional data. Such an example was discussed previously in the description of Stark's work. Then, once the model has been developed, it may offer a desirable alternative to living system experiments, where preparation time may be many hours, months, or days, and where surgical or chemical intervention may cause undesirable side effects. Such experiments can be implemented on the model, generally with little difficulty and little loss of time. The model can be used to "zero in" on a best experimental protocol, saving the animal experimentation for the final stages of exploration. Thus the model does not replace the need for animal experiments to finally validate methods and conclusions, but simply serves as a "short cut" to the final procedure, providing an easier, less expensive, and less time consuming alternative in the overall investigation.

The model can also be used to predict the effect of system changes and system sensitivities to structural and component changes. Using the model, it is a relatively simple matter to propose parameter alterations, and to observe the relative significance of these changes on the operating characteristics of the total system, as well as the sensitivity of the system to these changes. This is possible even for variables and parameters which cannot be observed directly in the physiological environment. This capability has important research and clinical applications, since it can provide a means for evaluating the probability of existence of various pathological states and may possibly suggest the etiology of a particular disease.

The physiological model can also serve as an effective adjunct in the training of bioengineers and medical scientists. The model can present problems in physiological dynamics in terms of cause-and-effect relationships between functioning parts of the system and total system operation. For example, it can be used to study the response of pathological states to various treatments. One important attribute of such a model is that a "patient" can be constructed with any desired

pathological condition, and the student can be exposed to this patient in much the same way as he would explore a clinical case. Thus the student can investigate many varieties of disease states, propose and validate a host of possible treatment protocols, and develop conceptual information about pathological dynamics, all in a single model of the physiological system of interest. At present, however, such computerized models of physiological system dynamics are not generally available, but tutorial, inquiry-response and steady-state simulations are available and finding growing acceptance in the educational community.

3. DEVELOPMENT OF MODELS. The development of a model can be broken down into four phases. These are block diagram formulation, data collection, mathematical description of the data, and computer simulation. The first step is the development of a block diagram based on the known physical principles of the system operation. This diagram should display the important characteristics of the system. This diagram may be too complex for initial simulation since it will probably include secondary functions which are not critical to overall performance. In addition, the diagram may contain physiological variables whose quantitative relationships are either not available in the literature or are extremely difficult, if not impossible, to determine by physiological experimentation. Therefore a revised "simplified" block diagram must be developed. This is generally a qualitative description of system behavior, and at this point quantitative relationships must be obtained.

Physiological experiments must now be performed in order to derive dynamic input-output relationships for each block of the model, unless these data are already available from prior work. Static characteristics may provide useful information for model development, but they cannot provide the information necessary for a complete description of system behavior. The design of the experiments should consider the particular subject (e.g., human, dog, rat, etc.), observation times based on system response times, quality and availability of data analysis and processing techniques (e.g., chemical assays), effect of the procedures on altering system physiology (e.g., surgical and chemical intervention), and overall cost of the procedure. Thus the block diagram model acts as a guide in designing the physiological experiments.

In order to use the experimental data, a mathematical description of the data must be obtained. These may be functions of time when considering system dynamics. If, for example, the blocks of the model are assumed to represent linear subsystems or linearized approximations to non-linear operation, the final mathematical representation for each block will be a transfer function $T(s)=\frac{Y(s)}{X(s)}$, where Y(s) and X(s) are the Laplace Transforms of the output and input, respectively, of the block. The time-domain description of these functions may be obtained using curve-fitting techniques.

This overall mathematical structure can be simulated on an analog or digital computer as an aid in exploiting the model. Once a simulation is developed both normal and pathological cases can be investigated by changing either potentiometer settings (analog simulation) or data values (digital simulation). Both analog and digital computers have advantages and disadvantages in their application. The

analog computer is the most direct form of simulation since the basic operations such as integration and multiplication are carried out continuously in either real time or a directly scaled version of real time. The disadvantages of this form of simulation are the necessity for amplitude and time scaling, and the complexity of the wiring or patching which occurs as the order of the system increases. Digital computer implementation on either large scale machines (e.g., IBM 370) or small scale minicomputers (e.g., DEC PDP-8) is another route for computer modeling. The simulation languages available for use on these machines (CSMP, MIDAS, ISL/8) provide a direct method for simulating an analog computer on the digital computer facility without the drawbacks of patching wires or time and amplitude scaling. Disadvantages of large digital computer simulation are the general unavailability of on-line interactive operation of the simulation languages and long turn-around times. Using a minicomputer can avoid these difficulties, but limited computer availability may be a problem. However, as costs decrease and machine capability increases minicomputers are becoming more widely available in biomedical research and education facilities.

4. CASE STUDIES. Three case studies will be presented to demonstrate the use of model development in designing experiments to study overall system function, subsystem operation, and parameter evaluation. In particular, the glucose-insulin and testosterone endocrine systems, and the respiratory system will be discussed.

4A. GLUCOSE-INSULIN HOMEOSTASIS. The development of the glucose-insulin model demonstrates the use of modeling in the design of experiments in a situation similar to that of Stark's approach to pupillary dynamics (3,4).. The glucose homeostatic system consists of a complex interaction between subsystems regulating hormonal release, glucose storage, and glucose utilization. Each such perfusion region can be viewed as a combination of controller and plant working together to control glucose and insulin levels. The pancreas and liver may be considered primary controllers due to their function under both hypoglycemic and hyperglycemic conditions, while plant function is represented by peripheral tissue activity. A block diagram of the primary interacting mechanisms of glucose-insulin control is presented in Fig. 1.

Although a quantitative description of total system function can be obtained from overall input-output measurements (e.g., system plasma responses), a clear understanding of individual subsystem function and interaction within the intact closed loop system can only be obtained if each block is itself described quantitively. The modeling approach emphasizes this fundamental observation, and focuses one's attention on those experimental procedures which will yield the input-output data necessary for subsystem development in a dynamic sense. Total system response data is widely available in the literature. For example, fundamental glucose tolerance test results can be used to relate system glucose response to glucose input over the time base of the test. However, the data needed to describe each physiological block in the figure is not generally available. A study of the model led to the development of an experimental protocol which satisfied both modeling requirements and physiological constraints involved in monitoring system variables for glucose-insulin control. Simultaneous input and output plasma concentrations for glucose and insulin were obtained for the liver, pancreas, and periphery over a fixed time sequence following glucose and insulin stimulus, respectively. These data were used to derive mathematical functions describing input and output dynamics

for each block of the closed loop. A set of normoglycemic glucose and insulin concentration curves in response to a glucose load are shown in Fig. 2. The impulse-like glucose load drives the total system into a temporary hyperglycemic condition, which elicited a pancreatic insulin response. These experimental results indicate an overreacting pancreatic insulin output, which is mediated by hepatic insulin clearance. Glucose levels rose very rapidly throughout the system, but began to decrease as insulin levels increased. Glucose concentrations returned to normal resting levels in a decaying oscillatory pattern, as would be expected of an underdamped higher-order system.

The curves of Fig. 3 and 4 describe arterial and hepatic concentration of glucose and insulin following insulin loading. The additional parameter of elapsed time after surgery is also included in these figures. The early post-operative (2 hours after surgery) response is more sensitive and less stable than the late post-operative (between 2 and 14 days after surgery) response. Arterial glucose levels decrease almost 70% from resting levels and return more slowly in the EPO than the LPO cases. Similarly, hepatic settling time is much greater in the EPO case. It is also initially highly oscillatory, perhaps indicating a very sensitive, lightly damped system. Such differences between the EPO and LPO cases suggest a possible test for degree of recovery after surgery.

Thus, the modeling procedures have been used as a guide in the design of an experimental protocol which was used to obtain the data necessary for determining true in-vivo relationships between subsystem variables. In addition, these subsystem studies have indicated the possibility of developing additional diagnostic criteria based on dynamic glucose subsystem response.

4B. TESTOSTERONE DYNAMICS. As another example of modeling of physiologic systems, the testosterone system is considered (5,6). Testosterone, the male sex hormone, gives the male his secondary sexual characteristics such as hair distribution, skin texture and voice quality. Fig. 5 represents a complete block diagram for the testosterone control system. Testosterone is secreted by the gonads and adrenal cortex and is produced peripherally through conversion of precursors. Hypothalamus-pituitary activity provides the primary control of testosterone secretion through the action of releasing factors and the hormones FSH, LH and ACTH[d]. In conjunction with this, testosterone removal mechanisms such as tissue storage and metabolism determine blood testosterone concentration.

This block diagram contains several effects which can be considered "second order". These include FSH control, testosterone secretion and the "short feedback" pathway in which the hypothalamus secretion of releasing factors is controlled by the blood FSH and LH concentrations. As described earlier, this total qualitative model is considered too complex for use in the initial modeling effort. A simplified block diagram, shown in Fig. 6 was developed in which second order effects were eliminated.

As in the glucose-insulin case, an experimental protocol was developed to obtain mathematical descriptions of each block of the figure. As an example, to mathematically describe the testosterone disappearance block, an experiment was

d. FSH: Follicle-stimulating Hormone
LH: Luteinizing Hormone
ACTH: Adrenocorticotrophic Hormone

designed in which radioactively labelled testosterone was rapidly injected intravenously into a rat and blood samples were obtained at specific times following injection. These blood samples were analyzed for radioactivity and the resulting data is shown in Fig. 7. Since the experimental procedure limits all input excitations to small perturbations about normal circulatory steady state levels, the model can be considered to be linear. Thus, the curve of Fig. 7, which is the "step response" of the testosterone disappearance block, can be used to generate a transfer function for this subsystem. The analog simulation of this transfer function is shown in Fig. 8. Similar procedures lead to transfer functions and simulations for the other blocks of the model.

Once a working simulation is developed, experiments are performed on the model to validate its performance characteristics and to improve knowledge of system behavior. This additional information can be used to create a more refined model. If little quantitative information is available, experiments on the model may suggest physiological experiments to be performed to obtain such information. The open loop response of each block of the testosterone model compared favorably with experimental results. Closed loop tests were then performed on the model. As an example, consider exciting the model with a step of voltage at the input of the testosterone disappearance block. This corresponds physiologically to a rapid intravenous injection of testosterone at times t=0. Responses are observed at the outputs of the LH disappearance and testosterone disappearance blocks, corresponding physiologically to the blood LH and testosterone concentrations, respectively. The results are shown in Fig. 9, which displays the deviations from baseline of these curves. As can be seen, the blood testosterone level begins at the injected level and returns to baseline with some oscillation within 24 hours after injection. The blood LH concentration begins below baseline in order to compensate for the increased testosterone level. The LH concentration then returns to baseline, again with a slight oscillation, within 24 hours after injection.

These results are as expected using a qualitative knowledge of system behavior, but there are no quantitative physiological data available with which to check the results. It is therefore necessary to perform physiological experiments to generate such quantitative data.

4C. RESPIRATORY FUNCTION. A digital computer simulation of respiratory function has been developed, based on the block diagram representation of Fig. 10 (7,8,9). This diagram, unlike that of the original Grodin's model, includes all that is known about respiratory function and control, at least in a qualitative sense. Once the overall system is developed, each subsystem must be described individually, and the appropriate interaction must be included so that the combined subsystems response to a simulated physiological input such as intrapleural pressure would closely resemble those of the living system. Just as in the glucose-insulin study, the overall complex model was initially developed qualitatively, and then each subsystem was studied individually and described mathematically. Unlike the glucose case, an experimental protocol was not necessary, since each block was described from the basic physics of the system function, and the specific parameter values were

already available in the literature. Of particular interest is the interconnection of the subsystems representing respiratory mechanics, alveolar mixing of respiratory gases, and diffusion between the alveolar space and the pulmonary capillary bed.

A simplified version of the mechanics section is shown in Fig. 11. This model includes the trachea-bronchi resistive pathway and the storage compartment of the lung. Also shown in the figure is the program listing used to represent the mechanics system dynamics. The program was written in the ISL/8[e] simulation language on a DEC PDP-8E minicomputer. Typical results of this simulation are shown in Fig. 12. A more detailed model of respiratory mechanics has also been developed. It includes trachea resistance, non-linear bronchial resistances, and non-linear bronchial and lung compliances. In addition, it includes flexible airway (bronchial) tissue inertance. The effect of airway inertance on overall system function has been questioned in previous studies. The inertance parameter is not easily measured or changed in the actual living system, but it can easily be varied in the computer simulation. This was done on the detailed mechanics model and the results are shown in Fig. 13. This figure represents air flow into the lung, with inertance values as a parameter of the study. The curves indicate that inertance variation has no effect on the overall flow characteristics of curve shape and timing, but does have a small effect on maximum and minimum levels of total respiratory flow. Thus the model has been used as a subject of an experimental procedure when the actual experiment on a living system was not possible. Of course, the results and conclusions of such an experiment can only be as good as the model, and thus the validity of the model must be determined prior to such experimentation.

The diffusion model can also be used to illustrate an application of modeling to experimental design and parameter evaluation. Fig. 14 represents the general model, consisting of a single compartment lung and multicompartment pulmonary capillary bed. Unlike previous models of pulmonary gas diffusion, however, this model represents the oxygen-hemoglobin interaction within the pulmonary blood as a storage (and hence capacitance) phenomenon, and not as a diffusion resistance. This difference in model concept suggested looking at the standard laboratory test used to evaluate diffusion capacity (a resistance-like element), and to develop variations of this standard test to see if the new approach is really a reasonable one. Variations of breath-holding time in the single-breath test resulted in a diffusing capacity parameter which decreased linearly with breath holding time, when plotted on semi-log paper. This variation in "diffusing capacity" can be explained using the hemoglobin storage concept developed for this model, but is not easily explained using the original concept of diffusion resistance. Thus, a modeling approach could be used to develop improved interpretations of standard clinical tests. This is yet another application of physiological modeling.

5. CONCLUSIONS. This presentation has utilized several case studies to demonstrate the use of model development in designing experiments to study overall system function, subsystem operation, and parameter evaluations. In particular, the glucose-insulin system, testosterone system, and respiratory system were discussed.

e. ISL/8: an Interactive Simulation Language developed for the DEC PDP-8 minicomputer by Interactive Minisystems, Inc., Kennewick, Washington 99336.

REFERENCES.

1. Grodins, F.S., J.S. Gray, K.R. Schroeder, A.L. Norins, R.W. Jones, "Respiratory Responses to CO_2 Inhalation: a Theoretical Study of a Nonlinear Biological Regulator", J. Appl. Physiol. 7:283, 1954.

2. L. Stark, "Stability, Oscillations, and Noise in the Human Pupil Servomechanism", Proc. IRE 47:1925, 1959.

3. Finkelstein, S.M., M. Bleicher, S. Batthany, J. Tiefenbrun, "In-Vivo Modeling for Glucose Homeostasis", IEEE Trans. on Biomed. Eng. 22:47, 1975.

4. Tiefenbrun, J., S.M. Finkelstein, W. Shoemaker, "Glucose Homeostasis and the Glucose-Insulin Feedback System in the Postoperative State", Adv. Exp. Med. Biol. 33 209, 1973.

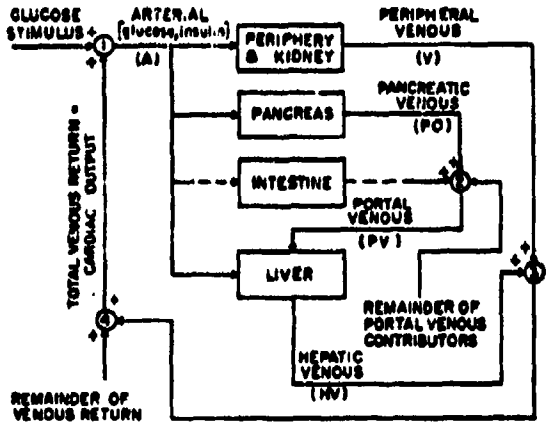
5. Reisman, S.S., "Mathematical Model of the Testosterone Control System", Ph.D. Dissertation in Bioengineering, Polytechnic Institute of New York, 1974.

6. Reisman, S.S., S.M. Finkelstein, "An Example of the Use of Computer Simulation in Physiologic Systems", Trans. Comp. Educ. Div., ASEE 7:13, 1975.

7. Finkelstein, S.M., W. Blesser, L. Braun, H. Lyons, "Simulation of the Effects of Airway Tissue Inertance on Respiratory Mechanics", Proc. 25th ACEMB 14:364, 1972.

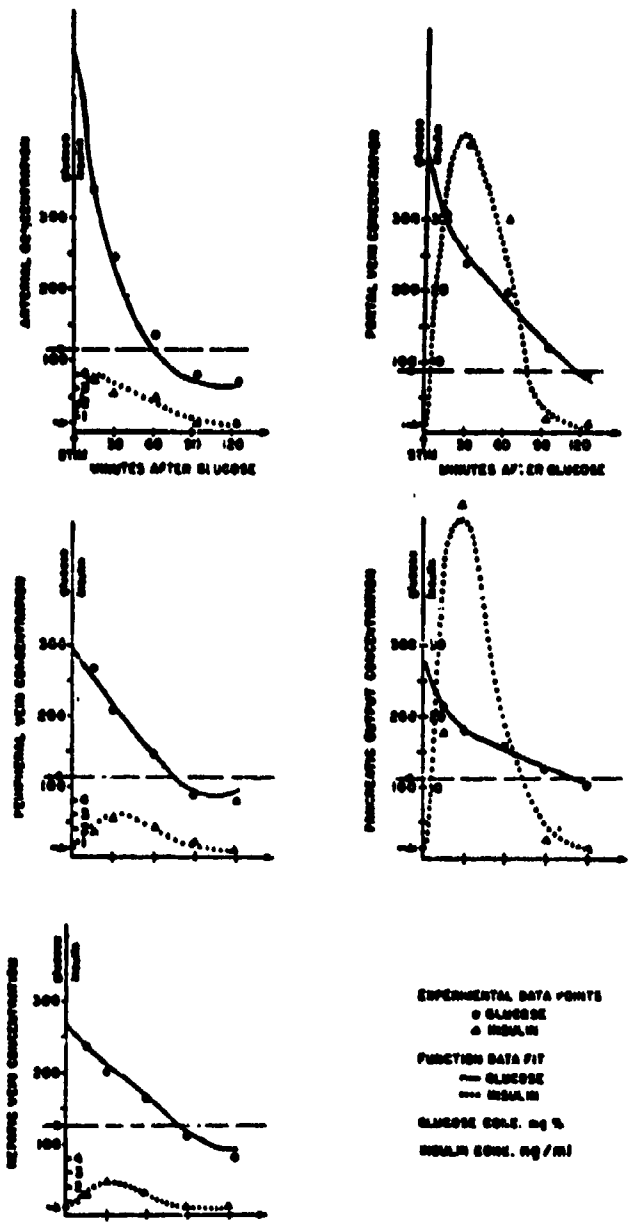
8. Braun, L., W. Blesser, S. Finkelstein, H. Lyons, "A Model of Gas Diffusion in the Pulmonary System", Proc. 26th ACEMB 26:354, 1973.

9. Albergoni, V., C. Cobelli, G. Forresin, "Interaction Model Between the Circulatory and Respiratory Systems", IEEE Trans. on Biomed. Eng. 19:108, 1972.



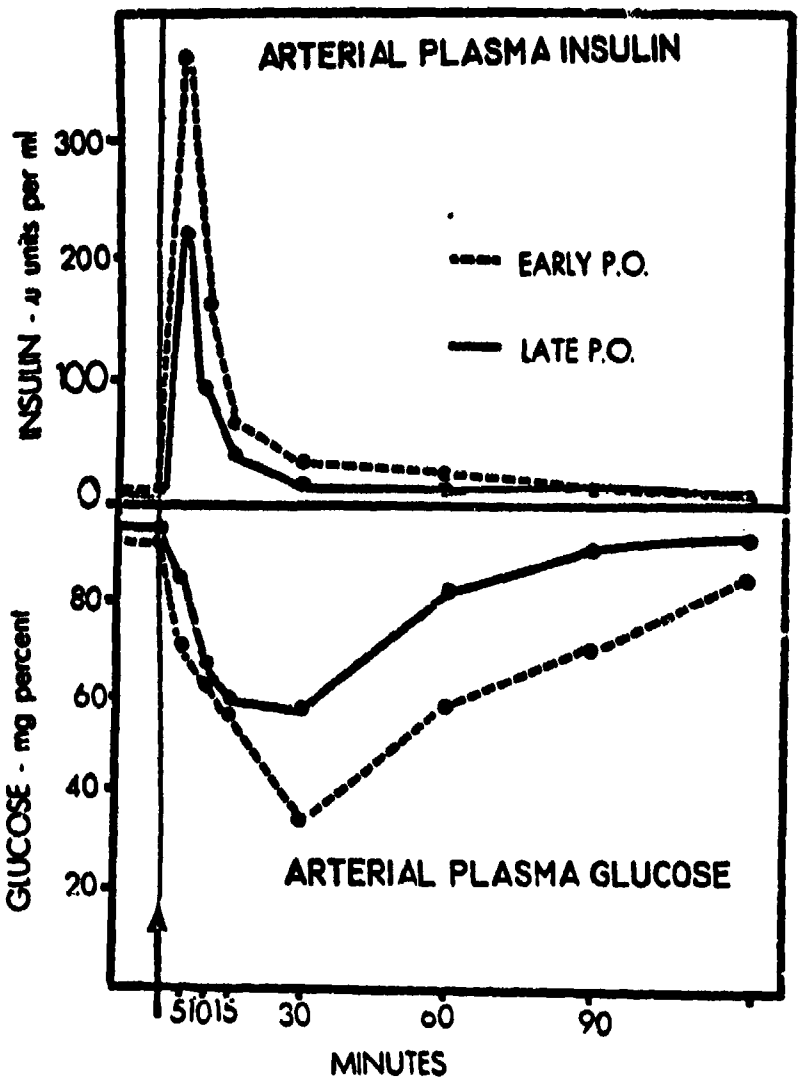
A model for glucose-insulin homeostasis.

Fig. 1.



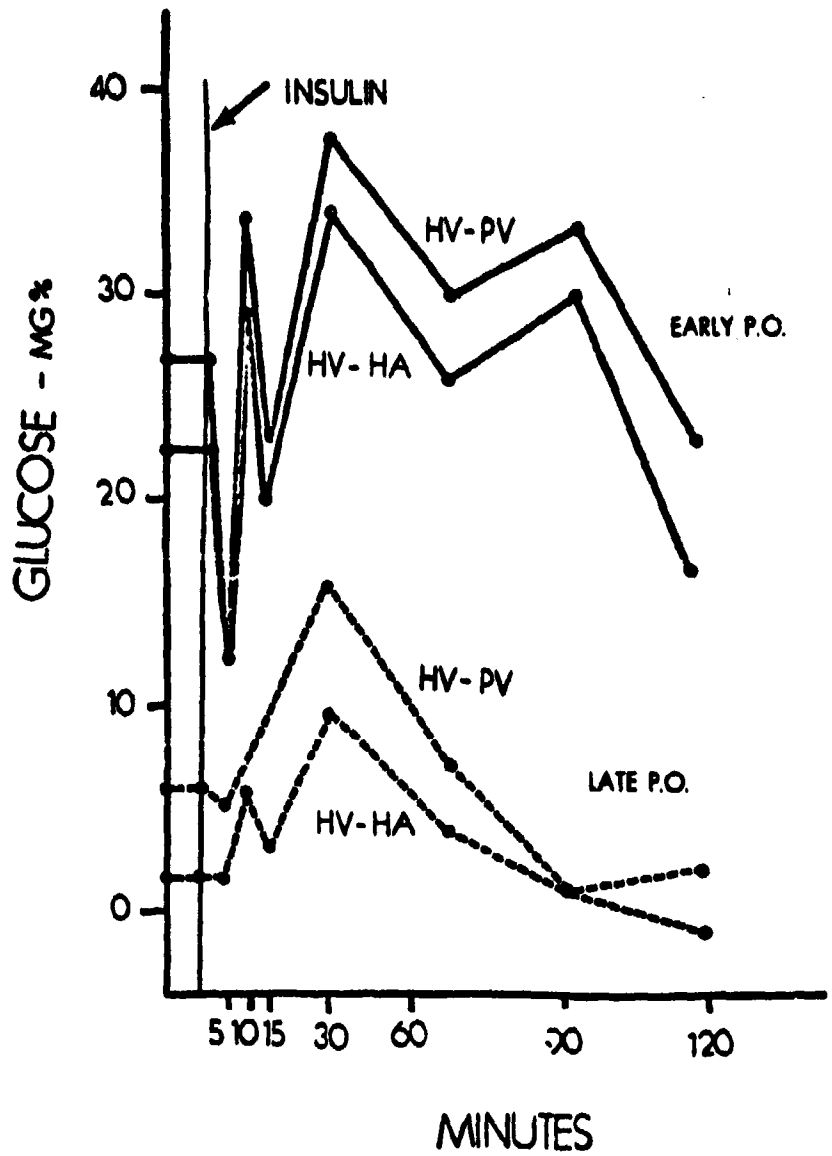
Glucose and insulin concentration dynamics for a normoglycemic subject.

Fig. 2.



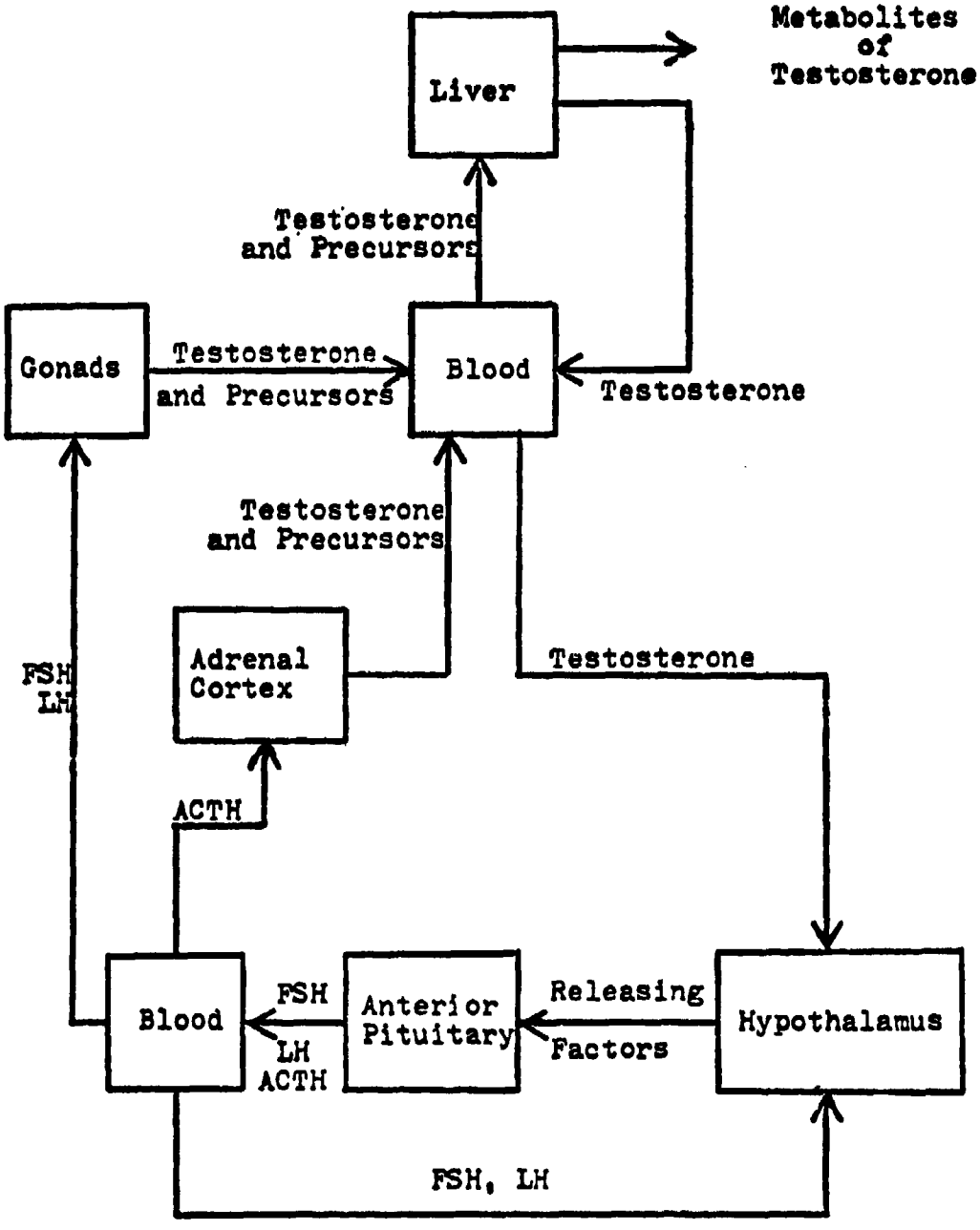
Arterial plasma glucose and insulin respose to an insulin load, for early and late postoperative studies

Fig. 3.



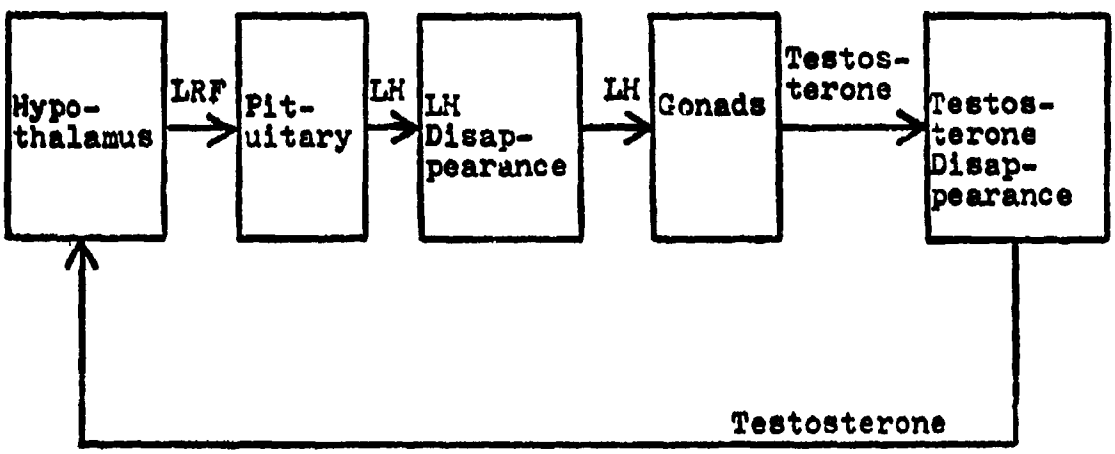
Hepatic glucose response to insulin loading for early and late post operative studies.

Fig. 4.



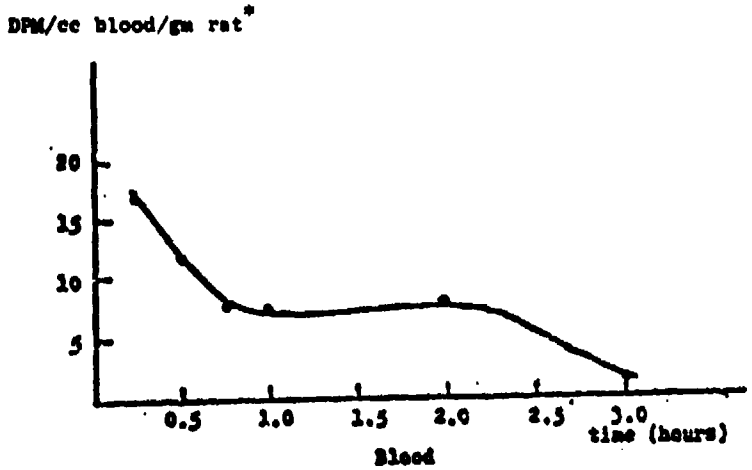
Major control pathways of testosterone concentration

Fig. 5.



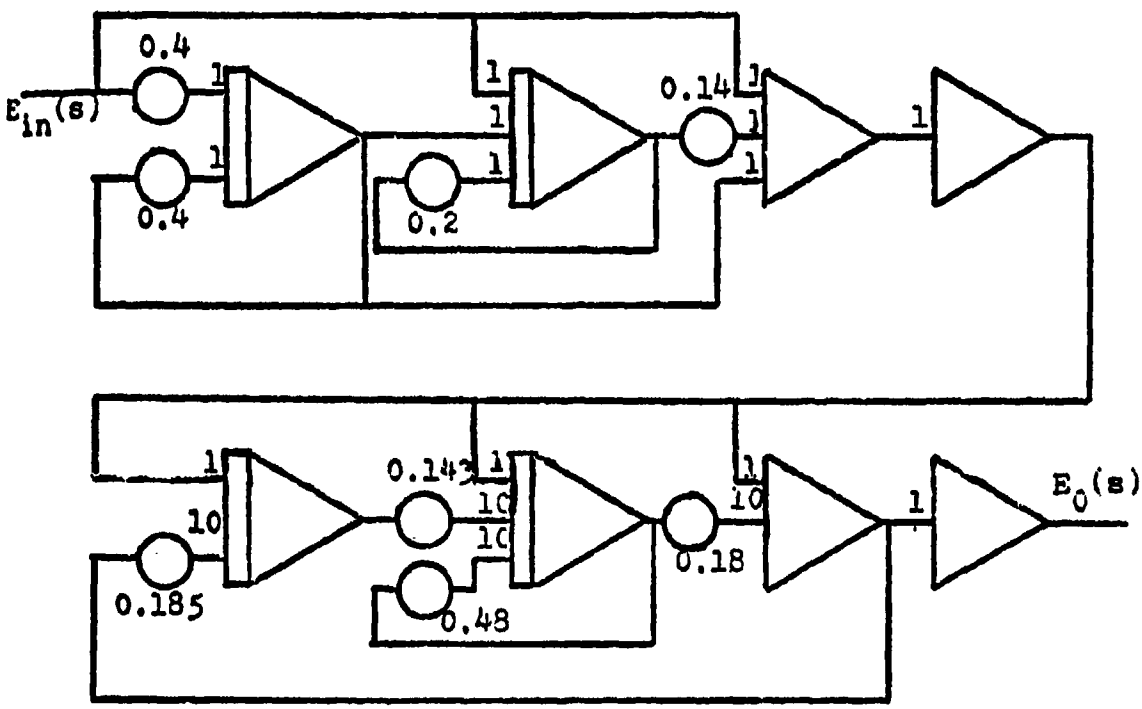
Simplified Block Diagram of
Testosterone Control System

Fig. 6.



Experimental Testosterone Uptake Results in Organs of Adult Female Rats
*All results are normalized to an injected level of one microcurie.

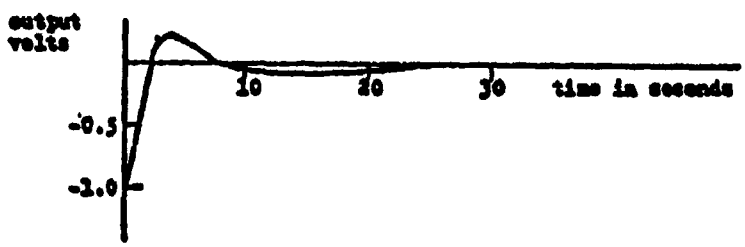
Fig. 7.



Analog Simulation for $T_T(s)$

Fig. 8.

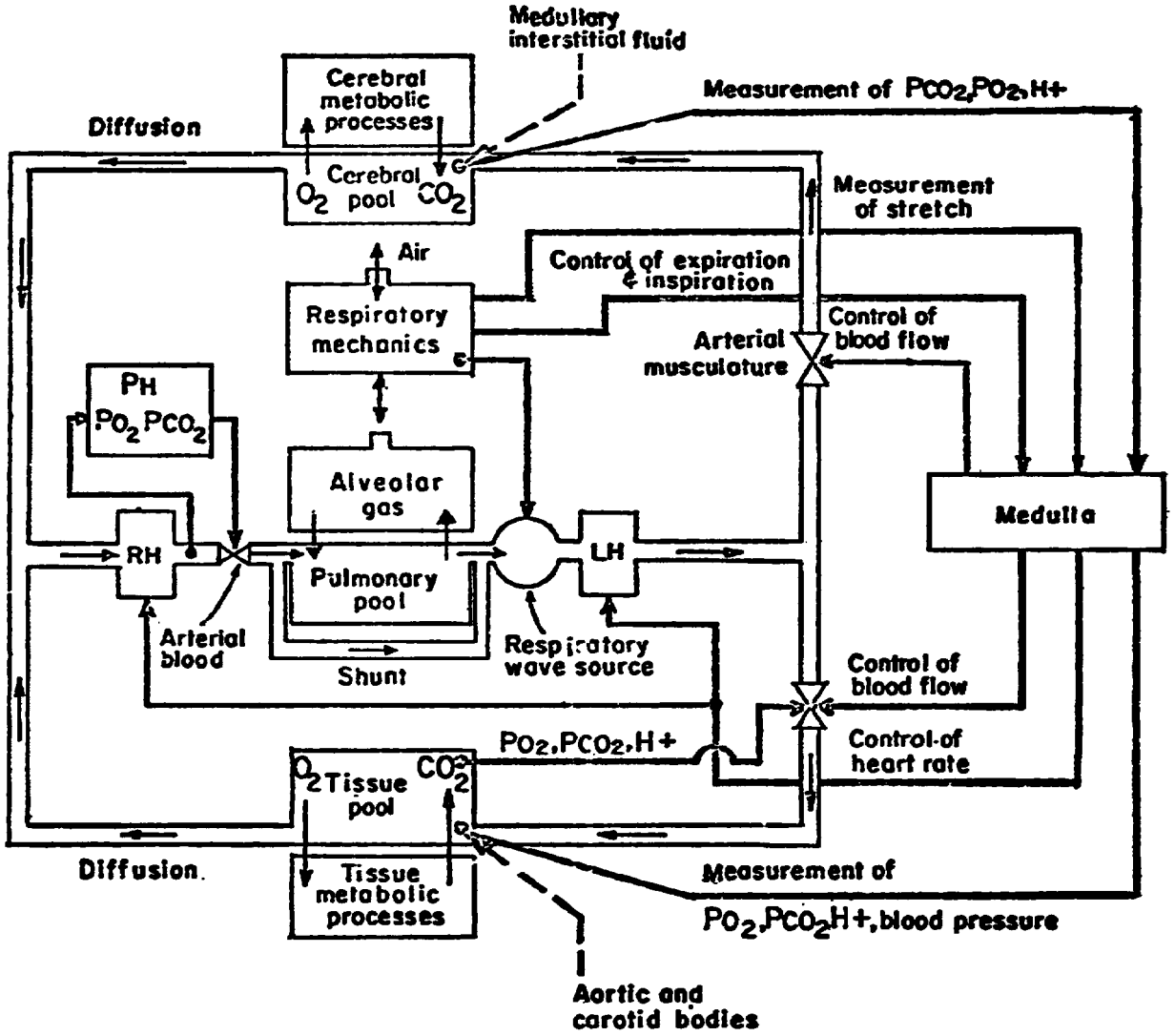
Output of Testosterone Disappearance Block
Due to Step Input of Testosterone
Note: 1 volt = 0.01 ug testosterone



Output of LH Disappearance Block Due to
Step Input of Testosterone
Note: 1 volt = 8.67×10^{-3} ng NIH-LH-S11/ml blood

Responses of the Control System to a Rapid
Intravenous Injection of Testosterone
Note: One second of computer time is equivalent
to one hour of system time.

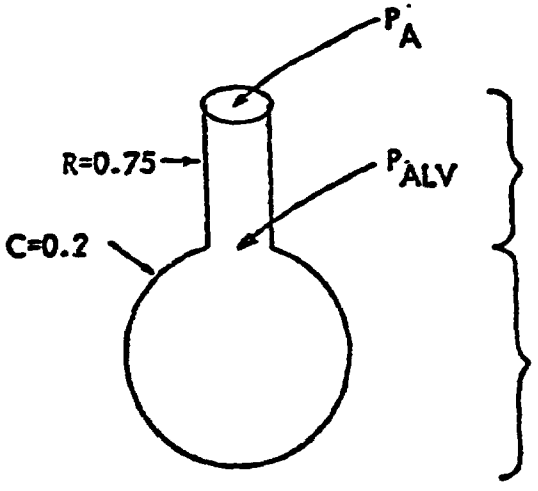
Fig. 9.

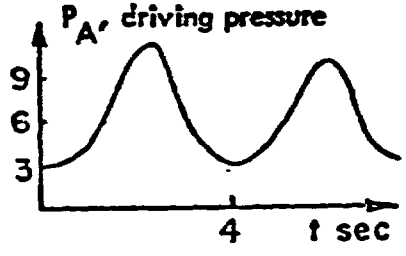


A Block Diagram of the Human Respiratory System

Fig. 10.

PHYSICAL MODEL





EQUATIONS

AIR FLOW THROUGH RESISTIVE PATHWAY:

$$FLOW = \frac{P_A - P_{ALV}}{R}$$

VOLUME OF AIR ENTERING STORAGE COMPARTMENT:

$$VOL = \int (FLOW)\, dt$$

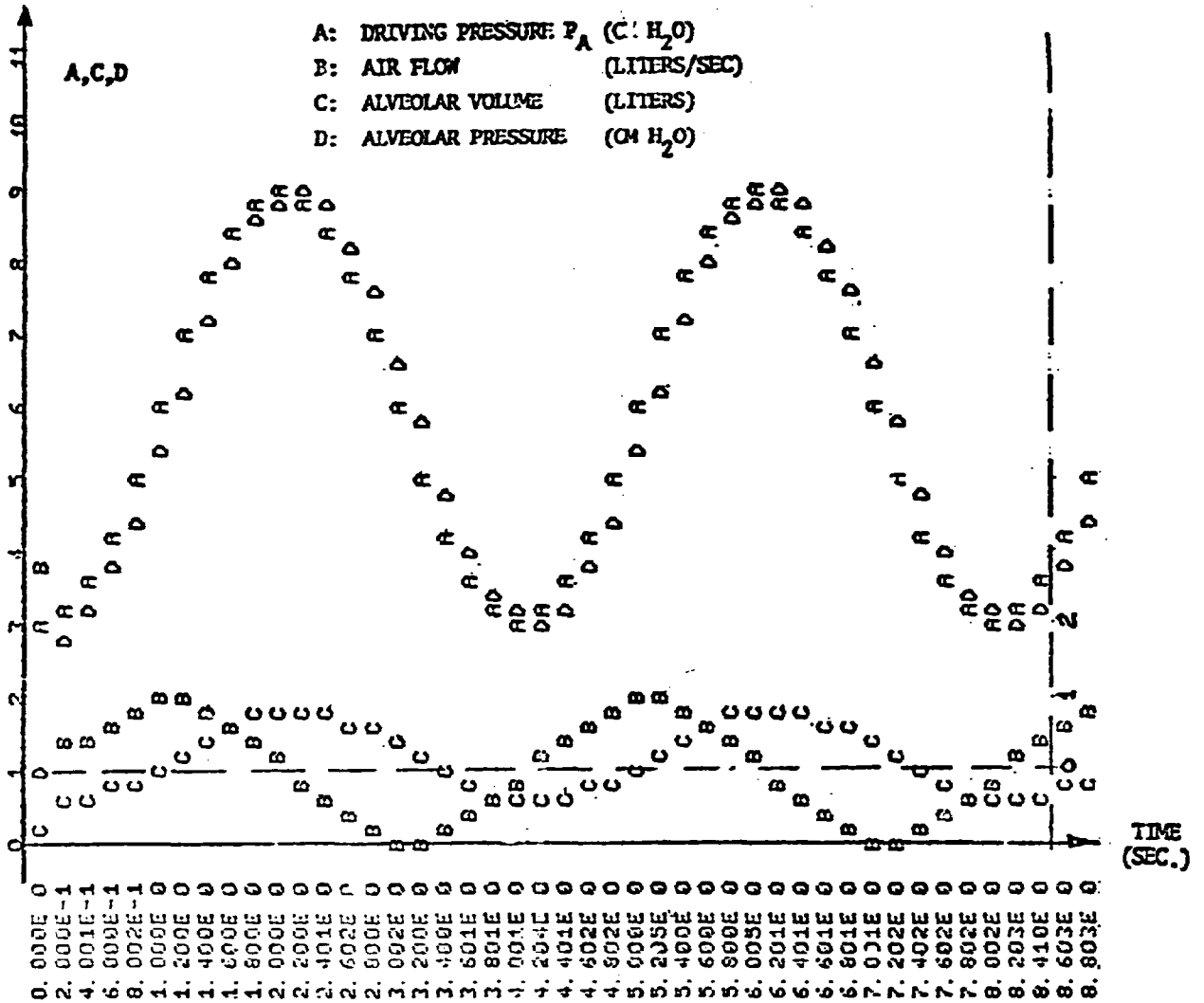
PRESSURE IN STORAGE COMPARTMENT:

$$P_{ALV} = \frac{1}{C}(VOL)$$

```
0    TME   0
1    POT   1
2    POT   2
3    POT   3
4    POT   4
5    POT   5
6    POT   6
7    POT   7
8    MUL   0. .1
9    COS   8
10   MUL   2, 9
11   SUB   3, 10
12   INT   4, 15
13   MUL   6, 12
14   SUB   11, 13
15   MUL   5, 14
16   POT   16
17   ADD   16, 15
18   HDR   PA
                    FLOW . VOL . PALV
19   PRI   11, 17, 12, 13
20   FIN   0, 7
21   END   21
```

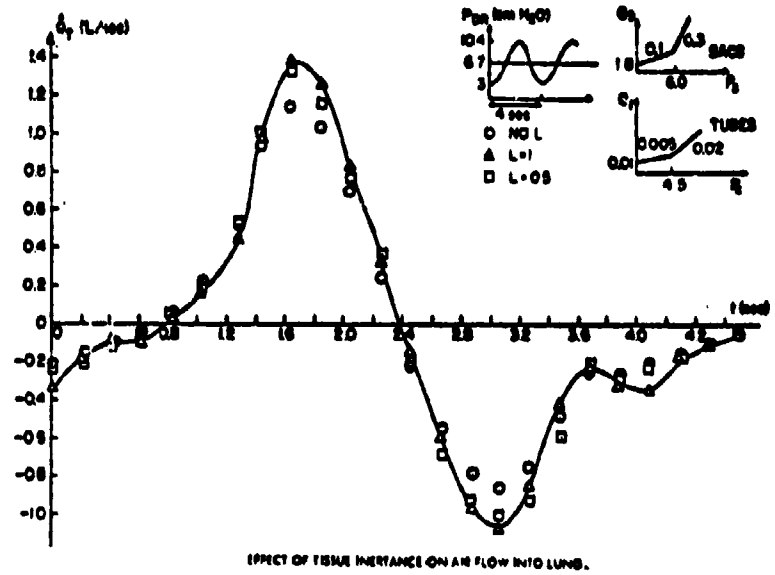
Model, Equations and ISL Listing for Simple Model

Fig. 11.



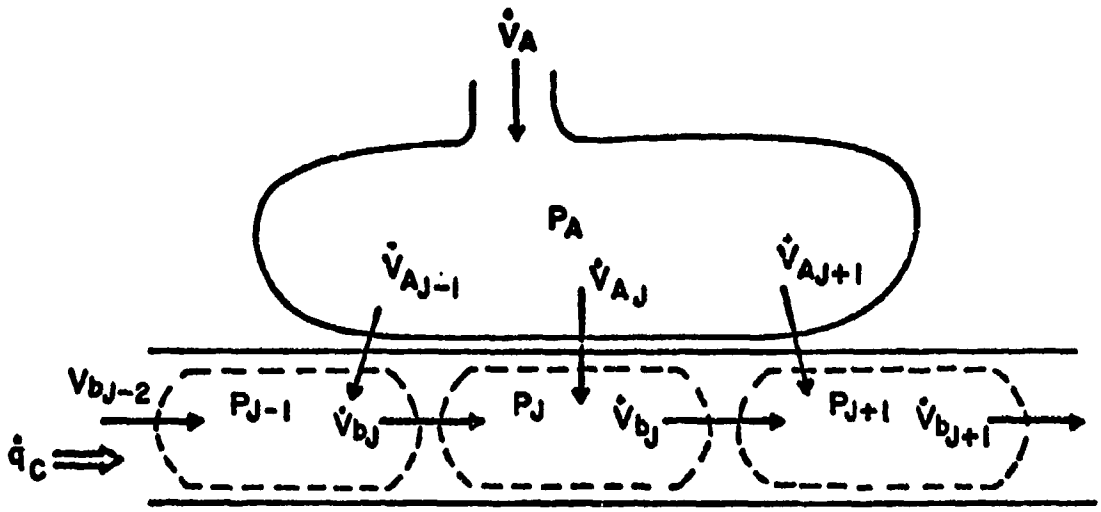
ISL Output Plot for Example Run

Fig. 12.



EFFECT OF TISSUE INERTANCE ON AIR FLOW INTO LUNG.

Fig. 13.



Alveolo-Capillary System Under Study

Fig. 14.

A DESIGN FOR THE DETECTION OF SYNERGY IN DRUG MIXTURES

P. V. Piserchia
B. V. Shah
Research Triangle Institute
Post Office Box 12194
Research Triangle Park, North Carolina

ABSTRACT. In Biometrics [September, 1969], P. S. Hewlett gives a definition of synergy based on the curvature of isobars of drug mixtures. Specifically, if $X(\theta)$ and $Y(\theta)$ represent doses for two drugs A and B which correspond to an $ED(\theta)$ response level (i.e., a proportion θ of all individuals tested will show the specified response) and if $(\lambda X(\theta), (1-\lambda)Y(\theta))$ represents a dose of a mixture consisting of a proportion λ of $X(\theta)$ and $(1-\lambda)$ of $Y(\theta)$, then synergy is absent or present according to whether the proportion $P(\lambda)$ of individuals responding to the dose $(\lambda X(\theta), (1-\lambda)Y(\theta))$ equals or exceeds θ for various values of λ; that is,

$$P(\lambda) > \theta \text{ for some } \lambda \text{ implies synergism.}$$

An immediate consequence of this definition which we prove is:

Suppose X_0 and Y_0 are two doses (not necessarily equivalent) of A and B. Consider the straight line connecting X_0 and Y_0 and written as $X = \lambda X_0$, $Y = (1-\lambda)\, Y_0$, $0 \le \lambda \le 1$. Then, if there exists a λ_0 such that

$$P(\lambda_0) = P(\lambda_0 X_0, (1-\lambda_0)\, Y_0) > \max\{P(X_0,0), P(0,Y_0)\}$$

then there exists a nonlinear isobar and, hence, synergy is shown to occur.

The import of the above derives from the fact that a test for synergy in drugs may be performed with as few as three test groups (those receiving X_0 alone, those receiving Y_0 alone and those receiving $(\lambda_0 X_0, (1-\lambda_0)\, Y_0)$) and, perhaps more important, the doses X_0 and Y_0 need not be equivalent.

1. INTRODUCTION AND DEFINITION OF SYNERGY. In this paper, we shall consider the effects of two drugs, combined in various mixtures, on the responses of some biological system or organism. The principal question of interest is whether the phenomenon of synergism occurs. Following Bushby [1969], we say synergy between two drugs occurs when, acting together, they evoke the same response as when they act singly, but at lower concentrations, or their effects interact in a fashion which is to the advantage of the organism by producing an otherwise un-

attainable rise in biological activity.

Each of the above concepts is related to the nature of some mechanism of joint drug action. A substantial amount of effort has been devoted to the construction of mathematical and statistical models for joint drug action (see Plackett and Hewlett [1967] and Ashford and Smith [1965] for a suitable list of references). However, certain aspects of this research appear to be controversial and no comprehensive and overall acceptable model exists. One reason for this is due to the complex manner in which the effects of drug mixtures are manifested. To use the terminology of Hewlett and Plackett [1959] and Plackett and Hewlett [1967], the joint action of two drugs may be similar or dissimilar according to whether the primary sites of action for the two drugs are the same or different. Alternatively, the joint action may be non-interactive or interactive if one drug has either no influence or some influence on the biological activity of the other.

These distinctions have given rise to four situations as described in the following table:

	Similar	Dissimilar
Non-Interactive	Simple Similar	Independent
Interactive	Complex Similar	Dependent

Plackett and Hewlett [1967] further indicate that one criticism of the above classification is that the "action of two drugs, whether interactive or not, may in some sense be partially similar; similar and dissimilar actions should be regarded as at opposite ends of continuum of biological possibilities." Within this context, the concept of synergism is primarily related to whether the effects of drug mixtures is non-interactive or interactive regardless of its position along the continuum from similar to dissimilar. However, part of the controversy associated with this topic pertains to the equating of no synergism to only the simple similar situation. Hence, although there do exist a number of methods for fitting joint action models, an alternative approach to the concept of synergy which is widely acceptable to most research workers is required.

As a result, Hewlett [1969] has discussed the measurement of the potencies of drug mixtures in terms of isobars, a procedure used in pharmacology. To construct an isobar for two drugs, the doses of the drugs are measured respectively on actual physical scales (e.g., mg/cc) along the two axes and hypothetical points representing the dose pairs producing a fixed biological response are plotted (e.g., 50% of the individuals receiving such a drug mixture dose evoke some specified quantal response). Of course, in an actual situation these points would have to be determined experimentally; but, to elucidate the concept we shall presume that the desired set of points is already known. An example is shown in the figure below where the fixed points on the two axes

correspond to the doses for the two drugs separately which lead to a 50% response rate among the tested individuals.

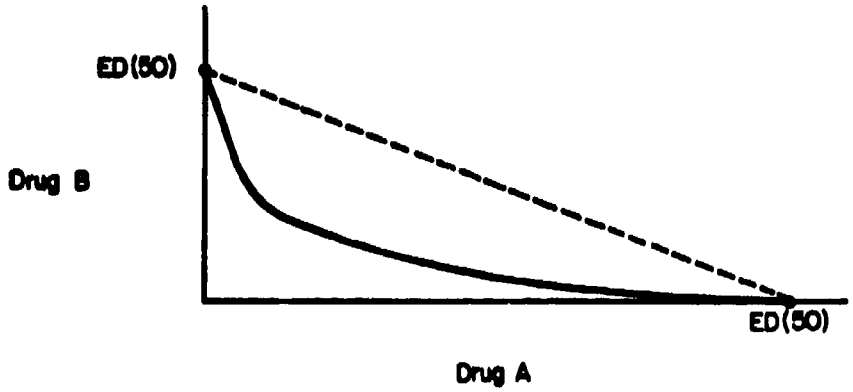


Figure 1. Hypothesized isobar for two synergistic drugs.

The curve in the Figure 1 is called an isobar. If it is a straight line, then one says that the two drugs show "additive action." On the other hand, if it falls below the straight line connecting the two fixed points, then one says that synergism (or potentiation) occurs. This definition tends to bypass the question of similarity or dissimilarity of the joint drug action but yet is consistent with lower concentrations evoking the same response which Bushby [1969] uses in describing synergy.

Hence, throughout the remainder of this paper, synergy will be viewed as curvative of isobars, giving rise to the following formal definition of synergy.

Let $P(X,Y)$ denote the proportion of individuals responding to a mixture of drugs A and B, where X = X units of A and Y = Y units of B.

Assume that $P(X,Y)$ obeys the following:

(a) $0 \le P(X,Y) \le 1$ for $X \ge 0$, $Y \ge 0$,

(b) $P(X,0)$ and $P(0,Y)$ are continuous and monotonically nondecreasing functions of X and Y, respectively.

If for a specific θ there exists an X or Y such that $P(X,0) = \theta$ or $P(0,Y) = \theta$, denote X as $X(\theta)$ and Y as $Y(\theta)$.

Now, suppose there exists a combination of A and B denoted as (X^*,Y^*) with $P(X^*,Y^*) = \theta^*$ (say), then the combination (X^*,Y^*) is said to be synergistic if one of the following conditions holds:

Condition 1: If neither $X(\theta^*)$ nor $Y(\theta^*)$ exist then (X^*,Y^*) is synergistic if $\theta^* > P(X,0)$ for all X and $\theta^* > P(0,Y)$ for all Y.

Condition 2: If either $X(\theta^*)$ or $Y(\theta^*)$, but not both, exist then (X^*,Y^*) is synergistic if $X^* < X(\theta^*)$ and $\theta^* > P(0,Y)$ for all Y, or, $Y^* < Y(\theta^*)$ and $\theta^* > P(X,0)$ for all X.

Condition 3: If $X(\theta^*)$ and $Y(\theta^*)$ both exist then (X^*,Y^*) is synergistic if

$$\frac{X^*}{X(\theta^*)} + \frac{Y^*}{Y(\theta^*)} < 1.$$

Briefly, condition (1) maintains that (X^*,Y^*) is synergistic if an otherwise unattainable rise in biological activity is achieved [Bushby, 1969]. Conditions (2), (3) are, formally, Hewlett's [1969] conditions for synergy.

2. IMPLICATIONS OF THE DEFINITION. An immediate consequence of the above definition is the following theorem and proof.

Theorem: Suppose X_0 and Y_0 are two doses (not necessarily equivalent) of drugs A and B. Consider the straight line joining $(X_0,0)$ and $(0,Y_0)$ and written as $X = \lambda X_0$, $Y = (1-\lambda)Y_0$, $0 \le \lambda \le 1$. Then, if there exists a λ_0 such that:

$$\theta_0 = P(\lambda_0 X_0, (1-\lambda_0)Y_0) > \max\{P(X_0,0), P(0,Y_0)\},$$

then $(\lambda_0 X_0, (1-\lambda_0)Y_0)$ is a synergistic combination of A and B.

Proof:

Case 1: Suppose neither $X(\theta_0)$ nor $Y(\theta_0)$ exist. Then, by the continuity assumption, $\theta_0 > P(X,0)$ for all X, and, $\theta_0 > P(0,Y)$ for all Y.

Hence, $(\lambda_0 X_0, (1-\lambda_0)Y_0)$ is synergistic by Condition 1.

Case 2: Without loss of generality assume $X(\theta_0)$ exists and $Y(\theta_0)$ does not. Then again, by the continuity assumption,

$$\theta_0 > P(0,Y) \text{ for all } Y.$$

Also, $P(X(\theta_0), 0) = \theta_0 > P(X_0,0)$, by assumption, and, through monotinicity, $X(\theta_0) > X_0$.

Therefore, $X(\theta_0) > X_0 > \lambda_0 X_0$ and $(\lambda_0 X_0, (1-\lambda_0)Y_0)$ is synergistic by Condition 2.

Case 3: If $X(\theta_0)$ and $Y(\theta_0)$ both exist then, $\theta_0 = P(X(\theta_0), 0) > P(X_0,0)$ and, $\theta_0 = P(0,Y(\theta_0)) > P(0,Y_0)$.

Hence, by the monotinicity assumption we have:

$$X(\theta_0) > X_0 \text{ and } Y(\theta_0) > Y_0.$$

Therefore,

$$\lambda_0 X(\theta_0) > \lambda_0 X_0 \text{ and } (1-\lambda_0)Y(\theta_0) > (1-\lambda_0)Y_0,$$

and,

$$\lambda_0 > \frac{\lambda_0 X_0}{X(\theta_0)} \text{ and } 1-\lambda_0 > \frac{(1-\lambda_0)Y_0}{Y(\theta_0)}.$$

Therefore,

$$\frac{\lambda_0 X_0}{X(\theta_0)} + \frac{(1-\lambda_0)Y_0}{Y(\theta_0)} < 1,$$

and $(\lambda_0 X_0, (1-\lambda_0)Y_0)$ is synergistic by Condition 3.

Graphically, the above theorem is represented in Figures 2 and 3.

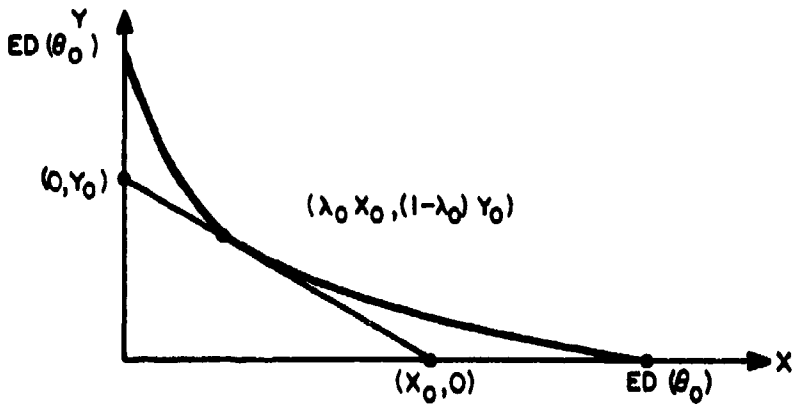


Figure 2. Isobar of a synergistic response. P (X,Y).

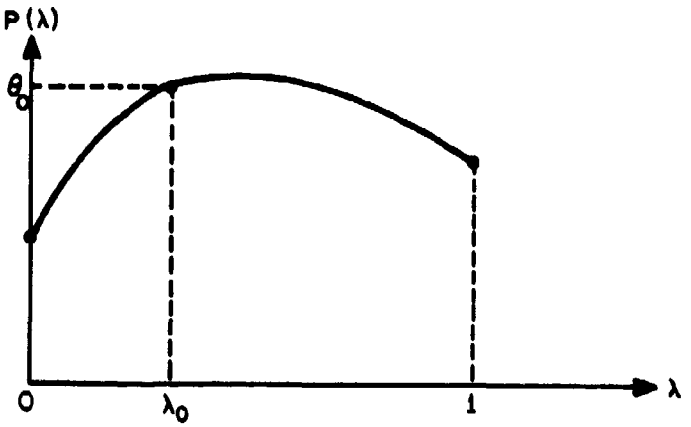


Figure 3. Synergistic response as a function of λ, (X_0, Y_0) fixed.

Notice the above does not require X_0 and Y_0 to be equivalent doses; however, it does require that $\max_{\lambda} P(\lambda)$ be greater than both end points. It is <u>not</u> sufficient to show $P(\lambda) > \lambda P(1) + (1-\lambda) P(0)$. An example should suffice.

Consider the response defined by

$$P(X,Y) = \log_e (X+Y+1) \quad \text{for } X + Y \leq e - 1,$$

$$= 1 \quad \text{for } X + Y > e - 1,$$

then, the isobars of $P(X,Y)$ are the lines $X + Y = \text{const}$. Clearly, straight line isobars and by definition an additive mixture. However, consider the response along any line of the form $X = \lambda X_0$, $Y = (1-\lambda) Y_0$ where $X_0 > Y_0$. We have,

$$P(\lambda) = P(\lambda X_0, (1-\lambda) Y_0) = \log(\lambda X_0 + (1-\lambda) Y_0 + 1)$$

$$= \log(\lambda(X_0-Y_0) + Y_0 + 1).$$

Certainly, $P(\lambda) > \lambda P(1) + (1-\lambda) P(0)$ for every $0 < \lambda < 1$, but yet, by definition, the mixtures are additive.

Figure 4 gives the geometry of the situation.

<u>3. OPTIMAL MIXING</u>. Associated with but not equivalent to synergy is the concept of the optimal mixing of two drugs.

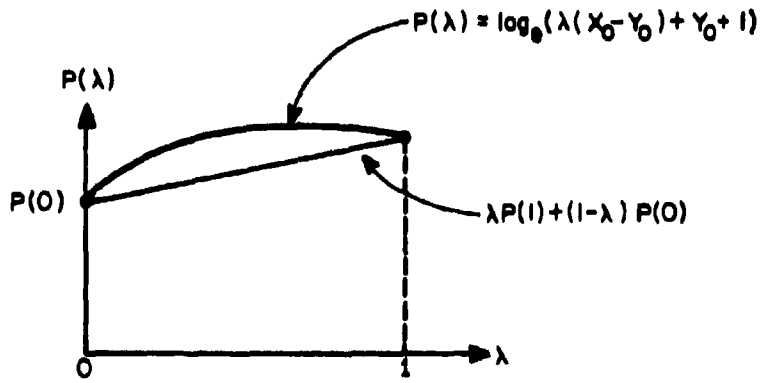


Figure 4. A non-linear, additive drug mixture.

We say two drugs have an optimal mixing rate if there is a ridge in the response, $P(X,Y)$, in a straight ling direction. If the projection of the ridge onto the (X,Y) plane is a line $Y = \rho X$ then we say X and Y have an optimal mixing rate $\rho = Y/X$.

The concept of optimal mixing is useful in establishing synergy. Suppose an optimal mixing rate exists. Then, if X_0 and Y_0 are any two does of X and Y, we have $\max_{\lambda} P(\lambda) = \max_{\lambda} P(\lambda X_0, (1-\lambda) Y_0)$ occurs at the intersection of the two lines:

(1) $X = \lambda X_0$, $Y = (1-\lambda) Y_0$,

(2) $Y = \rho X$.

Solving for λ, we obtain

$$\lambda = Y_0/(\rho X_0 + Y_0),$$

or equivalently,

$$X = X_0 Y_0/(\rho X_0 + Y_0),$$

$$Y = \rho X_0 Y_0/(\rho X_0 + Y_0).$$

It is to be noticed that optimal mixing is defined in terms of the parameter ρ and not in terms of λ. We mention this so as to avoid confusion in picking combinations of doses which are not on the line of optimal mixing. For instance, suppose optimal mixing occurs in a 1:1 ratio. Then, $\rho = 1$ and the line of optimal mixing is $Y = \rho X = X$. Now, suppose we choose doses X_0, Y_0 where $X_0 > Y_0$. Then in Figure 5, we have

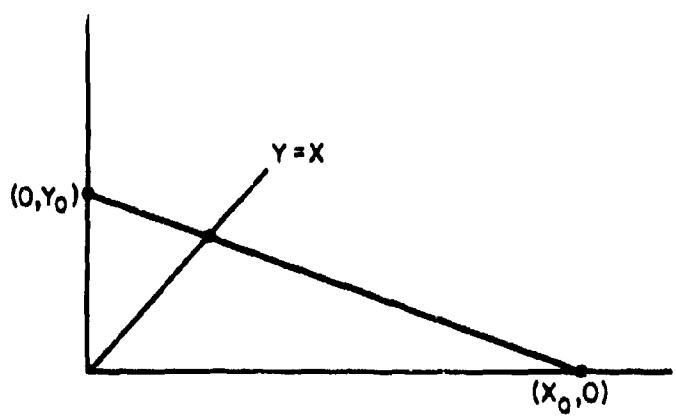


Figure 5. Representation of a three point design.

The maximum of $P(X,Y)$ along $X = \lambda X_0$, $Y = (1-\lambda)\, Y_0$ occurs at the intersection of $X = \lambda X_0$, $Y = (1-\lambda)\, Y_0$ and $Y = X$. It does not occur when $\lambda = 1/2$. Keeping this in mind, selection of combination doses becomes a more rational procedure.

4. DESIGN AND ANALYSIS. Having defined synergy, we now proceed to give certain methods useful in showing synergism if it exists.

The simplest design is the three point design. For a three point design, one chooses doses X_0 of A and Y_0 of B and a combination $(\lambda X_0,\ (1-\lambda)\, Y_0)$ of A and B. Synergism is then said to exist if one can show

$$P(\lambda) = P(\lambda X_0,\ (1-\lambda)\, Y_0) > \max\{P(X_0,0),\ (0,Y_0)\}$$

We propose to do this by testing:

$$H_0:\ P(\lambda) \leq \max(P(X_0,0),\ P(0,Y_0))$$

against the alternative:

$$H_1:\ P(\lambda) > \max(P(X_0,0),\ P(0,Y_0)).$$

The test statistics used will be the simple large sample normal test for differences between two binomial proportions. However, the critical region used will be of the form:

$$\Pr\left\{ \frac{\hat{P}(\lambda) - \hat{P}(X_0,0)}{\sqrt{\frac{\hat{P}(\lambda)\,\hat{Q}(\lambda)}{N_\lambda} + \frac{\hat{P}(X_0,0)\,\hat{Q}(X_0,0)}{N_X}}} > Z_{1-\alpha} \right\} \times$$

$$\Pr\left\{ \frac{\hat{P}(\lambda) - \hat{P}(0,Y_0)}{\sqrt{\frac{\hat{P}(\lambda)\,\hat{Q}(\lambda)}{N_\lambda} + \frac{\hat{P}(0,Y_0)\,\hat{Q}(0,Y_0)}{N_Y}}} > Z_{1-\alpha} \right\} = \alpha^2,$$

where $\hat{P}(X_0,0)$, $\hat{P}(0,Y_0)$ and $\hat{P}(\lambda)$ are the observed proportions of individuals responding at doses X_0 and Y_0 and combination $(\lambda X_0,\ (1-\lambda)\, Y_0)$, respectively, with $\hat{Q}(X_0,0)$, $Q(0,Y_0)$ and $\hat{Q}(\lambda)$ being the respective proportions not responding. Letting $\alpha^2 = .05$ we obtain $Z_{1-\alpha} = Z_{.78} = .760$. Letting $\alpha^2 = .01$, we have $Z_{1-\alpha} = Z_{.90} = 1.285$.

Notice that in the above, no assumption is made about the equivalence of X_0 and Y_0. This is not assumed because it is not necessary to choose equivalent doses to establish synergy. Also, no assumption is made about λ. Again this is done because no assumption concerning λ (other than $0 \leq \lambda \leq 1$) is necessary. However, intuitively, the efficiency of the test procedure should be greatest when $P(\lambda)$ is maximum. Therefore λ should be chosen such that the combination lies on the intersection of the line connecting X_0 and Y_0 and the line of optimal mixing as given in section 3 of this paper.

The Tables I-IV present minimum sample sizes needed to detect synergy for various values of $P_X = P(X,0) = P(0,Y) = P_Y$ and $P(\lambda) = P_\lambda > P_X$. The four tables give required sample sizes for significance levels .05 and .01 and power .80 and .90.

If we define $Z_{1-\alpha}$ and $Z_{1-\beta}$ as the $(1-\alpha)$-th and $(1-\beta)$-th percentage points of the normal (0,1) distribution respectively and if we let $\sigma_X = \sqrt{P_X(1-P_X)}$ and $\sigma_\lambda = \sqrt{P_\lambda(1-P_\lambda)}$ then the formula for determining N, the total sample size, is given by:

$$N = (\sqrt{2}\,\sigma_X + \sigma_\lambda)^2\,(Z_{1-\alpha} + Z_{1-\beta})^2/(P_\lambda - P_X)^2,$$

where α^2 is the significance level of the test and $(1-\beta)^2$ is the power of the test.

To determine N_X, N_Y and N_λ for a given N allocation is carried out by:

$$N_\lambda = N\sigma_\lambda/(\sqrt{2}\,\sigma_X + \sigma_\lambda),$$

and

$$N_Y = N_X = \frac{1}{2}\,(N - N_\lambda).$$

Integer values for N, N_X, N_Y and N_λ were determined by rounding off the values determined by the formulae so that $N_X + N_Y + N_\lambda = N$.

5. SUMMARY. Beginning with an intuitively appealing definition of synergy given by Hewlett [1969], we have attempted in this paper some exploration of the implications of this definition, tried to dispel certain naive notions concerning the analytic characterization of synergy and concerning the optimal mixing of drugs. Too, we have suggested a testing procedure to determine the existence of synergy and have given sample sizes required to detect it.

The techniques discussed in this paper are illustrated in the following example.

Suppose we wish to detect synergy in a mixture of drugs A and B. Further suppose we know 1 unit of A is approximately equivalent to 3 units of B and that A and B have an optimal mixing rate of 1 part A to 2 parts B. Now, denoting A as X and B as Y we have $X_0 = 1.0$, $Y_0 = 3.0$ and $Y = \rho X = 2X$. To derive the best combination of A and B we find

$$X = X_0Y_0/(\rho X_0 + Y_0) = .60 \text{ units of A,}$$

and

$$Y = \rho X_0Y_0/(\rho X_0 + Y_0) = 1.20 \text{ units of B.}$$

Now, suppose $X_0 = 1$ and $Y_0 = 3$ are approximately ED(.50)'s of A and B and it is suspected that the combination (.60, 1.20) gives an expected cure rate of .70. Then, for an $\alpha^2 = .05$ level test with power .80 we find $N = 144$ when $P_A = P_Y = .50$ and $P_\lambda = .70$. We find N_λ, N_Y and N_X by the following:

$$N_\lambda = N\sigma_\lambda/(\sqrt{2}\,\sigma_X + \sigma_\lambda)$$

$$= (144)\ (\sqrt{(17)(.3)})/(\sqrt{2} \times \sqrt{(.5)(.5)} + \sqrt{(.7)(.3)})$$

$$= 56.62.$$

$$N_Y = N_X = \tfrac{1}{2}(N - N_\lambda) = \tfrac{1}{2}(144 - 56.62)$$

$$= 43.68.$$

Hence, we take 56 experimental units for the combination (.60, 1.20) and 44 each for the individual applications of A (1 unit) and B (3 units).

Minimum Sample Size for Detecting Synergy

Table I

Significance Level .05 Power .80

$P_X=P_Y$ \ P_λ	.4	.5	.6	.7	.8	.9
.3	544	139	62	34	20	12
.4	0	596	147	63	32	17
.5	0	0	600	144	59	28
.6	0	0	0	555	126	46
.7	0	0	0	0	462	96
.8	0	0	0	0	0	315

Table II

Significance Level .05 Power .90

$P_X=P_Y$ \ P_λ	.4	.5	.6	.7	.8	.9
.3	751	192	84	45	26	15
.4	0	823	204	86	44	23
.5	0	0	830	198	81	37
.6	0	0	0	768	174	66
.7	0	0	0	0	637	132
.8	0	0	0	0	0	435

Table III

Significance Level .01 Power .80

$P_X=P_Y$ \ P_λ	.4	.5	.6	.7	.8	.9
.3	857	219	97	51	30	18
.4	0	940	232	99	51	28
.5	0	0	948	227	91	43
.6	0	0	0	877	199	74
.7	0	0	0	0	727	149
.8	0	0	0	0	0	496

Table IV

Significance Level .01 Power .90

$P_X=P_Y$ \ P_λ	.4	.5	.6	.7	.8	.9
.3	1113	284	126	68	39	23
.4	0	1223	301	129	66	35
.5	0	0	1232	293	119	57
.6	0	0	0	1139	258	95
.7	0	0	0	0	944	194
.8	0	0	0	0	0	645

REFERENCES

[1] Bushby, S. R. M., "Combined antibacterial action in vitro of trimethoprim and sulphonomides." Postgraduate Medical Journal, Supplement Vol. 45, November 1969, 10-17.

[2] Plackett, R. L. and Hewlett, P. S., "A comparison of two approaches to the construction of models for quantal responses to mixtures of drugs." Biometrics, March 1967, 27-43.

[3] Ashford, J. R. and Smith, C. S., "An alternative system for the classification of mathematical models for quantal responses to mixtures of drugs in biological assay." Biometrics, March 1965, 181-188.

[4] Hewlett, P. S. and Plackett, R. L., "A unified theory for quantal responses to mixtures of drugs: Non-interactive action." Biometrics, December 1959, 591-610.

[5] Hewlett, P. S., "Measurement of the potencies of drug mixtures." Biometrics, September 1969, 477-487.

SOME SEQUENTIAL DESIGNS FOR BINOMIAL CLINICAL TRIALS

L. Arkles
and
R. Srinivasan
Temple University, Philadelphia

ABSTRACT. The problem of selecting the best out of several treatments with dichotomous responses is considered in the framework of the Bechhofer sequential selection model with emphasis on minimizing the number of patients assigned to the inferior treatments. Adaptive sampling rules are proposed for the situations where the response to the treatments is delayed or where several patients have to be scheduled at each stage. Protocols which employ the new sampling rules with various termination rules considered in the literature are shown to be superior or comparable to those which employ the familiar Vector-at-a-Time or Play-the-Winner sampling rule in terms of the average sample number and the inferior treatment number.

1. INTRODUCTION AND DEFINITION OF SAMPLING RULES. Let $\Pi_1, \Pi_2, \ldots, \Pi_k$ be k ($k \geq 2$) binomial populations with respective unknown probabilities of success $p_1, p_2, \ldots, p_k$ where $p_1 \geq p_i$ for $i = 2,3,\ldots,k$. The problem of identifying the population with the largest probability of success, the 'best' population, has been extensively studied in the literature. In this paper we are mainly concerned with the sequential selection model for this problem as formulated by Bechhofer (1958) and Bechhofer, Kiefer and Sobel (1968), and adopted by Sobel and Weiss (1970) to the problem of clinical trials where several treatments with dichotomous responses are being compared.

The Bechhofer model assumes sequential sampling, and consists of a sampling rule which specifies the population to be sampled at any given stage and a termination rule which directs when to stop sampling and how to make the final choice of the best population. The selection is to be made subject to the P^*, Δ^* -admissibility requirement on the probability of correct selection (CS) that

$$P(CS) \geq P^* \text{ for } p_1 - \max\{p_2, p_3, \ldots, p_k\} \geq \Delta^* \qquad (1)$$

where P^* ($\frac{1}{k} < P^* < 1$) and Δ^* ($0 < \Delta^* < 1$) are prespecified constants.

In the context of clinical trials the Bechhofer model provides admissible protocols which assign patients to the treatments sequentially in time, one or more at each stage, until the best treatment is identified with a specified probability. Δ^* can be interpreted as the medically significant or detectible difference. For specified P^* and Δ^*, choice among the various possible admissible protocols is usually made on the basis of the (random) number N_i of patients assigned to treatment i ($i = 1,2,\ldots,k$) and the total number N of patients needed to reach a decision. More specifically, Sobel and Weiss (1970, 1972) base their comparisons on the loss functions

$$E(N) = \sum_{i=1}^{k} E(N_i), \quad \sum_{i=2}^{k} E(N_i) \tag{2}$$

and the risk

$$\sum_{i=2}^{k} (p_1-p_i)E(N_i)$$

the last two measures being given more importance for obvious ethical reasons.

It is convenient at this point to specialize our discussion to the case when $k = 2$; a major portion of this paper as well as most of the past work in this area is confined to the comparison of two treatments. The admissibility condition (1) now reads

$$P(CS) \geq P^* \text{ for } \Delta = p_1-p_2 \geq \Delta^*, \tag{3}$$

and the loss functions of interest, given in (2), become $E(N)$, known as the Average Sample Number (ASN), and $E(N_2)$, the Inferior Treatment Number (ITN).

Most of the protocols considered so far in the literature fall into two broad classes depending on the sampling rule employed. The older and more familiar sampling rule is the so-called Vector-at-a-Time (VT) rule which assigns patients to both of the two treatments at each stage, one to each treatment randomly, until a selection is made based on the termination rule. An essentially equivalent way of implementing the VT rule is to assign the first patient to one of the two treatments at random and then to alternate the treatments given to the subsequent patients as they arrive. It is readily seen that in any protocol which employs the VT rule, regardless of the termination rule used, we have $E(N_1) = E(N_2) = E(N)/2$.

Since one of the basic aims of a clinical trial is to reduce the ITN it was suggested by Zelen (1969) that sampling be done according to the so-called Play-the-Winner (PW) rule instead of the VT rule. The PW rule was

originally studied by Robbins (1956) as a data-dependent policy for the two-armed bandit problem. According to this rule the first patient to arrive is given one of the two treatments chosen at random. The ith patient (i = 2,3,....) is given treatment 1 (treatment 2) if the (i-1)th patient received treatment 1 (treatment 2) and it succeeded or if the (i-1)th patient received treatment 2 (treatment 1) and it resulted in a failure. Zelen investigated the performance of the PW sampling rule in the Anscombe-Colton model (Anscombe, 1963; Colton, 1963) for clinical trials and showed that in general it leads to a significant reduction in the number of patients who receive the inferior treatment.

Subsequently Sobel and Weiss (1970) and several others (See Hoel, Sobel and Weiss, 1975 for an excellent review) have shown that the PW rule is superior to the VT rule in the Bechhofer model in terms of reducing both the ASN and ITN for fixed P^* and Δ^*. Most of the emphasis here has been on devising different termination rules and comparing the resulting protocols with the already existing ones.

Despite its poor performance in terms of the ASN and the ITN, the VT sampling rule has some advantages in its implementation which are not shared by the PW rule. For example, in the PW rule, the allocation of any given patient to a treatment depends on the outcome of the preceding trial, and hence it is required that the response to the treatments be instantaneous or that the response be available by the time a new patient arrives; the VT rule, on the other hand, is applicable in situations of delayed response, and allows for the treatment of several patients at each stage.

One of the purposes of the present paper is to propose and study some sampling rules which are applicable in situations of delayed response. The simplest case here is when patients arrive twice as fast as the response to any one of the two treatments is made available. This is considered in Section 2. The Play-the-Clear-Winner (PCW) sampling rule introduced to handle this case is defined as follows: At the first stage, the first two patients to arrive receive treatments 1 and 2 respectively. At any given stage assignment of treatments is made either for two patients or for one patient depending on the outcome of the preceding stage. At the ith stage (i = 2,3,...) treatments 1 and 2 are assigned randomly to two patients if, at the (i-1)th stage, either (a) treatments 1 and 2 were assigned to two patients and they both resulted in a success or a failure or (b) treatment 1 or 2 was assigned to one patient and it resulted in a failure. At the ith stage (i = 2,3,...) treatment 1 (2) is assigned to one patient if, at the (i-1)th stage, either (a) treatments 1 and 2 were assigned to two patients and treatment 1 (2) resulted in a success and treatment 2 (1) resulted in a failure, or (b) treatment 1 (2) was assigned to one patient and it resulted in a success.

It can be easily verified that the PCW sampling rule is equivalent to the following rule: the first two patients to arrive receive treatments 1 and 2 randomly. The ith patient (i = 3,4,...) to arrive is given treatment

1 (2) if the (i-2)th patient either (a)received treatment 1 (2) and it resulted in a success or (b) received treatment 2 (1) and it resulted in a failure. This formulation implies that the PCW rule is equivalent to implementing two PW rules in parallel, one starting with treatment 1 and the other with treatment 2, a possible solution to the delayed response case suggested by Zelen (1969). This formulation also shows that the PCW rule is applicable in situations where the response to the treatments is instantaneous but two patients are to be scheduled to receive treatments at each stage.

The performance of protocols which employ the PCW sampling rule and various termination rules considered in the literature in connection with the PW rule is summarized in Section 2. Comparisons with the corresponding protocols which use the PW and the VT sampling rules are also presented. It is shown that the PCW rule is in general superior to the other two rules in the sense that it requires comparable or smaller ASN and ITN to reach a decision in addition to its greater generality over the PW rule. Numerical results on the comparisons are presented only for $P^* = 0.95$ and $\Delta^* = 0.2$.

The formulation of the PCW rule as two PW rules in parallel allows us to extend it to situations where m patients are to be scheduled at each stage or patients arrive m times as fast as the response to any one of the two treatments is made available. This is accomplished by simply implementing m PW rules in parallel,[m/2] starting with one of the two treatments chosen at random and the remaining starting with the other treatment. This method of dealing with the delayed-response situations was again essentially suggested by Zelen (1969). Section 3 deals with this rule (denoted PWP for Play-the-Winner-in-Parallel) for $m = 3$. In contrast to Section 2 only a very limited number of termination rules are considered here. Comparisons in terms of ASN and ITN indicate that the behavior of the PWP rule is similar to that of the PCW rule discussed in Section 2.

In Section 4 we return to the problem of selecting the best out of k ($k \geq 3$) binomial populations. The generalization of the VT sampling rule to three or more populations is straightforward. All of the k populations are sampled at each stage. Equivalently, the populations are randomly ordered at the outset and are sampled, one at each stage according to this order, sampling returning to the first population at the end of a cycle. A generalization of the PW rule, called the Play-the-Winner-Cyclical (PWC) sampling rule, appropriate for the present case was studied by Sobel and Weiss (1972). According to the PWC rule, the k populations are randomly ordered at the outset. Sampling starts with the first population. At the ith stage ($i = 2,3,\ldots$) the tth population ($t = 1,2,\ldots,k$) is sampled if, at the (i-1)th stage, either (a) the tth population was sampled and it resulted in a success or (b) the (t-1)th population (0th population being identified with the kth) was sampled and it resulted in a failure. Admissible protocols involving the VT and the PWC sampling rules and the so-called

inverse stopping rule were compared by Sobel and Weiss (1972) using the loss functions defined earlier in this section. They showed that the PWC rule was uniformly better than the VT rule for this stopping rule. Except for their work nothing is at present known about the behavior of the VT or the PWC sampling rule for other termination rules.

A natural generalization of the PCW rule to k populations is as follows: Sample all k populations at the first stage. At the ith stage (i = 2,3,...) sample only those populations which were sampled at the (i-1)th stage and resulted in a success. If no such population exists at the ith stage, then sample all the k populations again and continue the process. We shall refer to this sampling rule also as the PCW rule, and note that it is also applicable in situations where patients arrive twice as fast as the response to the treatments becomes available. In Section 4 we present some numerical results for the PCW rule for k = 3 with the inverse termination rule and some of its modifications applicable only to the VT and the PCW rules. It is shown that with inverse termination the PCW and the PWC rules behave more or less identically while the modified rules lead to improved protocols when employed with the VT or the PCW rules.

Throughout this paper numerical comparisons of the protocols are given only for $P^* = 0.95$, $\Delta^* = 0.2$ and a limited number of values of the parameters $p_1, p_2, \ldots, p_k$. More extensive comparisons as well as the analytical results pertaining to the protocols will be presented elsewhere.

2. THE PCW SAMPLING RULE FOR TWO BINOMIAL POPULATIONS. In this section we consider several termination rules proposed in the literature in connection with the PW sampling rule. The values of ASN and ITN are presented for admissible protocols ($P^* = 0.95$, $\Delta^* = 0.2$) which employ these termination rules and the VT, PW and PCW sampling rules for $\Delta = (p_1-p_2)/2 = 0.2$ and $p_0 = (p_1+p_2)/2 = 0(0.1)0.9$. The sample sizes corresponding to other values of these parameters are available but are not given here since the comparisons presented here reflect the general performance of the protocols quite adequately. Protocols are identified throughout by the sampling rule and the termination rule employed. For example, PCW3 refers to the protocol which uses the PCW sampling rule and Termination Rule 3. Symbols such as $P(CS|PCW3)$, $E(N_2|VT4)$ and $E(N|PW1)$ have their obvious meanings. For i = 1,2, the cumulative number of successes and failures on Π_i, at any given stage will be denoted by S_i and F_i respectively.

Termination Rule 1 (Sobel and Weiss, 1970). Sampling stops as soon as $|S_1-S_2| = r$, where r is chosen so as to make the resulting protocol admissible. The population with the larger number of successes is chosen as the better; in case $S_1 = S_2$, the better population is chosen at random.

For given P^* and Δ^*, the minimum values or r which make the protocols

VT1 and PW1 admissible have been determined by Sobel and Weiss (1970). This can be done for PCW1 using a similar method. For $P^* = 0.95$ and $\Delta^* = 0.2$, these are given by $r = 4$ for VT1, $r = 10$ for PW1 and $r = 8$ for PCW1. Exact expressions for the ASN and ITN of VT1 and PW1 are also given by Sobel and Weiss (1970). Similar expressions can be obtained for PCW1.

Termination Rule 2 (Sobel and Weiss, 1971). Sampling stops as soon as either S_1 or S_2 (or both) equals r where r is preassigned to make the protocols admissible. The population which achieves r successes first is declared the better. If both achieve r successes simultaneously, then the better population is selected at random.

It can be shown that, for all p_1, p_2, $P(CS|VT2) = P(CS|PW2) = P(CS|PCW2)$. Hence the same value of r would make all these three protocols admissible; r equals 20 for $P^* = 0.95$ and $\Delta^* = 0.2$. Sobel and Weiss (1971) have shown that $E(N|PW2) \leq E(N|VT2)$ and $E(N_2|PW2) \leq E(N_2|VT2)$ uniformly in p_1 and p_2. These inequalities can be shown to hold with PW2 replaced by PCW2.

The following termination rule is a modification of Termination Rule 2, and is applicable to the PCW and the VT sampling rules but not to the PW rule. It is defined in terms of the cumulative number of 'clear successes', S_i^c on Π_i ($i = 1,2$), defined by $S_i^c = S_i$ - (the number of times Π_1 and Π_2 were sampled together and they both succeeded).

Termination Rule 3. Sampling stops as soon as either S_1^c or S_2^c (or both) equal r. The population with the larger total number of successes is chosen as the better. If $S_1 = S_2$, then the better population is chosen at random.

For $P^* = 0.95$ and $\Delta^* = 0.2$, the r value which makes the protocol admissible equals 12 for PCW3 and 9 for VT3.

The next termination rule, originally studied by Hoel (1972) for the PW sampling rule, is based on the statistics $R_1 = S_1 + F_2$ and $R_2 = S_2 + F_1$.

Termination Rule 4. Sampling stops as soon as either R_1 or R_2 reaches a preassigned value r, and the population Π_i is selected as the better if R_i reaches r first for $i = 1,2$. With the PCW and the VT sampling rules, $r + 1$ may be reached before stopping. If both R_1 and R_2 reach simultaneously, as is possible with the PCW and the VT rules, the better population is selected at random.

It can be shown that $P(CS|PCW4) = P(CS|PW4)$. Hence, as in the case of Termination Rule 2, the same value of r would make both of these protocols admissible. For $P^* = 0.95$ and $\Delta^* = 0.2$, the minimum value of r equals 33 for PCW4 and PW4, and 29 for VT4.

Termination Rule 5 (Fushimi, 1973). Sampling stops as soon as either $|S_1 - S_2| = r$ or $F_1 + F_2 = s$. The population with the larger number of successes is chosen as the better, and in case $S_1 = S_2$, the better population is chosen at random.

For any given P^* and Δ^* there are in general several values of the pair (r,s) which would make the protocols VT5, PW5 and PCW5 admissible. Fushimi (1973) shows how the 'best' pair can be obtained for PW5 using the property that, as s tends to ∞, the present termination rule reduces to Termination Rule 1 and, as r tends to ∞, it reduces to Termination Rule 2. The 'best' choice of (r,s) corresponding to PCW5 can also be determined along the same lines.

Termination Rule 6 (Nordbrock, 1975). Sampling stops as soon as either $|S_1 - S_2| = r$ or $|\hat{p}_1 - \hat{p}_2| \geq \frac{s}{(F_1+F_2)}$ where $\hat{p}_i = \frac{S_i}{(S_i+F_i)}$; the population with the larger number of successes is chosen as the better, and in case $S_1 = S_2$, the better population is chosen at random.

The remarks made in connection with Termination Rule 5 regarding the choice of (r,s) apply here as well. (r,s) equals (8,4.2) for PCW6, (11,4.2) for PW6 and (4,3.8) for VT6 when $P^* = 0.95$ and $\Delta^* = 0.2$.

Table 1 summarizes our results on the ASN and the ITN of the protocols introduced above for $P^* = 0.95$, $\Delta^* = \Delta = 0.2$ and $p_0 = 0.1(0.1)0.9$. As mentioned earlier, the overall behavior of the protocols is adequately reflected by the results of this table. It can be seen that, except for a few exceptions (for example, for values of p_0 very close to 1), the PCW rule requires comparable or smaller sample sizes when compared to the VT or the PW rule. The increased generality of the VT sampling rule over the PCW rule, and that of the latter over the PW rule should also be kept in mind when comparing these protocols.

3. THE PWP SAMPLING RULE FOR TWO BINOMIAL POPULATIONS. The PWP sampling rule is considered here for Termination Rules 2 and 5 of the previous section. For $P^* = 0.95$ and $\Delta^* = 0.2$, r = 20 for PWP2, and (r,s) = (8,41) for PWP5. Table 2 gives the sample sizes for these two protocols corresponding to the same values of the parameters as in Table 1. It can be seen that the behavior of the PWP sampling rule is quite similar to that of the PCW rule.

4. THE PCW SAMPLING RULE FOR THREE BINOMIAL POPULATIONS. The PCW sampling rule for three binomial populations is considered here with Termination Rule 2 defined in Section 2, and two of its modifications applicable only to the PCW and the VT sampling rules. The protocol PWC2 has been studied by Sobel and Weiss (1972). Closed form expressions for $P(CS|PCW2)$ and $E(N_i|PCW2)$, $i = 1,2,3$, can be obtained using the method of Sobel and Weiss (1972). Numerical results on the probabilities of correct selection for various values of the parameters indicate that, as in the case of two populations, $P(CS|PCW2) = P(CS|PWC2)$ even though we have not been able to establish this. For $P^* = 0.95$ and $\Delta^* = 0.2$, the common value of r which makes the protocols PCW2 and PWC2 admissible is 28.

The modifications of Termination Rule 2 which we consider are quite similar to Termination Rule of Section 2 in that they are obtained by defining 'clear successes' appropriately. In the first modification, Termination Rule 3', we define T_1 = (number of times all three populations were sampled and either Π_1 and Π_2 or Π_1 and Π_3 succeeded and the other failed) + (number of times Π_1 and Π_2 or Π_1 and Π_3 were sampled and Π_1 succeeded and the other failed) + 2(number of times all three populations were sampled and Π_1 alone succeeded), and T_2 and T_3 symmetrically. Termination Rule 3' is then obtained from Termination Rule 2 by simply replacing S_i by T_i for $i = 1,2,3$. Similarly, Termination Rule 3" is obtained from Termination Rule 2 by replacing S_i by U_i for $i = 1,2,3$, where U_1 = (number of times all three populations were sampled and either Π_1 and Π_2 or Π_1 and Π_3 succeeded and the other failed) + (number of times Π_1 and Π_2 or Π_1 and Π_3 were sampled and they both succeeded) + 2[(number of times all three populations were sampled and Π_1 alone succeeded) + (number of times Π_1 and Π_2 or Π_1 and Π_3 were sampled and Π_1 alone succeeded) + (number of times Π_1 alone was sampled and it succeeded)], and U_2 and U_3 are analagously defined. The r values which make the Termination Rules 3' and 3" admissible for $P^* = 0.95$ and $\Delta^* = 0.2$ are respectively 24 and 37.

Table 3 summarizes the expected sample sizes for the protocols of this section for selected values of the parameters. As in the case of Tables 1 and 2, more extensive comparisons are available but are not presented. It is clear from Table 3 that PCW3' is to be preferred over the others.

ACKNOWLEDGEMENTS. All the numerical work involved in this paper was performed at the Temple University Computer Center. Part of the second author's research was supported by a Grant-in-Aid of Research at Temple University.

TABLE 1. EXPECTED SAMPLE SIZES FOR THE PROTOCOLS OF SECTION 2 FOR $P^* = 0.95$ AND $\Delta = \Delta^* = 0.2$.

P_0	$E(N_2)$			$E(N)$		
	PCW1	PW1	VT1	PCW1	PW1	VT1
0.1	33.0	40.5	20.0	73.0	91.0	40.0
0.2	29.5	35.7	19.8	66.0	81.5	39.6
0.3	25.8	30.9	19.2	58.7	71.9	38.5
0.4	22.0	26.0	18.7	50.8	61.9	37.4
0.5	18.0	20.9	18.2	42.7	51.5	37.0
0.6	14.3	15.8	18.7	35.0	41.2	37.4
0.7	10.9	11.0	19.2	28.0	31.2	38.5
0.8	7.8	6.5	19.8	21.9	22.0	39.6
0.9	5.0	2.2	20.0	17.0	13.4	40.0
	PCW2	PW2	VT2	PCW2	PW2	VT2
0.1	81.0	80.5	100.0	181.0	180.5	200.0
0.2	52.9	52.4	66.7	119.5	119.1	133.4
0.3	38.6	38.1	50.0	88.5	88.0	100.0
0.4	29.6	29.1	39.9	69.4	69.0	79.8
0.5	23.3	22.8	33.2	56.4	55.9	66.4
0.6	18.4	17.8	28.4	46.6	46.0	56.8
0.7	14.1	13.4	24.9	38.8	38.1	49.8
0.8	9.9	8.8	22.2	31.8	30.7	44.4
0.9	5.0	2.5	20.0	24.9	22.4	40.0
	PCW3		VT3	PCW3		VT3
0.1	49.0		45.0	109.0		90.0
0.2	34.7		33.0	77.9		66.6
0.3	27.2		28.0	61.8		56.0
0.4	22.2		25.4	51.5		50.8
0.5	18.5		24.7	44.1		49.4
0.6	15.4		25.4	38.2		50.8
0.7	12.5		28.0	32.9		56.0
0.8	9.3		33.0	27.6		66.6
0.9	5.0		45.0	21.0		90.0

TABLE 1. (Continued)

P_0	$E(N_2)$			E(N)		
	PCW4	PW4	VT4	PCW4	PW4	VT4
0.1	26.6	26.5	24.3	59.4	59.0	48.6
0.2	25.9	25.8	24.3	58.6	58.3	48.6
0.3	25.1	24.9	24.3	57.7	57.3	48.6
0.4	24.0	23.8	24.2	56.5	56.2	48.4
0.5	22.6	22.3	24.2	55.0	54.7	48.4
0.6	20.7	20.3	24.2	52.9	52.6	48.4
0.7	17.9	17.3	24.3	50.1	49.6	48.6
0.8	13.3	12.4	24.3	45.5	44.8	48.6
0.9	5.0	2.5	24.3	37.0	35.0	48.6
	PCW5	PW5	VT5	PCW5	PW5	VT5
0.1	20.3	20.4	19.7	45.2	45.9	39.4
0.2	20.5	22.0	19.9	46.2	50.1	39.8
0.3	20.0	22.7	20.4	45.6	52.6	40.8
0.4	18.8	22.2	21.2	43.6	52.7	42.4
0.5	16.8	20.3	22.2	39.9	49.8	44.4
0.6	14.0	16.9	23.3	34.4	43.5	46.6
0.7	10.8	12.2	24.4	28.0	34.1	48.8
0.8	7.8	7.1	24.9	22.0	23.9	49.8
0.9	5.0	2.3	25.0	17.0	14.2	50.0
	PCW6	PW6	VT6	PCW6	PW6	VT6
0.1	13.4	13.5	14.1	29.1	29.8	28.2
0.2	13.7	13.9	14.8	30.1	31.1	29.6
0.3	14.2	14.6	14.4	31.9	33.3	28.9
0.4	15.6	16.2	16.0	35.6	37.9	32.0
0.5	15.7	17.8	17.5	37.1	43.2	35.0
0.6	13.9	16.7	18.7	34.2	41.5	37.3
0.7	10.8	12.0	19.3	27.9	33.7	38.6
0.8	7.7	7.1	19.9	21.8	23.9	39.9
0.9	4.9	2.4	19.9	16.9	14.5	39.8

TABLE 2. EXPECTED SAMPLE SIZES FOR THE PROTOCOLS OF SECTION 3
FOR $P^* = 0.95$ AND $\Delta = \Delta^* = 0.2$

P_0	$E(N_2 \mid PWP2)$	$E(N \mid PWP2)$	$E(N_2 \mid PWP5)$	$E(N \mid PWP5)$
0.1	81.1	166.4	20.2	45.3
0.2	53.0	113.1	20.6	46.6
0.3	38.7	85.7	20.2	46.3
0.4	29.8	68.5	19.1	44.5
0.5	23.5	56.4	17.3	41.2
0.6	18.7	47.2	14.7	36.2
0.7	14.6	39.8	11.8	30.3
0.8	10.7	33.4	9.0	24.9
0.9	6.9	27.6	6.7	20.5

TABLE 3. EXPECTED SAMPLE SIZES FOR THE PROTOCOLS OF SECTION 4
FOR $P^* = 0.95$ AND $\Delta^* = 0.2$

P_1	$P_2=P_3$	$E(N_1)$				$E(N_2)=E(N_3)$			
		PWC2	PCW2	PCW3'	PCW3''	PWC2	PCW2	PCW3'	PCW3''
0.2	0	140.0	140.0	67.3	95.0	112.3	113.0	54.9	77.0
0.3	0.1	93.3	93.3	51.6	67.4	73.0	73.6	41.2	53.5
0.4	0.2	70.0	70.0	44.4	54.0	52.9	53.6	34.4	41.7
0.5	0.3	55.9	55.9	40.7	45.6	40.4	41.1	30.3	33.8
0.6	0.4	46.4	46.4	38.8	39.7	31.5	32.2	27.2	27.9
0.7	0.5	39.6	39.7	37.9	35.1	24.4	25.2	24.4	22.8
0.8	0.6	34.6	34.6	37.2	31.2	18.1	19.1	20.8	17.8
0.9	0.7	30.8	30.8	35.4	27.4	11.3	12.9	14.9	12.2
1.0	0.8	27.9	28.0	29.2	22.8	1.7	5.0	5.0	5.0

REFERENCES

Anscombe, F.J. (1963). Sequential medical trials, J. Amer. Statist. Assn. 58, 365-83.

Bechhofer, R.E. (1958). A sequential multiple-decision procedure for selecting the best one of several normal populations with a common unknown variance, and its use with various experimental designs, Biometrics, 14, 408-29.

Bechhofer, R.E., Kiefer, J., & M. Sobel (1968). Sequential Identification and Ranking Problems, Chicago, University of Chicago Press.

Colton, T. (1963). A model for selecting one of two medical treatments, J. Amer. Statist. Assn., 58, 388-400.

Fushimi, M. (1973). An improved version of a Sobel-Weiss play-the-winner procedure for selecting the better of two binomial populations, Biometrika, 60, 517-23.

Hoel, D.G., Sobel, M., & G.H. Weiss (1975). A survey of adaptive sampling for clinical trials, Advances in Biometry, Ed., R.M. Elashoff, New York, Academic Press.

Nordbrock, E. (1975). An improved play the winner sampling procedure for selecting the better of two binomial populations, to appear in J. Amer. Statist. Assn.

Robbins, H. (1956). A sequential decision problem with a finite memory, Proc. Nat. Acad. Sci., 42, 920-23.

Sobel, M., & G.H. Weiss (1970). Play-the-Winner sampling for selecting the better of two binomial populations, Biometrika, 57, 357-65.

Sobel, M., & G.H. Weiss (1971). Play-the-Winner rule and inverse sampling in selecting the better of two binomial populations, J. Amer. Statist. Assn., 66, 545-51.

Sobel, M., & G.H. Weiss (1972). Play-the-Winner rule and inverse sampling for selecting the best of $k \geq 3$ binomial populations, Ann. Math. Statist., 43, 188-226.

Zelen, M. (1969). Play the winner rule and the controlled clinical trial, J. Amer. Statist. Assn. 64, 131-46.

PREDICTIVISM AND SAMPLE REUSE

Seymour Geisser

School of Statistics
University of Minnesota

ABSTRACT. This paper emphasizes the paramount importance of prediction as opposed to estimation and reviews a variety of general structures for implementing the predictivistic outlook. It also stresses in particular the newly devised predictive sample reuse method as a highly flexible and versatile tool in low structure situations. An illustration is given to a simple survival situation.

1. INTRODUCTION. The fundamental thesis of this paper is that the inferential emphasis of Statistics, theory and concomitant methodology, has been misplaced. By this is meant that the preponderance of statistical analyses deals with problems which involve inferential statements concerning parameters. The view proposed here is that this stress should be diverted to statements about observables. With regard to parameters we take the narrow view which relegates them at most to be components of a statistical model that are not capable of being observed or potentially observed. This is not necessarily to deny them their utility in many hypothetical frameworks but there has been a strong tendency to exaggerate their importance in statistical inference. Even such a compelling "parameter" as the speed of light is in some sense ostensibly capable of being measured (observed) though perhaps subject to error. In this sense it is at least a potentially observable entity. Other values which often are misdesignated as parameters are those defined as a function of a finite number of observables or potential observables which typically occur in sample survey situations. For example we may be trying to "estimate" the total response of a specific finite population by observing some random portion of that population. The unobserved responses are presumably potentially observable (or the randomization is meaningless) and it is maintained that we are basically predicting them or some function of them. This is certainly within the realm of prediction though it is generally referred to as estimating a parameter of a finite population. Hence these two previously mentioned cases, measuring some physically meaningful constant and estimating functions of observables are within the realm of predictivism. It is our contention that in other cases the introduction of a convenient parametric statistical model seems to impel statisticians to reformulate an experimenter's often imprecisely framed question concerning the data into a parametric analysis even when the parameters are completely artificial constructs. We then proceed to foist upon the unwary client "precise" statements about these too often nonexistent entities. This tendency is reinforced because we have too long been subjected to solutions to hypothetical problems which invariably begin -- "suppose we are interested in the estimation of a parametric function BLAH(θ)." This stress on parametric inference made fashionable by mathematical statisticians has been not

This work was supported in part by U.S. Army Grant DAHC04-74-G-0216.

only a comfortable posture but also a secure buttress for the preservation of the high esteem enjoyed by applied statisticians because exposure by actual observation in parametric estimation is rendered virtually impossible. Of course those who opt for predictive inference i.e. predicting observables or potential observables are at risk in that their predictions can be evaluated to a large extent by either further observation or by a sly client withholding a random portion of the data and privately assessing a statistician's prediction procedures and perhaps concurrently his reputation. Therefore much may be at stake for those who adopt the predictivistic or observabilistic or aparametric view. But its relevance is clear.

It was the burden of a previous paper Geisser (1971) to argue that most problems currently cast in terms of parametric estimation and testing could be more informatively reformulated in a predictivistic mode. A general catalogue of such problems was presented there and the Bayesian inferential approach stressed. In this paper we shall discuss the problem of prediction per se from a variety of structures ranging from high to low depending upon the amount of information infused into the model. In particular we will stress a new low structure approach termed predictive sample reuse.

2. HIGH STRUCTURE. The high structure approach to statistical prediction involves the tight apparatus of a prior distribution for the parameters involving known hyperparameters and a specified likelihood, i.e. a joint sampling distribution of observables, past and future, as it were. Hence we need assume that $(X_1,\ldots,X_N; X_{N+1},\ldots,X_{N+M})$ or in a more compact notation $(X^{(N)}; X_{(M)})$ has joint distribution $F(x^{(N)}; x_{(M)}|\theta)$ where θ is a set of unknown parameters. Further, a prior distribution on θ, say $G(\theta|\tau)$, is also assumed where the set of hyperparameters τ is known. The posterior distribution of θ is then based on the observed $X^{(N)} = x^{(N)}$,

$$G(\theta|x^{(N)},\tau) = \frac{F(x^{(N)}|\theta)\, G(\theta|\tau)}{F(x^{(N)}|\tau)} \tag{2.1}$$

where

$$F(x^{(N)}|\tau) = \int F(x^{(N)}|\theta)\, d\, G\,(\theta|\tau). \tag{2.2}$$

This then permits the calculation of the predictive distribution of $X_{(M)}$ given $X^{(N)}$ and τ, resulting in

$$P(x_{(M)}|x^{(N)},\tau) = \int F(x_{(M)}|x^{(N)},\theta)\, d\, G(\theta|x^{(N)},\tau) \tag{2.3}$$

where

$$F(x_{(M)}|x^{(N)},\theta) = \frac{F(x_{(M)},x^{(N)}|\theta)}{F(x^{(N)}|\theta)} \tag{2.4}$$

The denominator of the above being the marginal sampling distribution of the observed random variables $X^{(N)}$. In essence, (2.3) represents the ultimate in statistical prediction and everything else is a summary of one kind or another of this distribution function. If point prediction is of interest then one might choose as a point predictor the predictive expectation of (2.3)

$$E(X_{(M)}|X^{(N)} = x^{(N)}, \tau) \tag{2.5}$$

or the median or the mode of (2.3) or whatever ensues from a particular loss function.

Often in this approach there is a necessary relaxation of the assumption that τ is known. This is generally handled in one of two ways. First it is often the case that little loss in terms of incoherence is engendered by assuming an improper prior for the hyperparameter τ. Hence a new predictive distribution is obtained by calculating

$$P(x_{(M)}|x^{(N)}) = \int P(x_{(M)}|x^{(N)},\tau)\, d\, G\, (\tau) \tag{2.6}$$

A second approach, usually associated with empirical Bayes procedures, is to "estimate" τ from the marginal distribution $F(x^{(N)}|\tau)$ given in (2.2) by maximum likelihood or the method of moments or any other convenient procedure. This then results in an approximate predictive distribution $P(x_{(M)}|x^{(N)},\hat{\tau})$ and a point predictor, say, $E(x_{(M)}|x^{(N)},\hat{\tau})$.

Historically there have also been two other high structure approaches. The first by Fisher (1956) was termed fiducial inference and the second Fraser (1968) termed structural inference. These generally require for their implementation, a much more restrictive sampling distribution and an assumption of complete ignorance concerning θ which in turn implies the absence of τ. Here one would calculate the fiducial or structural distribution $\varphi(\theta|x^{(N)})$ and then compute the predictive distribution of $X_{(M)}$,

$$P_{\varphi}(x_{(M)}|x^{(N)}) = \int F(x_{(M)}|x^{(N)},\theta)\, d\, \varphi\, (\theta|x^{(N)}). \tag{2.7}$$

This type approach is at most valid only under stringent assumptions.

Many statisticians have questioned its validity entirely. Recently Barnard (1975) has developed a pivotal approach to parametric inference. His approach, as demonstrated by Hinklay (1975), can easily be adapted to a predictivistic mode by finding predictive pivots. It appears also to be capable of incorporating certain types of prior information.

3. INTERMEDIATE STRUCTURE. The classical (Neyman-Pearson) approach only assumes $(X^{(N)};X_{(M)}) \sim F(x^{(N)},x_{(M)}|\theta)$, i.e. a sampling distribution and enough structure on the distribution so that one can compute, independent of θ,

$$\Pr\left[X_{(M)} \in A(X^{(N)})\right] = p.$$

This of course is not a probability statement for $X^{(N)} = x^{(N)}$, as in the Bayes approach. Here p represents the degree of confidence that $X_{(M)} \in A\,(x^{(N)})$, p being a valid probability in the sense of the long-term frequency of repetitions from the joint set of random variables $(X^{(N)};X_{(M)})$. In other words, p is the proportion of times in the long run that $X_{(M)} \in A\,(X^{(N)})$ and is interpreted as the confidence one has in $X_{(M)} \in A\,(x^{(N)})$ once $X^{(N)} = x^{(N)}$ has been observed. This is usually referred to as a tolerance interval in the statistical literature. For example, if we are dealing with the problem of predicting the $N+1$ observation X_{N+1} from the first N observations, $X_1,\ldots,X_N$ and assume that $\{X_i\}$ $i = 1,\ldots, N+1$ are iid $N(\theta,1)$ then one notes that for

$$\bar{X}_N = N^{-1}\sum_{i=1}^{N} X_i$$

$$\bar{X}_N - X_{N+1} \sim N(0,\ 1+N^{-1}). \tag{3.1}$$

From (3.1) we obtain

$$\Pr\left[a \le \frac{X_{N+1}-\bar{X}_N}{\sqrt{1+N^{-1}}} \le b\right] = \Pr\left[\bar{X}_N + a\sqrt{1+N^{-1}} \le X_{N+1} \le \bar{X}_N + b\sqrt{1+N^{-1}}\right] \tag{3.2}$$

$$= \Phi(b) - \Phi(a) = p,$$

where $\Phi(y)$ is the standard normal distribution function.

While (3.2) is a probability statement, once we observe $\bar{X}_N = \bar{x}_N$ and calculate the limits, this now becomes a confidence statement and has only the restricted interpretation discussed before.

A point predictor is usually obtained by inserting in $E(X_{(M)}|X^{(N)}=x^{(N)},\theta)$

an estimate $\hat{\theta}(x^{(N)})$ for θ - the expectation being taken over the conditional sampling distribution.

Another approach, having its roots in Fisher's work (1956), termed predictive likelihood, has recently been independently introduced by Hinkley (1975) and Lauritzen (1974). Here as in the fiducial approach, sufficiency though in an extended sense, plays the key role. It is assumed that $(x^{(N)};X_{(M)})$ have likelihood $L(x^{(N)};x_{(M)}|\theta)$ which admits a totally sufficient reduction of the data. In the case of independent and identically distributed random variables a minimal sufficient reduction need only be available. In this latter case as pointed out by Fisher (1956), a minimal sufficient statistic is a function of the individual sufficient statistics from any portion of the entire sample. The concept of a totally sufficient statistic introduced by Lauritzen (1974) permits extension of this result to the more general case of dependence.

Let $s_N = s(X^{(N)})$ and $s_{N+M} = s(X^{(N)},X_{(M)})$ be the set of totally sufficient statistics for θ based on the random variables to be observed and those that are to be observed and predicted, respectively. Then one can obtain, independent of θ, the conditional probability function

$$f(s(x^{(N)})|s(x^{(N)},x_{(M)})) \tag{3.3}$$

which is now defined as being proportional to the predictive likelihood i.e.

$$f(s_N|s_{N+M}) \propto \text{prlk}\ (x^{(N)}|x_{(M)}). \tag{3.4}$$

This is then treated as is the usual $L(x|\theta)$ where now $X_{(M)}$ takes on the role of θ. For the fixed value $x^{(N)}$, the predictive likelihood orders the plausibility for various values $X_{(M)} = x_{(M)}$. For a simple example, consider X_i, $i = 1,\ldots, N + M$ as Bernoulli iid random variables where $P(X_i=1) = 1-P(X_i=0) = \theta$. If r out of the first N are 1's, we can order possible predictive values for the number of 1's, say t, in the next M trials. Defining $R = \sum_{i=1}^{N} X_i$, $T = \sum_{i=1}^{M} X_{N+i}$, which are sufficient, we can compute in a simple fashion

$$P\left[R=r|R+T = r+t\right] = \frac{\binom{N}{r}\binom{M}{t}}{\binom{N+M}{r+t}} \propto \text{prlk}\ (r|t) \tag{3.5}$$

which is used to order the plausible values for $t=0,\ldots,M$.

A point predictor can conceptually be obtained by maximizing the predictive likelihood. In the case where $M > 1$ and the random variables are iid, it is clear that $\text{prlk}(x_{(M)})$ will have multiple maxima due to the exchangeability of the likelihood. This must be so and should be no cause for concern. In the previous example though, there may be a unique maxima at some value of t and be adequate if t is to be predicted. It is clear, however, that if the individual $X_{N+1},\ldots,X_{N+M}$ are to be predicted and the maximum was at $t = t_o$, say, then every partition of $x_{N+1},\ldots,x_{N+M}$ into t_o 1's and $M-t_o$ 0's would also yield identical maxima of the $\text{prlk}(x_{(M)})$.

For a variety of interesting applications of predictive likelihood to standard statistical situations, the reader is referred to Hinkley (1975).

4. LOW STRUCTURE AND ASSESSMENT. Before actually discussing techniques available in low structure situations it will be useful to review a very old and informal method of considerable value in comparing point predictors. Suppose several predictors are suggested for a set of data, then a fruitful comparison of them may be accomplished by a validation technique. The sample $x^{(N)}$ is randomly divided into two parts $x^{(N-n)}=(x_1,\ldots,x_{N-n})$ and $x^{(n)}= (x_{N-n+1},\ldots,x_N)$ called the construction sample and the validation sample respectively. Assume also that associated with each sample point x_i is a known value z_i. The data analyst then computes the competing predictors from the construction sample obtaining, say, $\hat{x}_{ji}(x^{(N-n)},z^{(N-n)};z_j) = \hat{x}_{ji}$ as the i^{th} predictor for the value x_j at known value z_j, $j = N-n+1,\ldots, N$; $i = 1,\ldots,K$ where K represents the number of predictors to be compared, and $z^{(N-n)} = (z_1,\ldots,z_{N-n})$. First the residuals $\hat{x}_{ji} - x_j = r_{ji}$ are computed and then the empirical distribution functions of residuals are plotted for each predictor. A comparison of these empirical distribution functions will shed much light in determining which predictor is most appropriate. Sometimes when the validation sample is not very large a relevant summary measure of the predictive discrepancy is adequate for comparison. For example we might compute the predictive mean squared error $s_i^2 = (N-n)^{-1} \sum_{j=N-n+1}^{N} r_{ij}^2$ $i=1,\ldots,K$. This procedure is generally useful only when a reasonably large number of observations is at hand. This is often not the case. Also the procedure seems inefficient in that it does not extract all of the information in the data. To overcome this a technique which is referred to as simple cross-validation may be substituted.

Let $x_j^{(N-1)} = (x_1,\ldots,x_{j-1}, x_{j+1},\ldots,x_N)$ with corresponding $z_j^{(N-1)} = (z_1,\ldots,z_{j-1}, z_{j+1},\ldots,z_j)$ be the data set with the j^{th} observation omitted. Now for each predictive function we compute the predictor $\hat{x}_{ji} = \hat{x}_{ji}(x_j^{(N-1)}, z_j^{(N-1)}, z_j)$ for the omitted observation x_j and repeat this for $j=1,\ldots,N$ for each predictor obtaining $r_{ij} = \hat{x}_{ji} - x_j$. Similarly as in the validation set up, we are in a position to compare for each predictor its empirical distribution function or a relevant summary measure of predictive discrepancy. However in the case of simple cross validation we have N residuals for each predictor instead of n as in the validation case. One caution is in order -- in the validation case the residuals are dependent only by virture of the same predictive function while in the simple cross-validation some further algebraic dependence creeps in as a result of using the data repetitively. On the other hand the simple cross-validation assessment uses all of the data while the validation assessment only uses a sample of the data. Notwithstanding, the cross-validatory assessment procedure is certainly very useful for the comparison of predictors generated from various structural assumptions as the basic dependence is the same for all of them.

However there are situations where specification of a particular sampling distribution and the resultant predictor based on such assumptions may be fraught with peril. When a particular sampling paradigm becomes difficult or impossible to identify, and yet prediction is necessary, data analytic techniques based on minimal assumptions need come to the fore. One such technique, termed predictive sample reuse (PSR), Geisser (1974a, 1975a) or cross-validatory choice, Stone (1974a), is currently a leading candidate for a satisfactory resolution of this low structure case. It may also be of service in what are basically higher structure situations as we will detail later. First of all the PSR method, when flexibly used, is very likely to be robust for a variety of sampling paradigms. A second feature is that it simulates the predictive process upon itself in some optimal fashion often using some structural hints. It is even capable in one of its manifestations of comparing a variety of approaches. Essentially the goal is to predict a future observation or set of such, or some function of them. For the purposes of this exposition we shall restrict ourselves to a single future observation with a form arbitrarily chosen for predicting it as

$$x = x(X,Z,z;\alpha) \quad \alpha \in \Omega \tag{4.1}$$

where α is some set of unknown values, $X = (x_1,\ldots,x_N)$ represents a sample of size N and with each x_i is associated a known z_i, and $Z = (z_1,\ldots,z_N)$. It must be stressed that in this approach α is not a platonic ideal nor in any sense a true value of paramount importance. It is to be regarded as merely a convenient way of forming a predictive function. Let $P_i^{(N-n)}$ represent the

i^{th} partition of the sample N-n retained and n omitted observations $0 < n \leq M$, where M is the largest integer such that the predictive function (4.1) can be formed with N-M observations. More precisely, the observational set X and the set Z with which it is associated are partitioned such that

$$P_i^{(N-n)} = (X_{ir}^{(N-n)}, Z_{ir}^{(N-n)}; X_{io}^{(n)}, Z_{io}^{(n)}) \tag{4.2}$$

is the i^{th} partition belonging to a set Γ of partitions relevant to a particular schema of observational omissions where $(X_{ir}^{(N-n)}, Z_{ir}^{(N-n)})$ and $(X_{io}^{(n)}, Z_{io}^{(n)})$ represent the N-n retained and n omitted data sets, respectively. Let the total number of such partitions be $P(N, n, \Gamma)$, or simply P. The specified predictive function is then applied to the retained observations for prediction of the omitted observations for each partition with the unknown set of values α estimated by means of optimizing an average discrepancy measure, say,

$$D_{N,n}(\alpha) = P^{-1}n^{-1} \sum_{i \in \Gamma} d(X_{io}^{(n)}, \hat{X}_{io}^{(n)}(X_{ir}^{(N-n)}, Z_{ir}^{(N-n)}, Z_{io}^{(n)}; \alpha)) \tag{4.3}$$

where each element in the set $\hat{X}_{io}^{(n)}$ is the form of the predictive function and d is a measure of the discrepancy of the set of values $X_{io}^{(n)}$ from the set of predicted values $\hat{X}_{io}^{(n)}$ for given α. $D_{N,n}(\alpha)$ is then optimized with respect to α in some sense. On the basis that this leads to a solution say, $\hat{\alpha}$, we obtain the predictor $\hat{x} = x(X,Z,z;\hat{\alpha}) = \hat{f}$.

When predictive functions are to be compared irrespective of their generation one can use a cross-validatory assessment. For a given discrepancy measure we could consider for the i^{th} partition the set of retained observations and associated values $(X_{ir}^{(N-n)}, Z_{ir}^{(N-n)})$ and partition this into two sets $(X_{irr}^{(N-2n)}, Z_{irr}^{(N-2n)}; X_{iro}^{(n)}, Z_{iro}^{(n)})$. From this reduced set of N-n observations and associated values we would, as previously, obtain an $\hat{\alpha}_i$ and compute the discrepancy (not necessarily based on the same d as was used to obtain the predictor) between the values predicted for the n omitted observations and the actual observations themselves. Repeating this for each i we would then compute an overall discrepancy measure

$$D_{N-n}^{*} = P^{-1}n^{-1} \sum_{i \in \Gamma} d(X_{io}^{(n)}, \hat{X}_{io}^{(n)}(X_{ir}^{(N-n)}, Z_{ir}^{(N-n)}, Z_{io}^{(n)}; \hat{\alpha}_i)) \tag{4.4}$$

for each predictive function. This measure then would be relevant to assessing either different predictive functions or various estimators of α in terms of predictive discrepancy for the same predictive functions. We also note that comparisons other than the average D^{*}_{N-n} can be utilized, e.g., empirical distributions of the discrepancy can be compared for several predictors. A variety of applications of PSR can be found in the following papers, Geisser (1974a, 1974b, 1975a, 1975b), Stone (1974a, 1974b). Here we shall only present one such very simple application involving a data based predictor which is to be combined with limited prior information. Let the predictive function be

$$f = \alpha\, h(X) + (1-\alpha)\, g \qquad 0 \le \alpha \le 1 \tag{4.5}$$

where g represents a prior guess at the value to be predicted and $h(X)$ the data based predictor. We shall use the squared discrepancy measure, with a one-at-a-time omission schema so that

$$D_{N,1}(\alpha) = N^{-1} \sum_{j=1}^{N} (\alpha h_j + (1-\alpha)\, g - x_j)^2 \tag{4.6}$$

where h_j is of the form h, but based on $N-1$ observations, i.e. x_j has been omitted. Maximization of $D_{N,1}(\alpha)$ with respect to α yields

$$\begin{cases} \hat{f} = h & \text{if } \hat{\alpha} \ge 1 \\ = g & \text{if } \hat{\alpha} \le 0 \\ = \hat{\alpha}\, h + (1-\hat{\alpha})g & \text{otherwise} \end{cases} \tag{4.7}$$

where

$$\hat{\alpha} = \frac{\sum_{j=1}^{N}(h_j-g)(x_j-g)}{\sum_{j=1}^{N}(h_j-g)^2}. \tag{4.8}$$

In particular if $h = \bar{x}$ then for $s^2 = (N-1)^{-1}\sum_{j=1}^{N}(x_j-\bar{x})^2$ and $t^2 = \frac{N(\bar{x}-g)^2}{s^2}$

$$\begin{cases} \hat{\alpha} = \dfrac{t^2 - 1}{t^2 + (N-1)^{-1}} & \text{if } t^2 > 1 \\ = 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

This procedure has the property that if the sample mean is within one sample standard deviation of the mean from the prior guess g one uses g otherwise one uses the linear combination. Further as the distance between the sample mean and g increases relative to the sample standard deviation, greater weight is attached to the sample mean. Moreover as N increases the predictor tends asymptotically to the sample mean.

In many applications it would appear that observational omissions one-at-a-time are appropriate. However there are some applications where this may not be the case. This point and others involving various schemata of omissions and choice of relevant partitions are discussed in Geisser (1975a).

There have also been various attempts to extend PSR point prediction to sets, intervals, and regions. It is not yet clear as to how satisfactory any of these methods are. Pertinent references are Geisser (1974b), Hinkley (1975), Butler and Rothman (1975).

5. AN APPLICATION. We now illustrate how some of the previous methodology might be applied in practice to what may be termed a simple survival situation. Suppose we have a random sample $X_1,\ldots,X_N$ on an exponential random variable X whose density is

$$f(x|\mu) = \mu e^{-\mu x} \quad \mu > 0,\ x > 0. \tag{5.1}$$

Further suppose our prior objective or subjective information is subsumed in a prior density for μ,

$$p(\mu) \propto \mu^{\delta-1} e^{-\gamma\mu}, \quad \gamma > 0,\ \delta > 0. \tag{5.2}$$

Here μ takes the place of θ in the high structure Bayesian approach and $\tau = (\delta,\gamma)$. Our interest is in predicting a value x_{N+1} for the random future observation X_{N+1} given the previous N observations $x^{(N)}$, say. Then the predictive density for X_{N+1} is easily calculated to be

$$\begin{aligned} f(x_{N+1}|x^{(N)}) &= \int p(\mu|x^{(n)})\ f(x_{N+1}|\mu)\ d\mu \\ &= (N+\delta)(N\bar{x}+\gamma)^{N+\delta}/(N\bar{x}+\gamma+x_{N+1})^{N+\delta+1} \quad z > 0, \end{aligned} \tag{5.3}$$

where $\bar{x}$ is the sample mean and $p(\mu|x^{(N)})$ is the posterior density of μ given the previous N observations $x^{(N)}$. Hence our forecast about X_{N+1} involves the hyperparameters γ and δ which enter the problem via the distribution of the parameter μ. Before any observations are taken one can

also find the predictive (marginal) density of the generic variable X, namely

$$\begin{aligned} f(x) &= \int f(x|\mu)p(\mu)d\mu \\ &= \delta\gamma^{\delta}/(\gamma + x)^{\delta+1}, \quad x > 0. \end{aligned} \tag{5.4}$$

Hence it is convenient and more appropriate from the predictive view to think about these hyperparameters in terms of predicting X before any observations are taken rather than in how they modulate the assumed prior distribution of μ. Therefore, prior to the sample, we have

$$\begin{cases} E(X) = \gamma/(\delta - 1) = g \\ Var(X) = \delta\gamma^2/(\delta - 2)(\delta - 1)^2 = g^2(1 + \alpha)/(1 - \alpha) \end{cases} \tag{5.5}$$

where $\alpha = (\delta - 1)^{-1}$.

Clearly $Var(X)$ exists for $0 < \alpha < 1$, and $E(X)$ exists for $\alpha > 0$ while the distribution exists for all $\alpha \notin [-1,0]$. Hence if one could frame his prior opinions about the potentially observable values of X in terms of its expectation and variance then one can easily execute the whole predictive process by solving for the appropriate values δ and γ from (5.5) and substituting them in (5.3).

It is to be noted that (5.3) and (5.4) were obtained from (5.1) and (5.2). However, for the predictivist who would prefer to start from (5.1) and (5.4) in terms of convenience of framing his predictions this is somewhat awkward. Interestingly enough in this case starting with $f(x|\mu)$ and $f(x)$ is sufficient to obtain $p(\mu)$ and $f(x_{N+1}|\bar{x})$, which is a more logical and appealing approach for the predictivist. This is possible here because $f(x)$ is the unique Laplace transform of $\mu \cdot p(\mu)$.

Now as we mentioned previously positing all of these assumptions yields the requisite information for making probability statements about a future value provided that one has specified values for g and α. However while one may often be willing to hazard a guess at g, one may be far less willing to specify a value for α. So in further analysis of this problem we may be in a position such that some of the parameters of τ are assumed known and others unknown. Assume then that g is known but not α.

One approach for estimating α or δ is from the marginal density

$$\begin{aligned} f(x_1,\ldots,x_N|\delta,\gamma) &= \int f(x_1,\ldots,x_N|\mu)p(\mu|\delta,\gamma)\, d\mu \\ &= \frac{\Gamma(N+\delta)\,\gamma^{\delta}}{\Gamma(\delta)\,[N\bar{x}+\gamma]^{N+\delta}} \end{aligned} \tag{5.6}$$

Since we assume $g = \frac{\gamma}{\delta-1}$ is known we let $Y_i = g^{-1}X_i$ and obtain for $N\bar{y} = \sum_{i=1}^{N} y_i$

$$f(Y_1,\ldots,Y_N|\delta) = \frac{\Gamma(N+\delta)\ (\delta-1)^{\delta}}{\Gamma(\delta)\ [N\bar{y}+\delta-1]^{N+\delta}} \tag{5.7}$$

Clearly $\sum_{i=1}^{N} Y_i = S$ is sufficient for δ in the above likelihood. The density of S is then easily obtained to be

$$f(s|\delta) = \frac{(\delta-1)^{\delta}\ \Gamma(N+\delta)\ s^{N-1}}{\Gamma(N)\ \Gamma(\delta)\ (s+\delta-1)^{N+\delta}} \tag{5.8}$$

which implies that $\alpha S \sim \beta_2(\alpha S; N, \delta)$ a Beta distribution of the second kind. The method of moments essentially fails here to yield a sensible estimate e.g. $E(S) = N$, which is uninformative relative to δ or α. Use of higher moments tends to restrict the range of $\hat{\delta}$ and renders it unreasonable as an estimator. The reason that moment estimators are basically inappropriate here is that they assume the existence of the moments used and hence tend to presume a restriction on the range of δ, whose restriction on the outset is $\delta > 1$. One can use however maximum likelihood estimation. Hence we calculate

$$\frac{\partial \log f}{\partial\ \delta} = \log\frac{\delta-1}{s+\delta-1} + \frac{\delta}{\delta-1} + \frac{1}{\delta} + \frac{1}{\delta+1} + \cdots + \frac{1}{N-1+\delta} - \frac{N+\delta}{s+\delta-1} \tag{5.9}$$

and one would have to find by one means or another $\hat{\delta}$ satisfying $\frac{\partial \log f}{\partial\ \delta} = 0$. An explicit solution for $\hat{\delta}$ seems impossible to achieve. One can approximate (5.9) by using the Euler-Maclauren sum formula so that we obtain for large N

$$\frac{\partial \log f}{\partial\ \delta} \doteq \frac{\delta}{\delta-1} - \log\frac{\delta}{\delta-1} + \log\frac{N+\delta}{s+\delta-1} - \frac{N+\delta}{s+\delta-1} + \frac{1}{2\delta} - \frac{1}{2(\delta+N)} \tag{5.10}$$

This is still quite formidable and when set equal to zero still does not yield an explicit solution for δ.

We now show how PSR may be of service even in this high structure situation. Suppose we were to predict a single value x_{N+1} from (5.3) using the predictive mean

$$E(X_{N+1}|\bar{X} = \bar{x}) = (\alpha N\bar{x} + g)/(\alpha N+1). \tag{5.11}$$

Apply the PSR method for the estimation of α using (5.11) as a

predictive function and squared discrepancy with one-at-a-time omission schema so that

$$D_{N,1}(\alpha) = N^{-1}\sum_{j=1}^{N}\left(\frac{\alpha(N-1)\bar{x}_j+g}{\alpha(N-1)+1} - x_j\right)^2 \tag{5.12}$$

where $\bar{x}_j$ is the mean of the observation with x_j omitted. Minimization of $D_{N,1}(\alpha)$ with respect to α yields

$$\begin{cases} \hat{\alpha} = \dfrac{t^2-1}{N} & \text{for } t^2 > 1 \\ \hat{\alpha} = 0 & \text{for } t^2 \le 1 \end{cases} \tag{5.13}$$

where $t^2 = N(g-\bar{x})^2/s^2$ and $s^2 = N^{-1}\sum_{i=1}^{N}(x_i-\bar{x})^2$. Hence PSR may be used to generate estimates even in the high structure case. On the other hand using (5.11) and (5.12) as a predictive function and discrepancy measure respectively yields a PSR predictor

$$\hat{x}_{N+1} = (\hat{\alpha}\, N\bar{x} + g)/(\hat{\alpha}\, N+1) \tag{5.14}$$

that does not strictly depend on high structure assumptions. In fact it may be robust for a variety of high structure assumptions which result in a predictive expectation approximately equal to (5.11). Actually if one did not use any high structure hint for a predictive function for this problem but merely used a convex combination of sample mean and prior guess

$$x_{N+1} = \alpha^*\,\bar{x} + (1-\alpha^*) \qquad 0 \le \alpha^* \le 1, \tag{5.15}$$

then the result for $\hat{\alpha}^*$ was already obtained in section 4 as

$$\begin{cases} \hat{x}_{N+1} = \dfrac{(t^2-1)\,\bar{x} + \dfrac{N}{N-1}\,g}{t^2 + (N-1)^{-1}} & \text{if } t^2 > 1 \\ = g & \text{if } t^2 \le 1 \end{cases} \tag{5.16}$$

This may be contrasted with (5.14) when the value for $\hat{\alpha}$ is inserted which turns out to be

$$\begin{cases} \hat{x}_{N+1} = \dfrac{(t^2-1)\,\bar{x} + g}{t^2} & t^2 > 1 \\ = g & t^2 \le 1. \end{cases} \tag{5.17}$$

The predictor in (5.17) is weighted slightly more towards $\bar{x}$ than (5.16), but in fact they are asymptotically equivalent to order N^{-1}. In any practical example there would probably not be much to choose between them.

It is also to be noted that the intermediate structures are difficult or impossible to apply in situations such as this one where there may be some prior information that should be taken into account.

6. REMARKS. A somewhat abbreviated exposition of the predictivistic view has been presented. This view is not a mode of inference as such but can be implemented from a variety of inferential modes. It stems from the attitude that inferences should be restricted to potentially observable entities unless compelling reasons to contrary exist. In conformance with this view we have presented various ways, arising from different standpoints, of implementing the predictive approach. In particular a recently developed low structure approach PSR has also been delineated in somewhat greater detail, which should be of great value in many situations and need be added, we believe, to the toolkit of every statistician.

REFERENCES.

Barnard, G. A. (1975). New foundations of statistical inference. Unpublished lecture notes, University of Minnesota.

Butler, R. and Rothman, E.D. (1975). Intervals based on reuse of the sample. Technical Report 56, University of Michigan.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference. (1st Ed.) New York: Hafner.

Fraser, D. A. S. (1968). The Structure of Inference. New York: John Wiley and Sons.

Geisser, S. (1971). The inferential use of predictive distributions. In Foundations of Statistical Inference (V. P. Godambe and D. A. Sprott, eds.) 456-69. Toronto, Montreal: Holt, Rinehart and Winston.

Geisser, S. (1974a). A predictive approach to the random effect model. Biometrika, 61, 101-107.

Geisser, S. (1974b). A new approach to the fundamental problem of applied statistics. Technical Report 235, University of Minnesota.

Geisser, S. (1975a). The predictive sample reuse method with applications. J. Amer. Statist. Assoc. 70, 350, 320-328.

Geisser, S. (1975b). Bayesianism, predictive sample reuse, pseudo observations, and survival. Bulletin of the International Statistical Institute (to appear).

Hinkley, D. V. (1975). Predictive Inference. Technical Report 294, University of Minnesota.

Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models. Scand. J. Statist. 1, 128-34.

Stone, M. (1974a). Cross-validatory choice and assessment of statistical predictions. (with Discussion). J. Roy. Statist. Soc. B, 36, 111-147.

Stone, M. (1974b). Cross-validation and multinomial prediction. Biometrika, 61, 509-515.

VARIOUS METHODOLOGICAL APPROACHES TO PEER EVALUATIONS

Ronald G. Downey and Paul J. Duffy

U.S.Army Research Institute for the Behavioral And Social Sciences

Arlington, Virginia

When confronted with the prospect of drawing order out of complex human behavior in the equally complex world of work, two primary characteristics have marked traditional behavioral science research. First, heavy reliance has been placed upon human evaluations or ratings of other humans. Secondly, these performance or trait ratings have been predominantly gathered from a limited observational viewpoint, namely the supervisor. The technique outlined in the present paper does not deviate from the first of these characteristics; it does rely on human evaluation of other humans. However, it goes beyond the second characteristic by gathering such evaluative information from the additional perspective of an individual's peers. For purposes of the present paper, peers are operationally defined by their sharing of some common purpose (e.g., members of the same work group), and generally by the lack of a formally recognized authority relationship between them. The term associate will be used interchangeably with peer.

The history of peer evaluations can be traced back to post World War II work by Williams and Leavitt (1947).[1] The history of the technique can be traced back even further to the original work of Moreno (1934) and his development of the sociogram technique. Since that time, peer evaluations have been used for two primary purposes. The first of

[1]Some research efforts were reported before this, during and just after World War II. See, for example, Clarke (1946), U.S. Army Research Institute (Note 1), and U.S. Army Research Institute (Note 2) where peer evaluations were used as criteria for leadership studies.

these purposes is evaluative in the criterion sense (i.e., leadership effectiveness, job performance, etc.). The second purpose is evaluative in the sense of predicting future behavior or success (i.e., motivation to work, goal orientation, potential, etc.). Lindzey and Byrne (1968) have presented an excellent review of the use of social choice methodology of which peer evaluations are one type. More specialized reviews of the work are: Gibb (1961), Gibb (1969), Hollander (1954), Boulger and Colmen (Note 3), and Nadal (Note 4).

Aside from considerations about the use of peer evaluations, another major issue centers on what the dimension is which peers are evaluating. For instance, previous research has been directed at peer evaluations of leadership (Hollander, 1965), personality traits (Tupes and Christal, Note 5), and supervisor skills (Weitz, 1958) to name but a few of the dimensions which have been investigated. While we will not directly address the issue of which dimension is measured, it is probably the single most important decision the researcher makes in the design of the experiment.

Given this short background we will address two major areas which relate to the development of a peer evaluation system; first, methodological considerations and second, situational factors which could impact upon the evaluative process.

To facilitate understanding of the methodological issues, they will be described in terms of effects upon the major scaling techniques available, of which there are four: ratings, rankings, full nominations

and high nominations. A summary of the following discussion is provided in Figure 1.

Methodological Issues

The general paradigm of the rating technique calls for a group member to provide a rating of the relative amount or degree of the dimension under consideration possessed by every other group member. The ranking procedure simply requires each group member to rank order every other group member from high to low (or some other relevant continuum) on the dimension under consideration. The full nomination technique requires that each group member choose a specified number or proportion of the group as being either high, medium, or low on the dimension. In the present paper, the minor variation of this technique whenever middle or medium nominations are not required will also be referred to as full nominations. However, the case where only high nominations are elicited is reserved as a discriminably different technique for reasons to be elaborated in later portions of the paper. Several variations based on combinations of these basic techniques are forced distribution rankings or combinations of rankings and ratings or nominations. General scoring algorithms for the four primary techniques are presented below:

Ratings

$$\text{Score} = \frac{\Sigma r_{Rt}}{N}$$

Rankings

$$\text{Score} = \left(\frac{\Sigma r_{Rk}}{N}\right) \times \left(\frac{100}{N_T}\right)$$

Full Nominations

$$\text{Score} = \frac{\Sigma(1 \times r_L) + \Sigma(2 \times r_M) + \Sigma(3 \times r_H)}{N}$$

High Nominations

$$\text{Score} = \frac{\Sigma r_H}{N}$$

where r_{Rt} = Rating

r_{Rk} = Ranking

r_L = Low nomination

r_M = Mid (or no) nomination

r_H = High nomination

N = Number giving an evaluation

N_T = Total number in the group

By inspection, several characteristics of these formulae should be noted. All of these techniques produce scores which are, in general, independent of group size with the exception of the ranking formula in which case adjustment must be made for group sizes greater than 100. It can also be seen that the average score for a group using either a ranking or nomination technique is determined; the average score for the rating technique is free to vary.

Metric and Distribution

The metric and distributional properties of associate evaluations are directly related to the particular technique employed. With respect to the scaling properties of the various techniques, the rankings and both nominations from an evaluator are ordinal data (Stevens, 1951). The ratings from an evaluator are the most nearly

interval data although here also it can be argued that these are merely ordinal data. The scaling properties of the summated scores from the various techniques approximate interval data as the number in the evaluation group increases.

In addition, the 4 most common procedures will commonly produce different distributions, examples of which are displayed in Figure 2. Given the free response mode for ratings, they will often produce negatively skewed distributions due largely to group norms to inflate any evaluative procedure. The ranking procedure, if it were perfectly reliable, would produce a rectangular distribution with one person at each rank. Generally, less reliable rank scores will tend to be normally distributed with even less reliable scores producing a more leptokurtic curve, and a perfectly unreliable test producing a point distribution with everyone receiving an average rank equal to the middle rank. Full nomination scores produce a distribution which, if perfectly reliable, is tri-modal with one group receiving all high nominations, a group with all low nominations and the remainder having middle nominations or none at all. High nominations only produce a bi-modal distribution (not shown in Figure 2).

Basis of Comparison

Scores which result from the four primary techniques vary along another important dimension; that is, the internal process evoked in the evaluator upon which he makes his judgement. In one case, the evaluator compares the particular individual against some external

(to the group) frame of reference and assigns him to some category. In the second case, the evaluator compares the particular individual against some internal (to the group) frame of reference and makes a judgement of more or less and assigns him to the appropriate category. The external process can only be used with the rating procedure. The internal process can be used with the ratings, but it must be used with rankings and nominations. It should be noted that the internal process, in general, requires a moderate number of individuals in the group (more than 5). The direct implication of this distinction is that the external frame of reference allows both comparison between individuals across peer groups and the comparison of peer groups. The internal process does not allow comparison between individuals across peer groups unless the assumption is accepted that the groups are equal on the particular ability, trait or behavior.

The corollary of this implication is that population norms can be developed only through the use of a rating procedure and an external frame of reference.

Reliability

The reliability of associate evaluations has generally been determined by one of two methods, internal consistency or test-retest. Both methods are analogous to the same procedures in classical test theory (Lord and Novick, 1968).

The internal consistency of peer evaluations is the degree to

which members of a peer group agree with one another when observing an individual in a similar situation and at the same time. Using the multiple choice test paradigm, the evaluators are comparable to the test items and those who are being evaluated are comparable to the people taking the test. While Gordon (1969) has recommended the use of the alpha coefficient for estimating the internal consistency or reliability of peer evaluations, the most common procedure has been a split-half (or group) estimate. The split-half estimate is made by computing scores for all group members, randomly assigning peer group members to one of two groups, and then correlating the scores for each ratee from each group (See Hollander, 1957; and Downey, Note 6). The correlation is then adjusted for the total group size using the Spearman-Brown formula (Gulliksen, 1950). If small groups are used, a random split may not be possible and some technique for averaging the intercorrelations between evaluators could be used (Gulliksen, 1950).

The test-retest method of estimating reliability requires that group members evaluate each other at two different times. Scores from the two different evaluations are then correlated. Examples of this type of estimate are given in Hollander (1957), Downey (Note 6), and Downey (Note 7). Perhaps the most rigorous examination of reliability was done by Gordon and Medland (1965) where they varied both time and group doing the evaluations and found reliabilities in the 80's.

Research has generally found the reliability of peer evaluations to be in the .70 to .90 range, regardless of the reliability estimate employed. Research which has compared the various evaluative methodologies is rare, but, in general, has supported the view that all four methods are quite similar with maybe a slight advantage to ratings (See Suci, Vallance, and Glickman, Note 8; Downey, Note 6; and Hammer, Note 9). Even the use of a paired comparison procedure does not significantly improve reliability (Bolton, Note 10). The selection of a particular technique will rarely be decided by differences in reliability between the techniques.

Acceptability

A major factor in the success or failure of a particular research program is the degree of involvement and commitment to the program on the part of the participants, in other words, acceptability. Acceptability is generally studied as a specific issue of the particular program under investigation rather than comparative analyses of acceptability across techniques or situations. There is, therefore, little formal evidence of differences between techniques but many inferences can be drawn based upon the particular qualities of the technique. A major factor in the acceptability of a technique is the degree of perceived difficulty. From this point of view, both the rating and ranking of large numbers of people (greater than 20) can be time consuming and makes for difficult discriminations among the average members of the group. On the other hand, the nomination

procedure allows the individual to place a large number of people in a desired category and does not force him to make difficult discriminations.

The rating procedure is quite acceptable where the group is small and cohesive. The full nomination technique is acceptable for moderate to large size groups where not all individuals are well known to one another. The high nomination technique is even more acceptable because it does not require an individual to make negative evaluations.

A major determinant of the degree of acceptability is the degree to which group members are knowledgeable about the evaluation procedure, process, background and use. Downey (Note 11) found that acceptability improved as a function of an educational program. Two different types of attitudes were found; first, the degree to which peer evaluations were felt to be valuable and accurate estimates and, second, the degree to which they were acceptable for particular uses. Downey also found that the peer evaluations and acceptance were positively related, with larger relationships being found in the group with less information on the peer evaluation process.

Feasibility

Closely linked with the previous concept of acceptability is feasibility, or costs associated with the implementation and execution of a particular peer evaluation system. The major costs associated with a peer evaluation system are: (1) preparation of evaluation materials, (2) administration time, and (3) scoring cost. Prior to

the advent of automatic data processing procedures, the costs associated with any peer evaluation system with large groups or on a large scale were prohibitive. Merely in terms of bits of information collected, it can be seen that the number of evaluations is equal to N^2 where N is the number in the group. Figure 3 presents the comparative costs associated with each of the four techniques.

As can be seen from Figure 3, each of the 4 techniques incur equally high costs associated with the preparation of a list of the peers. It is important that all evaluators be provided with a full list of all other members of the peer group. The administration time for the full nomination technique is low due to the small number of decisions associated with making the low and/or high choices. An excessive amount of time is devoted to making fine discriminations in the ranking procedure, whereas a moderate amount of time is taken up by the rating of every individual.

The scoring of the peer evaluations normally requires access to some sort of automatic data processing facility in all but the smallest scale operations. The actual computer cost is virtually equal for all techniques, but they can differ substantially in terms of the costs associated with getting the evaluations into a data processable form. Costs vary by technique as a function of using either keypunching or optical scanning. Both the full and high nomination techniques involve low cost and ratings also have low costs associated with optical scanning. Rankings produce high costs in both keypunching

and optical scanning and ratings have high costs associated with keypunching. Generally, nominations produce the lowest costs overall followed by ratings with rankings having the highest costs overall. It should be noted that peer evaluation systems are relatively costly efforts which typically require more than minimal sophistication with data processing procedures.

Situational Factors

In addition to the methodological concerns of the various techniques presented in the previous section, there are also a variety of situational or contextual factors which can impact upon a peer evaluation system, often regardless of the specific technique under discussion. Among these factors are group size, informal group structures, demographic characteristics, group boundaries, hierarchical characteristics, friendships, length of association and type of interaction. Each of these factors will be discussed in turn and, where appropriate, specific mention will be made of their effect upon the various techniques.

Size

Very few attempts have been made to study the independent effects of group size. More often than not, what evidence there is for the effects of group size has been reported as a byproduct in studies directed at some other purpose. For example, Downey, Medland, and Yates (Note 12), used a peer nomination technique with groups of Army Colonels in 14 career groups which varied in size from 22 to 321.

Reliabilities varied from .63 to .94 and the rank order correlation between group size and reliability was .03. Downey (Note 7), in a sample of Army Rangers, compared peer ratings collected within squads ($n \cong 10$) with peer nominations collected on the same men within platoons ($n \cong 40$). Correlations between the two scores were in the .60's. However, there were indications that the platoon scores were both more reliable and more predictive of job performance.

As mentioned previously, from the standpoint of feasibility, both ratings and rankings would seem to be most appropriate for relatively small group sizes (i.e., approximately a dozen), while the nomination technique is virtually mandatory for large group sizes (i.e., greater than 50). From the standpoint of empirical results, it appears that small groups may produce unreliable scores with reduced validity. Alternatively, while it is rational to believe that there is an optimal upper size peer group, there is scant evidence to support this view.

Informal Group Structures

Given the well documented fact that within any formally defined group there may exist one or more informal subgroups defined by some sort of mutual self interest, the issue arises as to what effect these informal subgroups may have on a peer evaluation procedure conducted in the total group for a purpose other than finding what subgroup structure exists.

For example, the worst case would be one in which two equal-

sized informal subgroups existed within a total group and included each group member exclusively in one or the other. In such a situation, it can be assumed that one or both subgroups might make their evaluations solely on the basis of subgroup membership, i.e., on a basis other than the one intended. The net effect of such behavior is to attenuate the validity of the peer evaluation procedure, and it is most pronounced when both subgroups engage in such behavior. The effect diminishes if one of the groups does, in fact, provide evaluations on the dimension intended. The effect also diminishes as informal subgroup size decreases or as the number of subgroups increases.

In terms of technique, the effect of subgroup behavior will be pronounced if ratings or rankings are used with resultant scores most likely to be negatively skewed. The use of full nominations will tend to produce scores with decreased variance, and high nominations will produce the worst case with a drastic reduction in variance.

It is clear that subgroups of sufficient size can have an effect upon the final scores, and therefore the question is the incidence of such effects and whether there exists a mechanism for detecting its occurrence. The simplest procedure for checking for these problems is the repetitive production of reliability indices as part of the procedure for producing peer scores. If the reliability coefficients were to drop below .60, it would seem to indicate a problem and care should be taken in use of the evaluations. If the evaluation process

is part of an ongoing process, then the use of a two-way analysis of variance design with one factor being the types of raters and the other factor being the same types of ratees should be used. If a significant interaction were found, then a strong case could be made for peer scores being at least partially the result of group membership.

Demographic Characteristics

The use of peer evaluations with their reliance upon fallible human observers immediately raises the possibility of racial and sexual bias on the part of evaluators. This concern is especially crucial in view of recent problems associated with demonstrating the absence of bias in employment selection and classification measures as well as criterion measures.

The evidence concerning racial bias in peer evaluations is mixed and inconclusive. In a study dealing with Air Force recruits, Cox and Krumboltz (1958) found that subjects were rated higher by members of their own race, but the effect varied across groups and there was substantial agreement on rank order across races (r = .76). They conclude that the bias which might exist is far from complete and suggest that prior acquaintanceship of group members may account for the differences. In a similar study in the Army, deJung and Kaplan (1962) found similar results with ratings differing as a function of the rater's race. However, an analysis of covariance adjusting for a combined interest and math score showed that whites

did not give higher adjusted scores to whites or blacks, but blacks did give higher adjusted scores to blacks. Results were interpreted in terms of assignment of higher scores to close acquaintances which had more of an impact upon blacks rating blacks due to the smaller group size.

In a more recent study in an industrial training context, Schmidt and Johnson (1971) used a forced choice rating distribution in groups with approximately equal numbers of blacks and whites. No differences due to race were found.

The evidence suggests that peer evaluations can be subject to racial bias, but the effect is perhaps more strongly related to the interaction between friendship or acquaintanceship and the particular evaluation method used. The presence of substantial correlations between the rank orderings from each race indicates that a similar view prevailed. But, the use of ratings allows evaluators to assign unrelated scores to individuals whom they consider special in some way.

In terms of sexual bias, Mohr and Downey (Note 13) recently reported results from a small sample of Army officers which indicated that females scored lower than males on scores received from both males and females. If bias occurred, it was on the part of both groups. An interesting finding was that females' self-ratings were not related to either male or female evaluations but males' self-ratings were related to these evaluations.

This admittedly small number of studies appears to indicate that differences based upon race and sex can occur, but it is unclear whether these differences are attributable to race or sex group differences, to interaction patterns (i.e., friendships, etc.), to the specific methodology, or some combination of all of these factors. It would certainly be safe to say that researchers should be sensitive to the potential for such bias.

Group Boundaries

The discussion of peer evaluations has proceeded to this point as if it were clear just what is meant by a peer or associate group. Most researchers report their procedures in sufficient detail to show the general characteristics of the groups which were, in fact, used. However, given that there are a variety of overlapping and higher order groups in most real-life settings, the issue becomes that of defining some basic guidelines for selecting the appropriate rating group. It is clear that the selection of the evaluative group can be effected by such factors as the length and type of interaction, formal organizational structure, informal group structure, friendship patterns and, of course, the particular dimension being evaluated.

As has been the case for several of the preceding issues, there is little empirical data to guide the selection of the group. Rather, guidelines must be best guesses based on partial information from related data.

In the previously mentioned study by Downey (Note 7), it was

found that platoon evaluations produced more reliable and slightly more valid scores than squad evaluations, but the differences were potentially confounded by differences between both method and size. A study by Gordon and Medland (1965), in which individuals were evaluated at two different times by totally different groups of different structure, indicated a high degree of stability across the two evaluations. Even the method which was used to compute reliability indices, random splits of the primary group, supports the notion that group composition can be drastically altered without major problems arising in producing reliable and valid scores.

A concept related to that of group boundaries is that of hierarchies. For example, an Army platoon is made up of 4 squads, each headed by a squad leader. If the platoon is chosen as the peer group, the issue is whether the squad leaders should be included in the process. Folklore holds that the inclusion of such individuals will often work to their disadvantage, and therefore they should be excluded from the platoon peer group and included in a peer group of squad leaders.

Research by Levi, Torrance, and Pletts (1958) indicated no effects from including the formal leader in the peer evaluation process. Research by Downey (Note 14), in which the leaders of small combat units were included in the peer nomination process, indicated that the leaders spanned the full range of leadership potential scores. And, rather than being penalized, there was a positive relationship between formal position and peer evaluation scores (as there should

be if the selection procedure for leaders had any validity originally).

It should be pointed out that these data were experimental and the introduction of an operational system may change the situation depending upon the use to which the resulting evaluations will be put.

A rational solution to the problem should be guided by the following suggestions:

(1) Select the group to have sufficient size to overcome problems associated with primary groups.

(2) Group size should not be so large as to produce subgroups which may be relatively unknown to each other or be competing for similar resources and rewards.

(3) Groups selected should be somehow reasonably related to the dimension to be evaluated, e.g., if evaluation of leadership in a work setting is desired, select a work group and not a social group.

Friendship

Friendship has been a major research issue in the history of peer evaluations. This is another case where folklore has stated that peer evaluations are the product of friendship or popularity and are therefore not valid indications of the dimension under consideration. The impact of this bit of folklore has been that, with the exception of simple validity studies, this is probably the single most researched question associated with peer evaluations.

Wherry and Fryer (1949) were the first to address the issue. They reported that although there was a moderate degree of relation-

ship between friendship and a leadership criterion, the major portion of the predicted criterion variance was independent of friendship. They concluded that peer evaluations of leadership are not popularity contests. Studies by Gibb (1950) and Horrocks and Wear (1953) in college samples support Wherry and Fryer's findings. Borgatta (1954) also reported that leadership and popularity evaluations were related, but he failed to draw any conclusions. Several other studies have documented a moderate degree of relationship between friendship and peer evaluations of leadership Hollander, 1956; Hollander and Webb, 1955; Theodorson, 1957).

Downey (Note 6) recently presented evidence that the use of full nominations (with small numbers of high and low nominations required) reduced the correlation between friendship and leadership evaluations compared with forced distribution ratings.

It would seem that when an evaluator is faced with a choice of how to evaluate a friend, he will tend to select a friend rather than another person he considers of equal, or at least indistinguishable, merit. Therefore, the variance associated with friendship may be a source of systematic error primarily in the middle of the distribution. This systematic error variance will increase in large groups where some members are relatively unknown to each other or the interaction patterns are not fully established for all members.

Even in the face of the impressive research findings demonstrating the invalidity of the "popularity contest" issue, this remains as the

most consistent argument against the use of peer evaluations in an operational setting. A corrollary of this objection is the feeling that peer evaluators do not make the right choice, the best counter-argument to which is the impressive list of validity studies on peer evaluations.

Length of Association

When peer evaluations are considered for use in any situation, an important question concerns how long group members must have been in contact with each other before reliable and valid evaluations can be provided. For example, this issue is often raised in the context of transient training groups. Research is fairly consistent in finding that peers can make reliable and valid evaluations after a relatively short period of time (typically 3 to 6 weeks).

Subsidary to the overall issue is the question of the effect of including a new group member in an intact group. Mayfield (Note 14) has suggested that in such a situation there may be reason to suspect that a longer period of acquaintanceship is necessary for sufficient integration into the group to occur. A more generalized way of approaching the question is the extent to which a person is known or not known to other members of the group. Evidence has shown that an individual not well known to other members of the group will typically be evaluated as lying near the middle of the distribution within the group (Downey, Note 6).

In terms of technique, a nomination procedure is most likely to

decrease the error variance associated with acquaintanceship while ratings or rankings would tend to capitalize on the error variance and show a greater degree of relationship with such measures.

Type of Interaction

While the use of peer evaluations has been extensive over a span of more than twenty-five years, they have nevertheless been applied in rather limited situations. In fact, the majority of the research has been conducted with junior personnel in a military training context. Recent work outside the military by Weitz (1958) and subsequent follow-ups by Mayfield (1970; Note 15) has been conducted in industry with insurance salesmen. There has also been a recent effort to use a peer nomination process in a senior Army officer promotion system which produced supportive results (Downey, Medland, and Yates, Note 12). But, until more extensive research is conducted in broader organizational contexts with a wider selection of subject populations, the generality of the peer evaluation process is largely a matter of conjecture.

A related issue is the type of interaction required to produce valid evaluation. Freeberg (1969) reported a study in which peer evaluations were more highly related to a performance criterion when the interaction between peers was relevant to the dimension being evaluated. Bayroff and Machlin (Note 16) found that leadership evaluations could be made in an academic environment and were highly related to evaluations done after exposure to a situation where

leadership was displayed. Lewin, Dubno, and Akula (1971) indicated that video tapes supplied sufficient information for reliable evaluations and were highly related to evaluations from group members.

It would be safe to assume that peer evaluations of a variety of complex human behaviors can be rendered reliably after exposure of the peers to each other in situations which require the individual to interact either with the environment or with other people in work oriented or socially oriented situations. Further, it can be surmised that the validity of the evaluations will be a function of the degree to which the particular behaviors are relevant to the dimension under study. Hollander (1956) found that reliable evaluations were given after one hour of discussion between peers in a Naval OCS class, but that they had only a moderate degree of relationship with evaluations after 3 weeks, and were not as predictive of eventual job performance. This convergence of views by peers after a short period of exposure is probably a function of similar psychological maps of behavior on the part of peers, and the preliminary evaluations on limited information are subject to revision based upon further information. There would seem to be little advantage of one evaluative technique over another as long as the technique does not require the evaluator to make finer discriminations than are possible based on the type of interaction.

Summary

The peer evaluation technique has been used by researchers both as a criterion of complex human behavior and as an index of future potential. In either case, the particular dimension measured has varied considerably. The present paper reviewed the psychometric properties and related research findings of the four major techniques (ratings, rankings, full nominations and high nominations). Several important similarities and differences were indicated. For example, only ratings can produce comparable scores across different groups without extensive assumptions. In addition, results of research indicate little differences in measurement reliability between techniques. The limited findings also indicate that, in general, ratings and rankings are less acceptable and less feasible than either of the nomination techniques.

Furthermore, a review of both the documented and likely effects of various situational factors on the evaluation process indicated the potential for major problems unless the researcher is aware of the issues. While no direct relationship was found between group size and reliability or validity of the evaluations, it can be assumed that very small or large groups will produce less reliable and less valid scores. Group structure and individual differences were found to be sources of potential problems which must be monitored and dealt with by the researcher. The popular issues of friendship, acquaintanceship and type of interaction were reviewed, and there is little evidence that

they have a major impact on the validity of the scores. Indications are that all techniques are relatively impervious to a variety of situational factors with the nomination technique being perhaps the most versatile.

In brief, it has been shown that peer evaluations have been a fruitful tool in both research and application. Several issues regarding their use remain to be resolved, but there is sufficient evidence to suggest that these issues are soluble and do not detract from the conclusion that peer evaluations are a very powerful tool for discriminating complex human behavior.

	RATING	RANKING	FULL NOMINATION	HIGH NOMINATION
METRIC/ DISTRIBUTION	Interval with tendency toward negative skewness	Interval with tendency toward platykurtic curve	Interval with tendency toward leptokurtic curve	Interval with negative skewness
BASIS OF COMPARISON	Relative or absolute	Relative	Relative	Relative
RELIABILITY	Similar for all methods .70's - .90's			
NORMATIVE SCORES	Norms can be developed across groups	Sample specific unless assumption is made that samples are equal (have same X and S.D.) on the evaluated dimension.		
ACCEPTABILITY				
Small Groups--	HIGH	HIGH	HIGH	HIGHEST
Large Groups--	LOW	LOW	HIGH	HIGHEST
FEASIBILITY				
Small Groups--	GOOD	GOOD	GOOD (Groups > 10)	GOOD (Groups > 10)
Large Groups--	MODERATE TO LOW	LOW	GOOD	GOOD

Figure 1. Summary of Methodological Issues

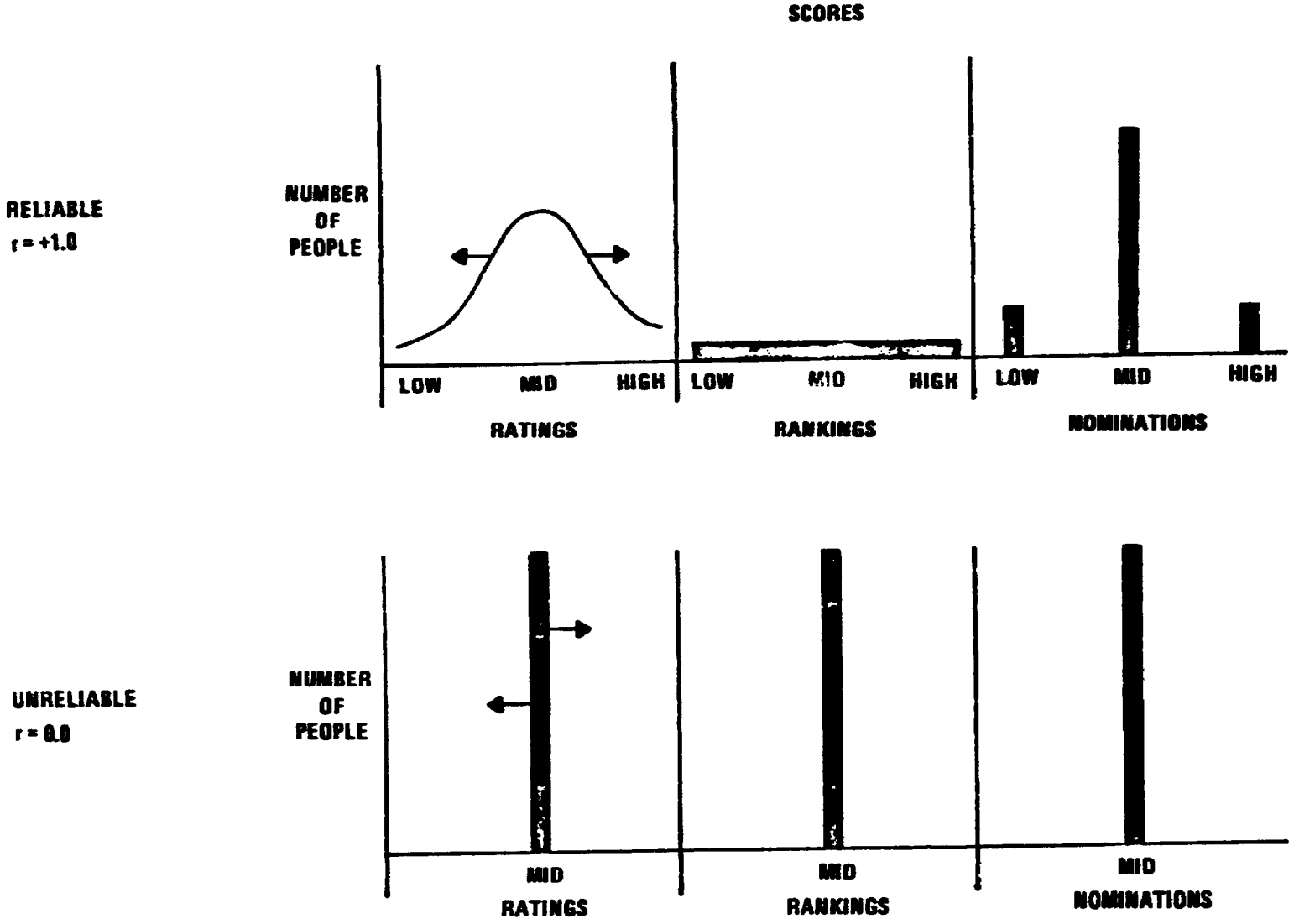


Figure 2. Score Distributions for Reliable and Unreliable Evaluations

	RATING	RANKING	FULL NOMINATION	HIGH NOMINATION
PREPARATION[1] (of Group Lists)	HIGH COST	HIGH COST	HIGH COST	HIGH COST
ADMINISTRATION (Time for evaluations)	MODERATE COST	HIGH COST	LOW COST	VERY LOW COST
SCORING/ REPORTING (data preparation processing)	KP[2]-HIGH COST	KP-HIGH COST	KP-LOW COST	KP-VERY LOW COST
	OS[3]-LOW COST	OS-HIGH COST	OS-LOW COST	OS-LOW COST
SUMMARY	MODERATE COST	HIGH COST	LOW COST	LOW COST

[1]Costs increase with group size.

[2]Key punching.

[3]Optical scanning.

[4]Group sizes of 10 or more.

Figure 3. Data Preparation, Administration and Processing Costs Associated With Peer Evaluation Methodologies[4]

Reference Notes

1. U.S. Army Research Institute for the Behavioral and Social Sciences. Selection of leaders: The status of the measurement of leadership (Report No. 444). Arlington, Va., April, 1943.

2. U.S. Army Research Institute for the Behavioral and Social Sciences. Validation of a program for selection of officers for retention in the peacetime army (Study No. 704). Arlington, Va., July, 1945.

3. Boulger, J. R., & Colmen, J. Research findings with peer ratings (Research Note No. 8). Washington, D.C.: Division of Research Peace Corps, 1964.

4. Nadal, R. A. A review of peer rating studies (Research Report No. 68-8). West Point, N.Y.: Office of Military Psychology and Leadership, United States Corps of Cadets, 1968.

5. Tupes, E. C., & Christal, R. E. Recurrent personality factors based on trait ratings (Research Report ASD-TR-61-97). Lackland Air Force Base, Texas: Personnel Laboratory, Aeronautical Systems Division, Air Force Systems Command, U.S. Air Force, 1961.

6. Downey, R. G. Associate evaluations: Nominations vs. ratings (Technical Paper 253). Arlington, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, 1974.

7. Downey, R. G. Utilization of associate nominations in a training environment: Ranger course (Research Problem Review, in press). Arlington, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, 1975.

8. Suci, G. J., Vallance, T. R., & Glickman, A. S. An analysis of peer ratings (Technical Bulletin No. 54-9). Newport, Rhode Island: Bureau of Naval Personnel, 1954.

9. Hammer, C. H. A simplified technique for evaluating basic trainees on leadership potential (Research Memorandum No. 63-10). Arlington, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, 1963.

10. Bolton, W. L. An application of constant sum paired comparison technique to peer ratings at an Army ROTC summer camp. Unpublished master's thesis, University of Tennessee (Knoxville), 1971.

11. Downey, R. G. Associate evaluations: Improving field acceptance (Research Memorandum No. 75-5). Arlington, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, 1975.

12. Downey, R. G., Medland, F. F., & Yates, L. G. Evaluation of a peer rating system for predicting subsequent promotion of senior military officers. Manuscript submitted for publication, 1975.

13. Mohr, E. S., & Downey, R. G. Are women peers? Manuscript submitted for publication, 1975.

14. Downey, R. G. Personal communication, October 10, 1975.

15. Mayfield, E. C. Peer nominations in the life insurance industry. In R. G. Downey & F. F. Medland (Chair), Peer ratings: Beyond validation studies. Symposium presented at the meeting of the American Psychological Association, Chicago, 1975.

16. Bayroff, A. G., & Wichlin, C. T. Development of criteria of leadership in ROTC. Unpublished manuscript, 1950. (Available from R. G. Downey, U.S. Army Research Institute for the Behavioral and Social Sciences, PERI-II, 1300 Wilson Boulevard, Arlington, Va. 22209.)

References

Borgatta, E. F. Analysis of social interaction and sociometric perception. Sociometry, 1954, 17, 7-32.

Clarke, J. M. Picking the 9,000. Infantry Journal, 1946, 59, 7-13.

Cox, J. A., & Krumboltz, J. D. Racial bias in peer ratings of basic airmen. Sociometry, 1958, 21, 292-299.

deJung, J. E., & Kaplan, H. Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. Journal of Applied Psychology, 1962, 46, 370-374.

Freeberg, N. E. Relevance of rater-ratee acquaintance in the validity and reliability of ratings. Journal of Applied Psychology, 1969, 53, 518-524.

Gibb, C. A. Leadership. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (2nd ed., Vol. 4). Reading, Massachusetts: Addison-Wesley, 1969, 205-282.

Gibb, C. A. The sociometry of leadership in temporary groups. Sociometry, 1950, 13, 226-243.

Gordon, L. V. Estimating the reliability of peer ratings. Educational and Psychological Measurement, 1969, 29, 305-313.

Gordon, L. V., & Medland, F. F. The cross-group stability of peer ratings of leadership potential. Personnel Psychology, 1965, 18, 173-177.

Gulliksen, H. Theory of mental tests. New York: John Wiley and Sons, 1950.

Hollander, E. P. Buddy ratings: Military research and industrial implications. Personnel Psychology, 1954, 7, 385-393.

Hollander, E. P. The friendship factor in peer nominations. Personnel Psychology, 1956, 9, 435-447.

Hollander, E. P. The reliability of peer nominations under various conditions of administration. The Journal of Applied Psychology, 1957, 41, 85-90.

Hollander, E. P. Validity of peer nominations in predicting a distant performance criterion. Journal of Applied Psychology, 1965, 49, 434-438.

Hollander, E. P., & Webb, W. B. Leadership, followership, and friendship: An analysis of peer nominations. The Journal of Abnormal and Social Psychology, 1955, 50, 163-167.

Horrocks, J. E., & Wear, B. A. An analysis of interpersonal choice relationships of college students. The Journal of Social Psychology, 1953, 38, 87-98.

Levi, M., Torrance, E. P., & Pletts, C. O. Sociometric studies of combat air crews in survival training. Sociometry, 1958, 21, 304-328.

Lewin, A., Dubno, P., & Akula, W. Face-to-face interaction in the peer-nomination process. Journal of Applied Psychology, 1971, 55, 495-497.

Lindzey, G., & Byrne, D. Measurement of social choice and interpersonal attractiveness. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (2nd ed., Vol. 2). Reading, Massachusetts: Addison-Wesley, 1968.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

Mayfield, E. C. Management selection: Buddy nominations revisited. Personnel Psychology, 1970, 23, 377-391.

Moreno, J. L. Who shall survive? Nervous and Mental Disease Monograph, 1934 (No. 58).

Schmidt, F. L., & Johnson, R. H. Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1971, 57, 237-241.

Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, Ch. 1.

Theodorson, G. A. The relationship between leadership and popularity roles in small groups. American Sociological Review, 1957, 18, 58-67.

Weitz, J. Selecting supervisors with peer ratings. Personnel Psychology, 1958, 11, 25-35.

Wherry, R. J., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criteria? Personnel Psychology, 1949, 2, 147-159.

Williams, S. B., & Leavitt, H. J. Group opinion as a predictor of military leadership. Journal of Consulting Psychology, 1947, 11, 283-291.

OBJECTIVE ANALYSIS OF CAMOUFLAGE VIA IMAGE INTERPRETERS

RONALD L. JOHNSON

US Army Mobility Equipment Research and Development Center

Fort Belvoir, Virginia 22060

ABSTRACT. In the past the assessment of camouflage effectiveness by its subjective nature has been difficult to objectively quantify. To accomplish this, 63 image interpreters analyzed imagery of a missile site. Subjects reported which visual cues enabled site detection and identification. There were 63 detections and 59 identifications with 13 visual cue categories for detection and 12 for identification. The frequency of response per category ranged from 41 to 1 for detection and 42 to 1 for identification. These frequencies were analyzed by the statistical technique "Minimum Contrasts" at a level of significance .05 and .01. This procedure objectively determined which target items were well camouflaged and which needed improvement.

I. INTRODUCTION.

The camouflage of military installations is becoming increasingly critical as both ground and aerial surveillance techniques improve. The goal of the camouflage is to increase the survivability of the installations, and,

simultaneously, to be cost effective. There is always the need for a reliable measure of the military worth of camouflage. This cannot be estimated, however, without quantifying the effects of the applied camouflage. In the past, this has been extremely difficult due to its inherent subjectivity. The purpose of this paper is to demonstrate a method for that quantification using the statistical technique "minimum contrast" to obtain an item analysis of the subjective cues identified by operational image interpreters.

II. DESIGN OF EXPERIMENT.

The SAM site selected for experimentation was situated in a German agricultural area. Three levels of camouflage were applied. The first was uncamouflaged. The second consisted of tone down painting all roads and buildings, plus construction of an adjacent decoy site. The third level was camouflaged by simulating the surrounding agricultural fields and trees. This was accomplished by using camouflage nets, directional plowing, grass herbicide, and supplementary planting of shrubs and trees. The decoy site in the second level was removed. Each of the three levels were photographed with 60% forward overlap using the following 5" format Kodak film:

Black and White Plus X Kodak #2402

Normal Color Kodak #2448

Color Infrared Kodak #2443

The resulting imagery was cut into strips approximately 15 frames long, the SAM site occupying at least two of the 15 frames. Sixty-three US Marine Corps image interpreters were given thirty minutes to analyze these film

strips. Each level of camouflage and each type of film were viewed by 7 randomly selected image interpreters. Each interpreter was used only once. The visual cues that enabled the image interpreters to make a detection and or an identification were recorded on the data sheet at the end of each test session.

III. EXPERIMENTAL RESULTS.

All 63 of the image interpreters detected the SAM within the alloted 30 minutes. Fifty-nine identified the site. The interpreters cited 13 visual cues which contributed to the site's detection and 12 other visual cues aiding site identification. The visual cues for both detection and identification are extrapolations of specific military aspects of typical cues of psychophysical stimuli materials such as size, shape, contrast, texture, and color. The cues cannot be identified in this report due to security classification, but are included in a confidential report by the author [1]. Tables 1 through 7 contain these detection cues averaged across different combinations of camouflage level and film type. In addition each table shows the frequency the cue was reported by the image interpreters and which cues are significantly different from each other at the .05 and .01 level. This test of significance was calculated using the technique of "minimum contrast" [2]. "Minimum contrasts" is a method to compare two proportions to determine whether the observed contrast is significant at the chosen level. The proportions for this study are the visual cue and the frequency the visual cue was cited by the interpreters as aiding them in site detection identification.

TABLE I

Significant Differences in Detection Between Visual Cues Averaged Across All Levels of Camouflage and Film Types.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														41
B	**													22
C	**													21
D	**													19
E	**	*	*	*										11
F	**	**	**	*										10
G	**	**	**	*										8
H	**	**	**	**										8
I	**	**	**	**										8
J	**	**	**	**										6
K	**	**	**	**										5
L	**	**	**	**	*	*								3
M	**	**	**	**	**	**	*	*	*					1

Cell Size = 63
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 2

Significant Differences in Detection Between Visual Cues Averaged Across Film Types, Uncamouflaged Level.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														13
B														10
C														7
D	**													4
E	**	-												2
F	**	-												3
G	**	**	-											1
H	**	-												3
I	**													4
J	**													4
K	**													3
L	**	-												3
M	**	**	-											1

Cell Size = 21
- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 3

Significant Differences in Detection Between Visual Cues Averaged Across Film Types, Tone-Down Plus Decoy Level.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														14
B	-													6
C	-													6
D														10
E	**			**										2
F	*			-										5
G	**			*										3
H	**			**										2
I	**			**										1
J	**	*	*	**		-								0
K	**			**										1
L	**	*	*	**		-								0
M	**	*	*	**		-								0

Cell Size = 21
- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Siginificant Difference at $\alpha = .01$

TABLE 4

Significant Differences in Detection Between Visual Cues Averaged Across Film Types, Full Camouflage Level.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														14
B	-													6
C														8
D	*													5
E	*													6
F	**													2
G	**													4
H	**													3
I	**													3
J	**													2
K	**													2
L	**	*	*	-	*									0
M	**	*	*	-	*									0

Cell Size = 21
- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 5

Significant Differences in Detection Between Visual Cues Averaged Across Camouflage Levels, Film Type - B&W Plus X.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														16
B	*													7
C	*													6
D	**													4
E	**													2
F	**													2
G	**													3
H	**													5
I	**	-												1
J	**	-												1
K	**													2
L	**	-												1
M	**	**	**											0

Cell Size = 21
- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 6

Significant Differences in Detection Between Visual Cues Averaged Across Camouflage Levels, Film Type - Color.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														13
B														9
C														7
D														7
E	**													4
F	**													3
G	**													4
H	**	**	**	**										0
I	*													5
J	**	*												2
K	**	**	-	-										1
L	**	**	**	**										0
M	**	**	-	-										1

Cell Size = 21
- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Signicant Difference at $\alpha = .01$

TABLE 7

Significant Difference in Detection Between Visual Cues Averaged Across Camouflage Levels, Film Type - Color IR.

	A	B	C	D	E	F	G	H	I	J	K	L	M	Frequency
A														12
B														6
C														8
D														8
E														5
F														5
G	**		*	*										1
H	**													3
I	**													2
J	**													3
K	**													2
L	**													2
M	**	*	**	**	-	-								0

Cell Size = 21
\- = Border Line Significance at $\alpha = .05$
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

Tables 8 through 14 contain the 12 visual cues which contributed to site identification averaged across different combinations of camouflage and film type. Cue frequency and significance at $\alpha = .05$ are again included as in the preceding tables. As before, the cues cannot be identified because of security classification.

TABLE 8

Significant Differences in Identification Between Visual Cues Averaged Across All Levels of Camouflage and Film Types.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													42
B	**												25
C	**												15
D	**	*											13
E	**	**											8
F	**	**											8
G	**	**											8
H	**	**	*	*									4
I	**	**	**	**									2
J	**	**	**	**									2
K	**	**	**	**	*	*	*						1
L	**	**	**	**	*	*	*						1

Cell Size = 59
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 9

Significant Differences in Identification Between Visual Cues Averaged Across Film Types, Uncamouflaged Level.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													17
B	**												8
C	**												2
D	**												7
E	**												5
F	**												6
G	**												4
H	**												2
I	**	*		*									1
J	**	*		*									1
K	**	**		**	*	*							0
L	**	**		**	*	*							0

Cell Size = 17
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 10

Significant Differences in Identification Between Visual Cues Averaged Across Film Types, Tone-Down Plus Decoy Level.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													12
B													6
C													7
D	*												4
E	**												2
F	**		*										1
G	**												2
H	**		*										1
I	**	*	**										0
J	**		*										1
K	**		*										1
L	**	*	**										0

Cell Size = 17
* = Significant Difference at α = .05
** = Significant Difference at α = .01

TABLE 11

Significant Differences in Identification Between Visual Cues Averaged Across Film Types, Full Camouflage Level.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													13
B													11
C	*												6
D	**	**											2
E	**	**											1
F	**	**											1
G	**	**											2
H	**	**											1
I	**	**											1
J	**	**	*										0
K	**	**	*										0
L	**	**											1

Cell Size = 17
* = Significant Difference at α = .05
** = Significant Difference at α = .01

TABLE 12

Significant Differences in Identification Between Visual Cues Averaged Across Camouflage Levels, Film Type - B&W Plus X.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													15
B	*												8
C	**												2
D	**												4
E	**												3
F	**												3
G	**	*											1
H	**												2
I	**	**											0
J	**	**											0
K	**	**											0
L	**	**											0

Cell Size = 17
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 13

Significant Differences in Identification Between Visual Cues Averaged Across Camouflage Levels, Film Type - Color.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													10
B													10
C													4
D													5
E	*	*											2
F	*	*											3
G	*	*											2
H	**	**											1
I	**	**											1
J	**	**											1
K	**	**											1
L	**	**											1

Cell Size = 17
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

TABLE 14

Significant Differences in Identification Between Visual Cues Averaged Across Camouflage Levels, Film Type - Color IR.

	A	B	C	D	E	F	G	H	I	J	K	L	Frequency
A													17
B	**												7
C	**												9
D	**												4
E	**												3
F	**		*										2
G	**												5
H	**	*	**										1
I	**	*	**										1
J	**	*	**										1
K	**	**	**				*						0
L	**	**	**				*						0

Cell Size = 17
* = Significant Difference at $\alpha = .05$
** = Significant Difference at $\alpha = .01$

IV. DISCUSSION.

A review of tables 1-7 demonstrates that the isolation of the critical visual cues for site detection was accomplished by the use of "minimum contrasts." Detection cue A was a significant factor in all tables for the detection of the SAM site. There was virtually no change in the importance of this cue in site detection when analyzed across levels of camouflage and film type. Therefore, more effort must be expended to prevent this cue from becoming a major factor in target detection. The addition of the decoy site adjacent to the SAM site had a pronounced effect in increasing the number of significant cues that allowed the image interpreter to detect the site (table 3). Visual cues E and F, and H through M did not have much effect on site detection either for level of camouflage or type of film analyzed. The number of cues leading to site detection was greater for the color and color infrared film than for the black and white film (tables 5-7). As is well known, more information is presented to the image interpreter on color and color infrared film than on black and white imagery.

Tables 8-14 indicate that the use of "minimum contrasts" to isolate the critical visual cues in the identification of the SAM site was successful. Visual cues important for site identification were different from those for site detection. Identification cues A and B were the most important except for camouflage level two containing tone-down and site decoy. For this

case, cues A and C were the most prominent in site identification (table 10). The effects of visual cue C were essentially nil for levels one and three (Tables 9 and 11). Visual cues D through L had little effect on site identification when analyzed by level of camouflage or type of film. Color infrared film generated more visual cues to target identification (Table 14) than both color and black and white films (Tables 12-13). We consider this to be due to the greater amount of information presented to the image interpreter with color infrared film than for the other two film types. The results indicated that this approach was a valid method to objectively analyze subjective cues.

V. SUMMARY AND CONCLUSIONS.

The problem faced in this study was to objectively analyze the effects of levels of camouflage on detection and identification. A SAM site was selected and photographed. Subjective visual cues were elicited from operational image interpreters in response to specially prepared classified packets of site photography. These cues for both detection and identification were grouped into categories and analyzed for significance using the technique of "minimum contrasts". This technique facilitated the quantification of the subjective cues used by image interpreters in site detection/identification for levels of camouflage and types of photographic film.

REFERENCES

1. Johnson, Ronald L., Evaluation of Joint Field Trial (JFT) Camouflage of Air Defense Site by Operational Photo Interpreters, Confidential, Countersurveillance and Topographic Division, Laboratory 4000, US Army Mobility Equipment Research and Development Center, Fort Belvoir, VA 22060, 7 August 1975.

2. Natrella, Mary Gibbins, Experimental Statistics, National Bureau of Standards, Handbook 91, Chapter 8, page 9, US Department of Commerce, Washington, DC, 1966.

A SIMPLE METHOD FOR DETERMINING THE UNRESTRICTED AVERAGE OUTGOING QUALITY LIMIT (UAOQL) OF A CONTINUOUS SAMPLING PLAN

Richard M. Brugger
RAM Assessment Division
Product Assurance Directorate
US Army Armament Command
Rock Island, Illinois

ABSTRACT. This paper provides a simple algorithm for determining the Unrestricted Average Outgoing Quality Limit (UAOQL) of a continuous sampling plan. The derivation of the algorithm is shown.

1.1 INTRODUCTION. As a prerequisite to a discussion of the UAOQL, some review of the fundamentals of continuous sampling is in order.

Most statisticians are familiar with the concept of sampling from a lot. For example, we might have a lot of one hundred items, from which a sample of size seven has been drawn. The acceptance decision for the lot will be based on the results found in the sample. For example, the rules of the sampling plan we are using might say that if two or fewer units out of the sample of size seven are defective, we shall accept the lot. If three or more units are defective, we shall reject the lot.

Under continuous sampling, we do not use the concept of sampling a certain number of units from a lot of material. Instead, we carry out inspection as the units are produced and flowing along the production line.

The prerequisites for using a continuous sampling plan are:

a. Moving product.

b. Ample physical facilities for 100% inspection when necessary.

c. Relative ease of inspection

d. A process capable of producing homogeneous material.

An example of a continuous sampling plan is Harold Dodge's CSP-1 [2]. Dodge was the original developer of continuous sampling plans, and published his first work on the subject in 1943. Under CSP-1, at the start of inspection, the screening crew inspects 100% of the units. When some prespecified number, i, of consecutive units are free of the defects concerned, that is, the defects for which we are inspecting, the screening crew is released from 100% inspection and the sampling inspector inspects a prespecified fraction, f, of the units, where the sample units are selected in a random manner as they pass the point of inspection. If a defective unit is found, 100% inspection is resumed, and the cycle repeats itself as necessary during the remainder of production.

We made mention of the values of i and f, which are specified for each individual CSP-1 plan. For example, we might have a clearance number, i, of twenty units, and a sampling frequency, f, of one in ten.

Some of the functional properties of a CSP-1 plan (or any CSP plan for that matter) that are usually of interest to the statistician are the following:

a. The Average Fraction Inspected, of AFI, which is the expected value of the fraction of material that will be inspected over an indefinitely long period of time when each unit has probability p of being defective.

b. The Average Outgoing Quality, or AOQ, which is the expected fraction of material that is defective in accepted material over an indefinitely long period of time when each unit has probability p of being defective.

c. The Average Outgoing Quality Limit, or AOQL, which is the maximum value of AOQ.

Thus far, we are talking about properties based on the assumption that each unit has probability p of being defective. This is of course a very restrictive assumption, since one might intuitively feel that in the real life situation, p would undergo some sort of variation over time. For this reason, statisticians concerned themselves with the problem of how to describe the mathematical properties of continuous sampling plans when p varied over time. In 1953, Lieberman [4] presented an analysis of

CSP-1 under the assumption that p was not constant for each unit. He determined that the worst situation would be the one where only good units reached the inspector during phases of 100% inspection, and only bad units reached the inspector during phases of sampling inspection.

The outgoing quality reflected by this worst possible situation eventually came to be called the Unrestricted Average Outgoing Quality Limit, or UAOQL. There is a very interesting paper on the UAOQL by Sackrowitz [5] in the April 1975 Journal of Quality Technology; however, Sackrowitz's definitions are somewhat different from what we will discuss here.

There are two general cases that we will consider: that situation where defective units found are removed from the flow of product and replaced with good units, and the situation where defective units found are removed from the flow of product but are not replaced with good units.

For the replacement case, White [6] carried out a quite complex derivation involving linear programming to show that for a broad class of plans, the UAOQL would result from that situation where for any phase of inspection of a plan, either all good units are submitted during every occurrence of the phase or all bad units are submitted during every occurrence of the phase. White [7] computed numerical solutions for plans from DOD Handbook H106. Endres [3], an employee of mine at the time, showed that this rule would apply also in the case where defective units were removed from the flow of product, but were not replaced with good units.

2. DISCUSSION. With the difficult mathematical proofs thus out of the way, the possibility of developing a simple algorithm presented itself [1]. The phases of inspection could be treated as states in a Markov chain. Remember that the UAOQL will result from a situation where for every occurrence of each phase, either only all nondefectives are submitted for inspection, or only all defectives are submitted for inspection.

Let us define configurations to be the values of

$y = (\phi(1), \ldots, \phi(k))$,

where

k = number of states

$\phi(j)$ = 0 if during occurrences of the phase represented by state j only nondefectives are submitted for inspection.

$\phi(j)$ = 1 if during occurrences of the phase represented by state j only defectives are submitted for inspection.

It is clear that for any plan of the type we are considering, then, there will be 2^k configurations. For even moderate sized values of k, the problem could be difficult if we had to consider every configuration. Fortunately, we can make the problem smaller.

Let us first go through the case where defective units are removed and replaced with good units.

Theorem 1: If a configuration exists such that for any state j

(i) $\phi(j) = 0$, and

(ii) State j is an absorbing barrier,

then this configuration need not be considered in determining the UAOQL.

Proof: Under the conditions stated in the theorem, the long run outgoing quality would be zero.

Theorem 2: If a configuration exists such that for any state j

(i) $\phi(j) = 1$, and

(ii) State j is an absorbing barrier,

then this configuration need not be considered in determining the UAOQL.

Proof : Under the conditions stated in the theorem, the long run outgoing quality would be zero.

We thus see that all configurations involving absorbing barriers can be disregarded.

Consider CSP-1. Let state 1 be the 100% inspection state and state 2 be the sampling state. We have the following configurations:

$$y_1 = (0, 0)$$

$$y_2 = (0, 1)$$

$$y_3 = (1, 0)$$

$$y_4 = (1, 1)$$

Configurations with $\phi(1) = 1$ or $\phi(2) = 0$ can be disregarded, since these would represent absorbing barrier situations. Therefore y_1, y_3, and y_4 can be disregarded. The remaining configuration, y_2, represents the situation under which the UAOQL occurs; no defective units are submitted during periods of 100% inspection, only defective units are submitted during periods of sampling inspection.

Let us now define another term.

A sequence of states which repeats itself indefinitely under the conditions imposed shall be called a cycle. For example, if a Markov chain consists of four states, and if a configuration results in a sequence of states (1, 2, 3, 4, 3, 4, 3, 4 ...), then (3, 4) is a cycle.

Theorem 3: The long run outgoing quality for a configuration involving cycles is equal to the average number of defectives passed in a cycle divided by the average number of units passed in a cycle.

Proof: The long run outgoing quality is

$$\lim_{m \to \infty} \frac{\sum_{i=1}^{m} \text{defectives passed in cycle } i}{\sum_{i=1}^{m} \text{units passed in cycle } i}$$

$$\lim_{m\to\infty} \frac{m(\text{AVERAGE NUMBER OF DEFECTIVES PASSED IN A CYCLE})}{m(\text{AVERAGE NUMBER OF UNITS PASSED IN A CYCLE})}$$

$$= \frac{\text{AVERAGE NUMBER OF DEFECTIVES PASSED IN A CYCLE}}{\text{AVERAGE NUMBER OF UNITS PASSED IN A CYCLE}}$$

Considering CSP-1 again, it has been shown that only configuration $y_2 = (0, 1)$ need be considered. Since the sequence of states (1, 2. 1, 2, ...) occurs, we may refer to (1, 2) as a cycle. Using Theorem 3, we may then say that

$$\text{UAOQL} = \frac{\text{AVERAGE NUMBER OF DEFECTIVES PASSED DURING 100\%}}{\text{AVERAGE NUMBER OF UNITS PASSED DURING 100\%}} + \frac{\text{AVERAGE NUMBER OF DEFECTIVES PASSED DURING SAMPLING}}{\text{AVERAGE NUMBER OF UNITS PASSED DURING SAMPLING}}$$

$$+ \frac{0 + (\frac{1}{f} - 1)}{i + (\frac{1}{f})} = \frac{1 - f}{fi + 1} ,$$

where i is the length of 100% inspection and f is the sampling frequency. Let us now consider the case where defective units found are removed but not replaced with good units.

Theorem 1': If a configuration exists such that for any state j

(i) $\phi(j) = 0$, and

(ii) State j is an absorbing barrier,

then this configuration need not be considered in determining the UAOQL.

Proof: Under the conditions stated in the theorem, the long run outgoing quality would be zero.

We see that this is the same as Theorem 1 for the replacement case.

Theorem 2': If a configuration exists such that for state 1 (corresponding to the first phase encountered in using the sampling plan)

(i) $\phi(1) = 1$, and

(ii) State 1 is an absorbing barrier,

Then this configuration need not be considered in determining the UAOQL.

Proof: Under the conditions stated in the theorem, no product would be passed at all, hence, outgoing quality would not be defineable.

Theorem 3': If the number of units passed in a cycle is greater than zero, then the long run outgoing quality for a configuration is equal to the average number of defectives passed in a cycle divided by the average number of units passed in a cycle.

Proof: Same as Theorem 3 for the replacement case.

Theorem 4': If a cycle passes zero units, it is only necessary, in determining the long run outgoing quality, to consider those states, if any, which occur before cycling begins.

Proof: The fraction defective of material passed by the inspection system would remain unchanged once cycling begins, since no more units would be passed. This theorem is useful when a 100% inspection state other than state 1 is an absorbing barrier.

As an example, let us consider the simple case of CPS-1 again under the nonreplacement assumption. We have the configurations

$$y_1 = (0, 0)$$

$$y_2 = (0, 1)$$

$$y_3 = (1, 0)$$

$$y_4 = (1, 1)$$

Configurations with $\phi(1) = 1$ or $\phi(2) = 0$ can again be disregarded, since these would represent absorbing barrier situations with no defective units passing. Again $y_2 = (0, 1)$ is the only configuration that need be considered. In the replacement case, then,

$$\text{UAOQL} = \frac{\text{AVERAGE NUMBER OF DEFECTIVES PASSED DURING 100\%}}{\text{AVERAGE NUMBER OF UNITS PASSED DURING 100\%}} + \frac{\text{AVERAGE NUMBER OF DEFECTIVES PASSED DURING SAMPLING}}{\text{AVERAGE NUMBER OF UNITS PASSED DURING SAMPLING}}$$

$$= \frac{0 + \left(\frac{1}{f} - 1\right)}{i + \left(\frac{1}{f} - 1\right)} = \frac{1 - f}{f(i - 1) + 1}$$

In our examples, we have used the simplest case, CSP-1. However, in practice, we have found that we can use this method for plans of some complexity in order to determine the UAOQL for either the replacement or the nonreplacement case.

REFERENCES

1. Brugger, Richard M., "The UAOQL of CSP-1, CSP-2, and CSP-R under the nonreplacement assumption, "QEM 21-240-9, Ammunition Procurement and Supply Agency, Joliet, Illinois, December 1967.

2. Dodge, Harold F., "A sampling inspection plan for continuous production", *Annals of Mathematical Statistics*, Vol 14, No.3, September 1943, pp 264-279.

3. Endres, Allen C., "The computation of the unrestricted AOQL when defective material is removed but not replaced," *Journal of the American Statistical Association*, Vol. 64, No. 326, June 1969, pp 665-668.

4. Lieberman, Gerald J., "A note on Dodge's continuous inspection plan", *Annals of Mathematical Statistics*, Vol. 24, No.3, September 1953, pp 480-483.

5. Sackrowitz, Harold, "A note on unrestricted AOQL's", *Journal of Quality Technology*, Vol. 7, No. 2, April 1975, pp 77-80.

6. White, Leon S., "Markovian decision models for the evaluation of a large class of continuous sampling inspection plans", *Annals of Mathematical Statistics*. Vol. 36, No. 5, October 1965, pp 1408-1420.

7. White, Leon S., "The evaluation of H106 continuous sampling plans under the assumption of worst conditions", *Journal of the American Statistical Association*, Vol. 61, No. 315, September 1966, pp 833-841.

SEMI MARKOV CHAINS APPLIED TO MARKOV CHAIN FUNCTIONALS PARTIALLY DEPENDENT ON RANDOM BACKWARD TIME SHIFTS

David L. Arp
Naval Weapons Center
China Lake, California

ABSTRACT. Given a Markov Chain (MC) model for a particular Continuous Sampling Plan (CSP), a method of restructuring its states into a simpler Semi Markov Chain (SMC) pattern is used to analyze MC functionals which are partially defined by random backward shifts in operational time.

Specifically, the usual MC model, for the Job Shop case of CSP-1, initially starts with an inspection phase of I states and thereafter cyclically alternates between it and a sampling phase. However, whenever sampling is terminated, this plan is modified by the additional requirement of a (limited) Downstream Inspection (DSI) of the previous I units followed by a phase transition determined by the outcome of such an inspection. For a production run of length N, this modification induces a corresponding one in the expected value of the associated basic functional: Fractio- Inspected [FI(N;1)]. Both modifications are handled here by 1.) slightly changing the usual SMC reduction and 2.) coupling this change with a new functional: Incremented Fraction Inspected [IFI(N;2)]. The expected value of the functional Total Fraction Inspected [TFI(N;2)] is then expressed as the expected value of the sum of two terms: the new functional and the (unmodified) functional, FI(N;2), defined on the altered SMC. In addition to comparing the long run expressions for TFI and AFI, a comparison is also made between TFI and the expression which results from the more familiar requirement of (limited) Upstream Inspection (USI).

In analyzing the above situation for finite N, two interpretations of DSI are subsequently studied. The first, based on possible inspection or manufacturing irregularities in both phases, is the scheme already referred to above. The second, based only on the putative assumption of sampling phase irregularities, is a less strict version. For N infinite, a comparison is made between the expected values of the two TFI's.

Since, in either of the two plans, TFI does not explicitly take account of multiple inspections of the same unit, other measures of plan performance are considered which do. To this end, the paper concludes with a study of the functional Fraction of Repetitions (FR), its first moment, and a variant functional. In order to deal with this functional, further modifications in the SMC model for CSP-1 are necessary.

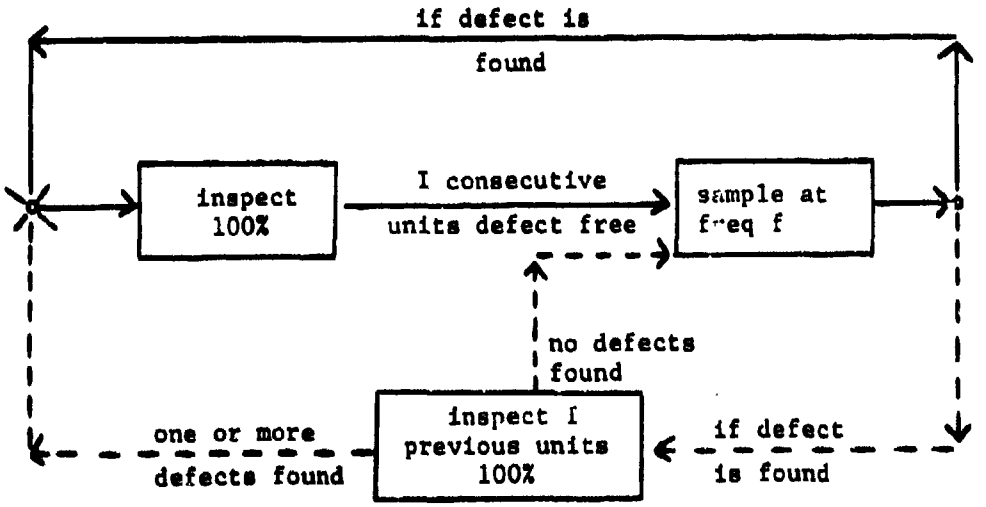
1.0 BACKGROUND.

1.1 Introduction. The principal subject of this paper is the study of variations in one member of a class of sampling plans and functionals defined on these variations. The class referred to is that of certain Continuous Sampling Plans (CSP) which are treated as finite state, irreducible, time homogeneous, and aperiodic Markov Chains (MC). The element referred to, classically denoted by CSP-1, is the simplest element of this class. In dealing with these MC models, four different kinds of standard groupings, called phases, of their states can be distinguished: screening (sc), unlimited sampling (uls), limited sampling (ls), and checking (ck). Using the terminology of phases, a given CSP can then be defined as a collection of two or more different phases (normally, one of which is sc) which are linked together in accordance with sampling frequency criteria. Throughout the bulk of the paper, only the two canonical phases making up CSP-1 will be considered; interest will be especially focused on structural changes in uls which are brought about by Downstream Inspection (DSI). At the end of Chapter 3, the checking phase will also be briefly considered since it can be regarded as Upstream Inspection (USI).

CSP-1 and the major variation in it, brought about by DSI, are portrayed in Figure 1.

Figure 1

CSP-1 and DSI



In Figure 1, CSP-1 consists of the top two boxes connected together with the solid lines. The DSI plan, denoted by CSP-12, is obtained from CSP-1 by replacing the top solid line by the dotted ones and adding the lower box. Two approaches will be used to handle this change.

The first approach, given in Chapters 2 and 3, consists in counting only the extra units inspected without regard to any inspection repetitions due to DSI. In the second approach, given in Chapter 4, all units inspected are also counted, but now including repetitions. Both approaches use, as the main tool, Semi Markov Chain (SMC) reduction of MC models which is now briefly described.

In describing SMC reduction, the term macrostate will be used to refer to an ensemble of MC states which is structured as a (discrete) SMC state (e.g., a canonical phase of a CSP). To be a macrostate, an ensemble must satisfy two conditions. 1.) The MC probability of entrance vector (pev) into the ensemble, given that such an entrance occurs, must be stationary and independent of the state from which the entrance is made. In other words, letting the ensemble S be composed of the k MC states, j, $1 \leq j \leq k$, we impose the condition that, for an arbitrary time n,

$$\underline{v}(n) = \underline{v}$$

where

$$\underline{v}(n) = (v_1(n), v_2(n), \text{---}, v_k(n))$$

$$v_j(n) = P[M(n) = j \,|\, M(n) \text{ in } S, M(n-1) \text{ not in } S]$$

and

$M(\cdot)$ is the MC process.

2.) Subject to the restrictions of 1.) for a given target macrostate, an exit can occur from a MC state of the ensemble into a MC state of the macrostate only if the first state communicates with the second in the underlying MC. To avoid a circular construction, we finally note that any MC state is, itself, a (trivial) macrostate.

Two different, but equivalent, methods can be used to construct such macrostates from MC states: the MC method, which is pedestrian, but straightforward, and the SMC method, which is more subtle but nearer to the general idea of SMC reduction. Under either method, MC functionals induce well defined SMC ones and the MC properties of time homogeneity, irreducibility, and aperiodicity are preserved [cf., 6.2 and 6.8].

In the MC method, the component states of a given macrostate and the possible exit macrostates are the transient states and the absorbing states, respectivaly, of an absorbing MC which is derived from a partitioning of the original MC. The possibly defective probability density function (pdf) of a transition of the macrostate to any one of the target macrostates is then just the weighted sum of the first entry probability functions, each weighed by the component in the stationary pev. In the more constructive SMC method, a given parent macrostate is considered to be made up of two or more smaller macrostates (including a MC state with or without self transitions). To such a division, the "MC method" is applied, only now to an absorbing SMC. The derived system of Backward Equations (see A.21), or, in simpler situations, direct combinatorial analysis is then applied to obtain the resulting first entry SMC probability functions. Their weighted sum, again weighed by components of the (induced) stationary pev, yields the pdf of the parent macrostate (to some one target macrostate). This latter method is easier to use and intuitively more appealing; it will be used almost exclusively throughout this paper except for a simple example of the MC method given at the end of Chapter 1. Furthermore, the SMC method, at any stage in its use, emphasizes the concepts 1.) of constructing from a given MC a class of SMC's which is partially ordered by filtration [6.2 and 6.7] and 2.) of using different elements of this class to attack either different problems or different stages of one problem which arise from the original MC.

Neither of these two methods should be confused with the process of lumping as it is defined in [6.13]. In fact, for CSP's, it is not possible to lump the states in each phase, in the above sense, into a new MC state. A more thorough presentation of SMC reduction, with many applications, can be found in [6.2]. What notation, definitions, and theorems concerning SMC's that are needed in this paper are taken from this reference and can be found in the Appendix. A more heuristic approach to SMC reduction (for the stationary case) together with further applications can be found in [6.4, 6.5, and 6.6] where it is called The Simplified Markov Chain Method.

In summary, the MC method can be stated as follows. Given the components, v_j, of the pev and the MC first entrance probability function

$$f^n_{j,A}$$

from j to a target macrostate (absorbing state) A, the equation for the pdf from S to A is (see Appendix for notation)

$$Q_{SA}(n) = \sum_{j=1}^{k} v_j f^n_{j,A} \cdot \qquad \text{(A1)}$$

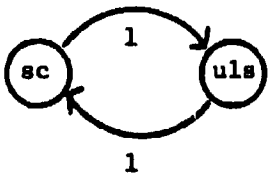
Similarly, the SMC method leads to the same form for the RHS of Eq. A1 in which the f's are replaced by SMC first entrance probability functions.

Another ubiquitous tool, used in concert with SMC reduction, is the z-transform. The transform is applied here to probability sequences rather than to the transitional matrices themselves. This approach is taken because, in practical applications, the ranks of the matrices are quite large (about 3×10^2 and greater). Thus the ranks of the complex functional matrices, obtained via the transform, would be so large that 1.) important relationships would be obscured and 2.) an analysis of them would be almost as difficult as that done without the transform. The salient features of the transform can be found in [6.3 and 6.12]. We record here only some basic notation that will be used with sequences treated as functions from the natural numbers to the reals. Given a sequence a(n), $\hat{a}(z)$ is its z-transform. Given sequences a(n) and b(n), a*b(n) is their convolution. $\delta_n(k)$ denotes the (Dirac) sequence which is one for the argument equal to n and zero otherwise; $\hat{\delta}_n(z) = 1/z^n$. $H_n(k)$ denotes the (Heaviside) sequence which is one for the argument greater than or equal to n and zero otherwise; $\hat{H}_n(z) = (1/z^n)(z/(z-1))$.

1.2 SMC(1) and FI(N;1). The basic premise used in modelling a CSP is that the underlying production process is a Bernoulli process with a constant probability of defective p (and probability of non-defective q = 1-p). In particular, the MC structure of CSP-1, which is fully described in [6.1, 6.2, and 6.4], arises from the sequential sampling scheme imposed on the above process with an operational time defined by the flow of non-repeating production units. The SMC model of CSP-1, derived from the MC model, is given in Figure 2 and is denoted by SMC(1).

Figure 2

SMC Model of CSP-1 (SMC(1))



For the model in Figure 2, I = clearance number for sc, f = sampling frequency for uls, p = probability of defective, q = 1-p, and we have the following statements expressed in

Theorem 1. Let sc = 1 and uls = 2. Then, SMC(1) is an irreducible SMC.

Proof. The SMC states are

$(1;\ \hat{Q}_{12}(z))$ and $(2;\ \hat{Q}_{21}(z))$,

where

$$\hat{Q}_{12}(z) = \frac{q^I(z-q)}{\phi(z)} \quad \text{and} \quad \hat{Q}_{21}(z) = \frac{\delta}{z-\beta} \tag{1.1}$$

In Eqs. 1.1, $\phi(z) = z^I(z-1) + \gamma, \gamma = pq^I$, $\delta = fp$, and $\beta = 1-\delta$.

The transitional matrix of the embedded MC is

$$\begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}$$

Even though it clearly has period 2, the SMC is none-the-less aperiodic [6.2 and 6.8]. It easily follows from the matrix that $\underline{e} = (1/2, 1/2)$ is the stationary (but not long run) vector.

Using the notation in the Appendix (see A.25),

$$\mu_1 = \frac{1-q^I}{pq^I} \quad \text{and} \quad \mu_2 = \frac{1}{fp}$$

The last two statements and A.25 imply

$$P_1(\infty;1) = \frac{\mu_1}{\mu_1+\mu_2} \quad \text{and} \quad P_2(\infty;1) = \frac{\mu_2}{\mu_1+\mu_2} \ .$$

Further details are found in [6.2] which finishes the proof.

The expressions $(1-q^I)/pq^I$ and $1/fp$, in Theorem 1, appear throughout the paper and will hereafter be abbreviated by the symbols μ_1' and μ_2', respectively. These special primed symbols are used to avoid confusion with standard notation (see A.25) and, at the same time, to serve as a reminder of their origin (ie, CSP-1).

The principal measure of plan performance for CSP-1 is the Fraction Inspected (FI) functional which is given in

Definition 1. For a production run of length N and sampling plan CSP-1,

$$FI(N;1) = 1 - \frac{\nu}{N} \sum_{j=0}^{N} C_2(j)$$

where $C_2(\cdot)$ is the characteristic function for state 2 = uls and $\nu = 1-f$.

Taking the expected value of the above functional, conditioned by an initial start in sc (Job Shop case), letting N approach infinity, and using the Ergodic Theorem, we have [6.1 and 6.2]

$$AFI(\infty;1) = 1-\nu P_2(\infty;1) \qquad (A2)$$

where the LHS of Eq. A2 is defined by

$$\lim_{N \to \infty} E_{sc}[FI(N;1)].$$

1.3 MC Method (An Example). The MC method will be briefly illustrated by applying it to the MC model of uls. This model consists of two MC states: SN, the non-inspection state and SI, the inspection state. The transitional matrix of the absorbing MC, derived from the MC model of any CSP having a uls phase, is

$$\begin{array}{c} \\ SN \\ SI \\ A \end{array} \begin{array}{c} \begin{array}{ccc} SN & SI & A \end{array} \\ \begin{bmatrix} \nu & f & 0 \\ q\nu & qf & p \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

where A is the only possible target phase to be entered. The pev for the ordered ensemble S = (SN, SI) is $\underline{y} = (\nu, f)$ which induces an initial probability vector $(\nu, f, 0)$ for the states (SN, SI, A), where A is the absorbing state, the other two being transient. Thus, from Eq. A1, we need to derive the expression

$$(\nu) f^{n}_{SN,A} + (f) f^{n}_{SI,A} \quad .$$

From the Chapman-Kolmogorov equation, a difference equation for the first entry probability functions can be derived. z-Transforming this difference equation, we obtain

$$\hat{Q}_{uls,A}(z) = \delta/(z-\beta)$$

where $\delta = fp$ and $\beta = 1-\delta$.

In a similar manner, $\hat{Q}_{sc,A}(z)$ can be derived using an (I+1) x (I+1) transitional matrix consisting of I transient and one absorbing states [6.2]. Also, for this latter function, see [6.10, Chp. 13] for a different derivation which is based on renewal theory and Bernoulli trials.

1.4 Notation and Terminology. Three essentially different plans will be studied in future chapters. They are denoted by CSP-12, CSP-13, and CSP-14. For ease in indexing functionals, CSP-1 will henceforth be denoted by CSP-11. SMC models associated with the above plans will be denoted by SMC(k), k a positive integer; in one case, a Markov Renewal Process (MRP) model is constructed for CSP-12 and is denoted by MRP(2). A MC state without self transitions will be called a trivial SMC state; one with self transitions will sometimes be considered as a (non-trivial) SMC state with a geometrically distributed holding time. A [functional] will usually mean E_{sc} [functional] for the models considered. In particular, with respect to some other set of models, A [·] could have an entirely different definition. Theorems, propositions, and definitions are numbered consecutively throughout the paper. Statement y of section x in the Appendix will be denoted by A.xy.

1.5 Acknowledgements. I would like to give credit to Mr. Richard M. Brugger who formulated the DSI concept and applied it, as a temporary measure, to tighten quality assurance in CSP-11 under practical shop conditions. This intriguing idea - inspecting against the natural flow of operational time - was originally communicated to the author by Mr. Gary Aasheim, an associate of Mr. Brugger. In closing this chapter, I would like to express my graditude to Leah K. Jones for the excellent typing of and editorial assistance on a fairly difficult subject.

1.6 Principal Results. For the quantities referenced below, $\delta = fp$, $\beta = 1-\delta$, and $\nu = 1-f$.

Eq. A2 gives AFI(∞;1); $P_2(\infty;1)$, μ'_1, and μ'_2 are given in Theorem 1.

Eq. B8 gives ATFI(∞;2); $P_2(\infty;2)$ is given in Theorem 4 and TFI is defined in Definition 2.

Eq. C8 gives ATFI(∞;3); $P_2(\infty;3)$ is given by Eq. C7.

Theorems 17 and 20 give AFR(∞;2); Theorem 18 gives AFR'(∞;2).

Theorem 7 compares ATFI(∞;2) and AFI(∞;1); Theorem 14 compares ATFI(∞;2) and ATFI(∞;3).

2.0 DSI - GENERAL. Having initially started in the screening phase (Job Shop case), if a defect is found in the sampling phase at time n, n > I, Downstream Inspection (DSI) requires 1.) a return to unit n-I with 100% inspection of the succeeding I units and 2.) entrance to the sampling phase (screening phase) if no (one or more) defects are found upon completion of 1.). DSI is portrayed in Figure 1, Chapter 1.

2.1 Introduction. If the DSI stage is, for the moment, intuitively looked on as a "pseudophase", the Total Fraction Inspected (TFI) can be obtained by treating it as a modification of FI(N;1). Conceptually, this modification can be broken down into two separate parts. The first is an additive fractional increase due to a sum each term of which, after multiplication by N, is equal to $\nu \min(k,I)$ where k+1 is the duration of the corresponding sampling phase segment. The second is a nonlinear decrease in FI(N;1) due to the transitional requiremtnts that come into force upon leaving the "pseudophase". The decrease occurs because, upon finding a defect, there is a chance of immediate (at least in the sense of operational time) return to the sampling phase rather than an automatic entrance to the screening phase which would otherwise take place in CSP-11. The finite probability of this immediate return results in a fractional increase in units not inspected and, therefore, a corresponding fractional decrease in units inspected.

These remarks lead to the following proposed solution. The nonlinear decrease can be dealt with by weaving the transitional requirements of the "pseudophase" into the SMC structure of CSP-11 thereby yielding a new SMC and its Fraction Inspected function, FI(N;2). The additive increase can then be easily handled by coupling a new Incremented Fraction Inspected functional, IFI(N;2), to FI(N;2). Adding these two functionals, we finally have

Definition 2. The **Total Fraction Inspected** is given by

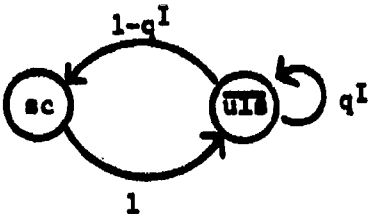
$$TFI(N;2) = FI(N;2) + IFI(N;2).$$

In this chapter, $ATFI(\infty;2)$ is found and compared with $AFI(\infty;h)$, $h = 1,4$. In Chapter 4, other functionals and SMC models are studied since the one considered here and its transient version, treated in Chapter 3, are not complete measures of plan performance.

2.2 MRP(2) and IFI(∞;2). For $ATFI(\infty;2)$, the solution proposed in the introduction suggests a model for CSP-12 given in Figure 3 and denoted by MRP(2). This model is a Markov Renewal Process whose definition is given in A.19 (also see A.28).

Figure 3

Model for CSP-12 (MRP(2))



Concerning the model in Figure 3, we have

Theorem 2. MRP(2) is a MRP. Letting sc = 1 and $\overline{uls}$ = 2, the states are

$$(1;\hat{Q}_{12}(z)) \text{ and } (2;\hat{Q}_{21}(z), \hat{Q}_{22}(z))$$

where

$$\hat{Q}_{12}(z) = \frac{q^I(z-q)}{\phi(z)} \qquad (2.1)$$

$$\hat{Q}_{21}(z) = \frac{\delta(1-q^I)}{z-\beta} \qquad (2.2)$$

and

$$\hat{Q}_{22}(z) = \frac{\delta q^I}{z-\beta} \qquad (2.3)$$

Proof. Eq. 2.1 follows from the model for CSP-11. Eqs. 2.2 and 2.3 follow from the model for CSP-11 and the introductory remarks to Chapter 2 since, upon completion of a DSI segment, the sampling phase (screening phase) is entered with probability q^I (probability $1-q^I$) with operational time playing no role. MRP(2) is a MRP by definition.

The definition of FI(N;2) is of the same form as that given for FI(N;1) in Definition 1. We now define the incremental functional in

Definition 4. Let $W(\cdot)$ be the following functional:

$$W(t) = \sum_{s=1}^{\ell} R_s + (1-C_2(t))R_{\ell+1}$$

where $\ell = N_2(t)-1$ and $R_s = \min(k,I)$ if the s^{th} exit from state 2 takes place $(k+1)$ time units from the s^{th} entrance. Then the Incremented Fraction Inspected functional for MRP(2) is

$$IFI(t;2) = \nu \frac{W(t)}{t}$$

where

$$\nu = 1-f.$$

Filtering out state 1 from MRP(2), we obtain the pdf of the renewal time for an occurrence of state 2 which is given by

$$Q'_{22}(t) = Q_{21}*Q_{12}(t) + Q_{22}(t).$$

Thus, averaging the time over one renewal cycle, we have

$$E[T] = \sum_{k=1}^{\infty} kP[T_{21} + T_{12} = k \text{ or } T_{22} = k]$$

$$= \sum_{k} k\, Q'_{22}(k)$$

$$= \sum_k k\, Q_{21} * Q_{12}(k) + \sum_k k\, Q_{22}(k).$$

From the mean value property of the z-transform, we must evaluate

$$-zD_z(\hat{Q}_{21}\hat{Q}_{12}) \text{ and } -zD_z(\hat{Q}_{22})$$

at $z = 1$. Calling the results of the evaluation m_1 and m_2, respectively, we have

$$m_1 = \frac{\bar{\delta}}{\delta}\ (\mu'_1 + \mu'_2) \text{ and } m_2 = \frac{q^I}{\delta}$$

where

$$\bar{\delta} = \delta(1-q^I).$$

Proposition 1.

$$E[T] = (1-q^I)\mu'_1 + \mu'_2 \qquad \text{(B1)}$$

Proof. See above.

Averaging $W(\cdot)$ over one cycle yields

$$E[W] = \sum_{k=1}^{I} kP[W=k]$$

$$= \sum_k k\delta\beta^k + I \sum_{j=0}^{\infty} \delta\beta^{I+1+j}$$

$$= \delta\beta D_\beta \left(\frac{1-\beta^{I+1}}{1-\beta}\right) + I\beta^{I+1}.$$

Since

$$D_\beta\left(\frac{1-\beta^{I+1}}{1-\beta}\right) = \frac{1}{\delta^2}\,(1-\beta^{I+1} - \delta(I+1)\beta^I),$$

substituting the RHS of this equation for the RHS above and simplifying, we have

Proposition 2.

$$E[W] = \frac{\beta}{\delta}\,(1-\beta^I) \qquad (B2)$$

Proof. From above.

We are now ready to prove

Theorem 3. For MRP(2),

$$IFI(\infty;2) = \nu\beta(1-\beta^I)\left(\frac{\mu_2'}{(1-q^I)\mu_1' + \mu_2'}\right),$$

[a.e.].

Proof. By the Strong Renewal Theorem [6.7, 6.9], we have

$$\lim_{N\to\infty} \frac{W(N)}{N} = \frac{E[W]}{E[T]}\ ,\ [a.e.].$$

The theorem follows from this result, Props. 1 and 2, Def. 3, and simplification.

Corollary.

$$AIFI(\infty;2) = \nu\beta(1-\beta^I)\left(\frac{\mu_2'}{(1-q^I)\mu_1' + \mu_2'}\right)$$

Proof.

$$AIFI(\infty;2) = \lim_{N\to\infty} \frac{E_{sc}[W(N)]}{N}\ .$$

2.3 SMC(2) and IFI(∞;2). By its very definition the functional $W(\cdot)$ depends on the sample paths of a MRP, including the self-transitions of a component state. In particular, the fundamental probability functions (see A.12) of the induced SMC (see A.28) are not sufficient to describe W since they don't record the self transitions of $\overline{uls}$. However, just as MRP(2) induces a unique SMC(2), W also induces a correspondingly unique functional, $W^*(\cdot)$, defined on the chain. We first prove

Theorem 4. MRP(2) induces a unique SMC, denoted by SMC(2).

Proof. From A.28, SMC(2) can be defined via its pdf's as follows [cf., 6.14]

$$\hat{Q}^*_{12}(z) = \hat{Q}_{12}(z) \tag{4.1}$$

and

$$\hat{Q}^*_{21}(z) = \left\{\sum_{j=0}^{\infty} \hat{Q}_{22}^j\right\} \hat{Q}_{21}$$

$$= \hat{Q}_{21}/(1-\hat{Q}_{22})$$

$$= \left(\frac{\overline{\delta}}{z-\beta}\right)\left(\frac{z-\beta}{z-(\beta+\delta q^I)}\right)$$

$$= \frac{\overline{\delta}}{z-\beta} \tag{4.2}$$

Recalling the definitions of μ_1 and μ_2 from Theorem 1, we have, from the derivatives of Eqs. 4.1 and 4.2,

$$\overline{\mu}_1 = \mu'_1 \quad \text{and} \quad \overline{\mu}_2 = \frac{1}{\overline{\delta}}$$

where

$$\overline{\delta} = \delta(1-q^I).$$

From A.25, we thus obtain

$$P_1(\infty;2) = \frac{(1-q^I)\mu_1'}{(1-q^I)\mu_1' + \mu_2'} \quad \text{and} \quad P_2(\infty;2) = \frac{\mu_2'}{(1-q^I)\mu_1' + \mu_2'} \tag{B3}$$

where 1 and 2 on the LHS's are sc and $\overline{uIs}^+$, respectively.

The transitional matrix and stationary vector of SMC(2) are the same as in Theorem 1 which finishes the proof.

Our goals now are to find $E[W^*]$ and $E[T^*]$. To this end, we prove

Theorem 5. The functional $W(\cdot)$ induces a well defined functional, $W^*(\cdot)$, on SMC(2).

Proof. W^* is implicitly defined through the following equations. Conditioning on the number of self transitions, j, of $\overline{uIs}$, we have

$$P[W^*=k] = \sum_{j=0}^{\infty} P[W^*=k|j]P[j]$$

$$= \sum_{j} a_j(k)$$

where

$$a_j(k) = P[W^*=k \text{ and } j \text{ repetitions}].$$

Noting that $a_j(k)$ can be defined in terms of $a_s(k)$, $s < j$, we can derive a set of equations relating the above a's. For ease in notation, we first define

$$\hat{B}(z) = \sum_{s=0}^{I-1} \left(\frac{\beta}{z}\right)^s .$$

Then, for $0 \le k \le (j+1)I$, k a fixed integer, we obtain the system given below.

$$a_j(k) = (\delta q^1) a_{j-1} * B(k) + (\beta q)^I a_j(k-I) \tag{5.1}$$

where

$$I \le k \le jI,$$

$$= (\delta q^I) a_{j-1} * B(k) \tag{5.2}$$

where

$$0 \le k < I,$$

$$= (\delta q^I)(\beta^{k-jI}) a_{j-1}(jI) \tag{5.3}$$

where

$$jI < k < (j+1)I, \text{ and}$$

$$= (\beta q^I) a_j(jI) \tag{5.4}$$

where

$$k = (j+1)I.$$

From Eq. 5.1, we have

$$\sum_{k=I}^{jI} \frac{a_j(k)}{z^k} = (\delta q^I) \sum_{k=I}^{jI} \frac{a_{j-1} * B(k)}{z^k}$$

$$+ (\beta q)^I \sum_{k=I}^{jI} \frac{a_{j-1}(k-I)}{z^k}$$

$$= X + Y \tag{5.5}$$

Adding zero on the RHS of Eq. 5.5 and changing indices in the term Y, we have

$$X + Y = X + \left(\frac{\delta q}{z}\right)^I \sum_{s=0}^{(j-1)I} \frac{a_{j-1}(s)}{z^s} + R - R$$

where

$$R = (\delta q^I) \sum_{k=0}^{I-1} \frac{a_{j-1} * B(k)}{z^k} \quad .$$

Grouping one R with X, using Eq. 5.2 to transform the second R, recalling that for j-1, $0 \leq k \leq jI$, and using the definition and convolutional property of the z-transform [6.3, 6.12], we have

$$\begin{aligned} \text{RHS}(5.5) &= (X+R) + Y - R \\ &= (\delta q^I)\hat{a}_{j-1}(z)\hat{B}(z) + Y \\ &\quad - \sum_{k=0}^{I-1} \frac{a_j(k)}{z^k} \quad . \end{aligned} \tag{5.6}$$

Again, using the definition of the z-transform, noting that in Y the sum is from 0 to (j-1)I while on the LHS(5.5) the sum varies from I to jI, and adding the last term of Eq. 5.6 to the LHS of Eq. 5.5, we obtain

$$\hat{a}_j(z) - \sum_s \frac{a_j(s)}{z^s} = (\text{LHS}(5.5) + R + S_j) - S_j,$$

where s varies from jI+1 to (j+1)I,

$$\begin{aligned} &= (X+R) + (Y+S'_{j-1}) - S'_{j-1} \\ &= (X+R) + \left(\frac{\delta q}{z}\right)^I \hat{a}_{j-1}(z) - S'_{j-1} \end{aligned} \tag{5.7}$$

where

$$S'_{j-1} = \left(\frac{\beta q}{z}\right)^I \sum_s \frac{a_{j-1}(s)}{z^s} ,$$

s varying from (j-1)I + 1 to jI.

From Eqs. 5.3 and 5.4, we find that

$$S'_{j-1} = \left(\frac{\beta q}{z}\right)^I S_{j-1}$$

$$= S_j .$$

Thus, the S_j term cancels out in Eq. 5.7 leaving us with the final equation

$$\hat{a}_j(z) = (\delta q^I)\hat{a}_{j-1}(z)\hat{B}(z) + \left(\frac{\beta q}{z}\right)^I \hat{a}_{j-1}(z) \qquad \text{(B4)}$$

Eq. B4 can now be solved iteratively, if desired, thereby proving Theorem 5.

W^* can also be explicitly defined in the same way as W (except that R_s^* can vary from zero to infinity). The importance of Theorem 5 is its use in Proposition 4.

Proposition 3. Let sc = 1 and $\overline{uls}^* = 2$. Then, we have for the renewal time, T^*, for state 2,

$$E[T^*] = \frac{(1-qI)\mu'_1 + \mu'_2}{(1-q^I)} \qquad \text{(B5)}$$

Proof.

$$E[T^*] = \sum_{k=0}^{\infty} kP[T^*=k]$$

$$= \sum_k kQ_{21}^* * Q_{12}(k)$$

$$= -zD_z(\hat{Q}_{21}^{*}\hat{Q}_{12}), \text{ at } z = 1.$$

Evaluating the last expression, we have the result.

Proposition 4. Averaging W^* over one renewal of state $\overline{uls}^*$, we have

$$E[W^*] = \frac{\beta(1-\beta^I)}{\delta(1-q^I)} \tag{B6}$$

Proof. The renewal time is given by T^* in Proposition 3 and has pdf $Q_{21}^{*}*Q_{12}$.

Summing $\hat{a}_j(z)$, in Theorem 5, from one to infinity, we have from Eq. B4

$$\hat{A}(z)-\hat{a}_0(z) = (\delta q^I)\hat{A}(z)\hat{B}(z) + \left(\frac{\beta q}{z}\right)^I \hat{A}(z) \tag{B7}$$

where

$$\hat{A}(z) = \sum_{j=0}^{\infty} \frac{a_j(z)}{z^j} .$$

From the mean value property of the z-transform and the definition of $\hat{a}_j(k)$, we thus have

$$E[W^*] = -zD_z\hat{A}(z) \text{ (at } z=1).$$

The proof is finished by evaluating the RHS of this last equation and simplifying.

We are now ready to prove the analogue of Theorem 3 (where the IFI functional is considered to be a quantity dependent on the plan but evaluated on the model) in

Theorem 6. For SMC(2),

$$IFI(\infty;2) = \nu\beta(1-\beta^I)P_2(\infty;2), \text{ [a.e.]}$$

where $2 = \overline{uls}^*$ and $P_2(\infty;2)$ is given in Theorem 4.

Proof. Again by the Strong Renewal (or Ergodic) Theorem,

$$\lim_{N\to\infty} \frac{W^*(N)}{N} = \frac{E[W^*]}{E[T^*]}, \quad [a.e.]$$

$$= \frac{\beta(1-\beta^I)}{\delta(1-q^I)} \quad \frac{(1-q^I)}{(1-q^I)\mu_1' + \mu_2'}.$$

from Eqs. B5 and B6,

$$= \beta(1-\beta^I) \quad \frac{\mu_2'}{(1-q^I)\mu_1' + \mu_2'}$$

$$= \beta(1-\beta^I)P_2(\infty;2),$$

from Eq. B3.

Multiplying by ν finishes the proof.

In particular, the equations in Theorems 3 and 6 agree, as they should.

Corollary.

$$AIFI(\infty;2) = \nu\beta(1-\beta^I)P_2(\infty;2)$$

Proof. The same as in the Corollary to Theorem 3.

2.4 TFI(∞;2) and Comparisons. Given the real number p varying over the open unit interval, the inequality "$1-q^I < 1$", Theorem 1, and Theorem 4 imply

$$P_2(\infty;1) < P_2(\infty;2)$$

for SMC(1) and SMC(2). We shall show a similar result for AFI(∞;1) and ATFI(∞;2). Before doing this, we record the following result

$$ATFI(\infty;2) = AFI(\infty;2) + AIFI(\infty;2)$$

$$= (1-\nu P_2(\infty;2)) + \nu\beta(1-\beta^I)P_2(\infty;2)$$

$$= 1-\nu P_2(\infty;2)(\delta+\beta^{I+1}) \qquad \text{(B8)}$$

Theorem 7. For p in the open interval $0 < p < 1$,

$$ATFI(\infty;2) \geq AFI(\infty;1) \text{ iff } \beta(1-\beta^I) \geq q^I\alpha_1'$$

where $\alpha_1' = \mu_1'/(\mu_1' + \mu_2')$.

Proof. From Eqs. A2 and B8, the statement is equivalent to

$$P_2(\infty;1) \geq P_2(\infty;2)(\delta + \beta^{I+1}).$$

This inequality is, in turn, equivalent to

$$(\delta + \beta^{I+1})(\mu_1' + \mu_2') \leq (1-q^I)\mu_1' + \mu_2'$$
$$= (\mu_1' + \mu_2') - q^I\mu_1'$$

Dividing through by $(\mu_1' + \mu_2')$, we have

$$(\delta + \beta^{I+1}) \leq 1-q^I\alpha_1'$$

or

$$1-(\delta + \beta^{I+1}) \geq q^I\alpha_1'.$$

However,

$$1-(\delta + \beta^{I+1}) = \beta(1-\beta^I).$$

Thus,

$$\beta(1-\beta^I) \geq q^I\alpha_1'$$

which finishes the proof.

For $p = 0$ or 1, the formulas in Theorem 7 are equal.

Another type of CSP, denoted here by CSP-14, is the plan obtained from replacement of DSI in CSP-11 by USI. For CSP-14, the SMC model is straightforward since the limited inspection scheme runs with the natural flow of operational time. For this model, we have

Proposition 5. Letting sc = 1, uls = 2, and ck (or USI) = 3,

$$P_2(\infty;4) = \frac{\mu_2'}{(1-q^I)\mu_1' + \mu_2' + I}$$

Proof. If $\underline{e}$ is the stationary vector, using the SMC model for the ck phase found in [6.2], we have

$$(1-q^3)\underline{e} = (1-q^I, 1, 1).$$

The rest of the proof easily follows from A.25 given that $\mu_3 = I$.

It clearly follows from Proposition 5 that

$$AFI(\infty;4) = 1-\nu P_2(\infty;4)$$

Thus, to compare AFI(∞;4) and ATFI(∞;2), it would suffice to compare the expressions which are analogous to those in Theorem 7. However, to avoid a long proof, it also suffices to give the following probabilistic argument.

Upon finding a defect in the sampling phase, I new units are inspected with CSP-14 while, on the other hand, at most νI new units are inspected under CSP-12. Since the transitional probabilities are the same from the limited inspection (pseudo) phase in both plans, the proof is finished.

3.0 DSI - TRANSIENT. Two interpretations of DSI for the transient case are treated in this chapter. The first version is the transient case of DSI, already dealt with in Chapter 2 for infinite N. That is, DSI is applied to both phases of CSP-11 with constant "pseudophase" transitional probabilities. In contrast to the first version, the second plans "pseudophase" transitional probabilities to sc (or uls) monotonically increase (or decrease) with increasing duration in the sampling phase, until truncated by $1-q^I$ (or q^I). One can infer from this monotonicity that DSI is applied only to the sampling phase in the following sense. If a defect is found during a sampling segment, k + 1 time units from entrance to this particular segment, then only the previous τ units are to be inspected, where $\tau = \min(k, I)$. Upon completion of this modified DSI, uls is entered if no defects are found (with probability q^k); otherwise, sc is entered (with probability $1-q^k$).

3.1 Introduction. The analysis of each version involves three stages. However, for convenience in the final section, a fourth stage is added for the second version.

In the primary stage, the modified sampling phase is partitioned into I + 2 SMC states which are consecutively labelled 0 through I and b. The purpose of this splitting is the derivation of an expression for the monotonically increasing portion of the functional W(·).

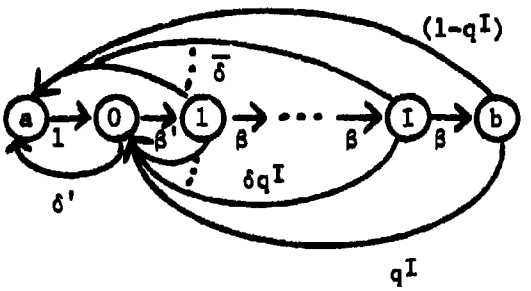
In the secondary and tertiary stages, SMC states 1 through I are recombined into a preliminary macrostate, c'; it, in turn, is combined with SMC state 0 to form the final SMC state, c. The purpose of these latter two manipulations is to facilitate the derivation of an expression for the truncated portion of W(·) by avoiding complex sums of products of characteristic functions.

The chapter concludes with a comparison between the TFI functional of each version for infinite N (or t).

3.2 Strict DSI. In order to analyze the transient case of DSI, the SMC model, shown in Figure 4, is used. It is denoted by SMC(3).

Figure 4

Model for CSP-12 (SMC(3))



$sc = a$

$\bar{\delta} = \delta(1-q^I)$, $r = 1-\delta q^I$, $\delta' = \bar{\delta}/r$, and $\beta' = \beta/r$

Concerning this model, we have

Theorem 8. SMC(3) is an irreducible SMC.

Proof. The z-transformed pdf's of the states making up SMC(3), together with their corresponding transitional probabilities in the embedded MC, are given below.

$$\hat{Q}_{k,k+1}(z) = \beta/z,\ q_{k,k+1} = \beta,\ \text{for } 1 \le k \le I-1$$

$$\hat{Q}_{ka}(z) = \bar{\delta}/z,\ q_{ka} = \bar{\delta},\ \text{for } 1 \le k \le I$$

$$\hat{Q}_{k0}(z) = \delta q^I/z,\ q_{k0} = \delta q^I,\ \text{for } 1 \le k \le I$$

$$\hat{Q}_{0a}(z) = \bar{\delta}/(z-\delta q^I),\ q_{0a} = \delta'$$

$$\hat{Q}_{01}(z) = \beta/(z-\delta q^I),\ q_{01} = \beta'$$

$$\hat{Q}_{a0}(z) = q^I(z-q)/\phi(z),\ q_{a0} = 1$$

$$\hat{Q}_{ba}(z) = \bar{\delta}/(z-\beta),\ q_{ba} = 1-q^I$$

$$\hat{Q}_{b0}(z) = \delta q^I/(z-\beta),\ q_{b0} = q^I$$

The equations follow from SMC(2) by observing that $\overline{uIs}^*$, since its holding time pdf is geometric, can be regarded as a MC state which jumps to itself and sc with probabilities $\bar{\beta}$ and $\bar{\delta}$, respectively.

Ordering the states of SMC(3) in the same manner, from left to right, as they are ordered in Figure 4, we obtain the linear system of equations from the matrix equation $\underline{e} = \underline{e}T$, T the embedded MC transitional matrix.

$$\delta' e_0 + \bar{\delta} \sum_{j=1}^{I} e_j + (1-q^I)e_b = e_a \qquad (8.1)$$

$$e_a + \delta q^I \sum_{j} e_j + q^I e_b = e_0 \qquad (8.2)$$

$$\beta' e_0 = e_1$$

for $1 \le k \le I-1$,

$$\beta e_k = e_{k+1}$$

and

$$\beta e_I = e_b$$

From this system, exclusive of Eqs. 8.1 and 8.2, we obtain

$$e_k = \beta^{k-1}\beta' e_0, \quad 1 \le k \le I \tag{8.3}$$

and

$$e_b = \beta^I \beta' e_0 \tag{8.4}$$

Eqs. 8.1, 8.3, and 8.4 imply

$$\delta' e_0 + \beta'(1-q^I)(1-\beta^I)e_0 + (1-q^I)e_0 = e_a$$

or

$$\frac{1-q^I}{1-\delta q^I} e_0 = e_a \tag{8.5}$$

Since $\underline{e}$ is normalized, we have from the sum of its components, Eqs. 8.3, 8.4, and 8.5

$$e_0 = \frac{\delta(1-\delta q^I)}{G} \tag{8.6}$$

where

$$G = (1+\delta)(1-\delta q^I) - \beta^{I+2}.$$

Thus, Eqs. 8.3, 8.4, 8.5, and 8.6 imply

$$e_a = \frac{\delta(1-q^I)}{G}, \quad e_0 = \frac{\delta(1-\delta q^I)}{G}$$

$$e_b = \frac{\delta\beta^{I+1}}{G} \text{ , and } e_k = \frac{\delta\beta^k}{G}$$

where $1 \le k \le I$.

Differentiating the $\hat{Q}$'s, multiplying through by minus one, and evaluating the results at $z = 1$, we have (adding terms where appropriate)

$$\mu_a = \frac{(1-q^I)}{\gamma} \text{ , } \mu_0 = \frac{1}{1-\delta q^I}$$

$$\mu_b = \frac{1}{\delta} \text{ , and } \mu_k = 1$$

for $1 \le k \le I$.

This finishes Theorem 8.

Corollary. For SMC(3),

$$\alpha_0 + \sum_{j=1}^{I} \alpha_j + \alpha_b = P_2(\infty;2) \qquad (8.7)$$

where the LHS refers to SMC(3) and the RHS refers to SMC(2).

Proof.

$$\mu_0 e_0 + \sum_{k=1}^{I} \mu_k e_k + \mu_b e_b = (\delta+\beta)/C$$

$$= 1/G$$

$$\mu_a e_a + 1/G = \left(\frac{1-q^I}{\gamma}\right)\left(\frac{\bar{\delta}}{G}\right) + \frac{\gamma}{\gamma G}$$

$$= \frac{1}{GP_2(\infty;2)}$$

Thus

$$\text{LHS (8.7)} = GP_2(\infty;2)(1/G)$$

$$= P_2(\infty;2)$$

Relative to SMC(3), we have

Definition 5. The monotonically increasing portion of W(t), divided by t, and considered as being defined on SMC(3) is

$$\frac{W'(t)}{t} = \frac{\sum_{n=0}^{t-1} \sum_{k=1}^{I} kC_k(n)(1-C_{k+1}(n+1))}{t} .$$

Thus we can also write

$$IFI'(t;2) = \nu \frac{W'(t)}{t} .$$

Operating on this equation and the RHS of the equation in Definition 5 with $E_{sc}[\cdot]$, we obtain

$$AIFI'(t;2) = \delta \frac{\sum_{n=0}^{t-1} \sum_{k=1}^{I} kP_{ak}(n)}{t} \qquad (C1)$$

which can be evaluated by using the z-transformed Backward Equations for SMC(3); see [6.1] for an example of such an evaluation. Letting t approach infinity, we have

$$AIFI'(\infty;2) = \delta \sum_{k=1}^{I} k\alpha_k \qquad (C2)$$

Since, from the last part of Theorem 8 and from A.25

$$\alpha_k = \left(\frac{\delta\beta^k}{G}\right)(\mu_k)(GP_2(\infty;2))$$

$$= \delta\beta^k P_2(\infty;2)$$

and

$$a_b = \left(\frac{\delta\beta^{I+1}}{G}\right)(\mu_b)(GP_2(\infty;2))$$

$$= \beta^{I+1}P_2(\infty;2)$$

we have, from Eq. C2,

$$\delta \sum_k k a_k = \beta\delta^2 P_2(\infty;2)\, D_\beta\left(\frac{1-\beta^{I+1}}{1-\beta}\right)$$

$$= \beta(1-\beta^I-\delta I\beta^I)P_2(\infty;2)$$

From A.27,

$$W''(b)a_b = \delta I a_b$$

where $W''(t)$ is the constant part of $W(t)$. Therefore, adding the last two expressions and performing the indicated operations, we have

$$\lim_{t\to\infty} \frac{E[W(t)]}{t} = \beta(1-\beta^I)P_2(\infty;2),$$

a result which agrees with that obtained in Chapter 2.

In order to deal with the constant part of the functional for finite t, we proceed to reduce SMC(3) to a more manageable model as described in Section 3.1.

Stage two consists in filtering out the states 1 through I in SMC(3), an operation which leads to a new model: SMC(4). The details and results of collapsing SMC(3) into SMC(4) are given in

Theorem 9. SMC(4) is an irreducible SMC obtained from SMC(3).

Proof. Let c' be the ordered ensemble composed of the states 1 through 1. Noting that the pev of c' is (1,0,0,---,0), I-1 zeros, we

apply combinatorial analysis to get (dropping the argument s)

$$\hat{Q}_{c'0} = \hat{Q}_{10} + \hat{Q}_{12}\hat{Q}_{20} + \text{---} + \hat{Q}_{12}\hat{Q}_{23}\cdot\text{----}\cdot\hat{Q}_{I0}$$

$$= \frac{\delta q^I}{z} \sum_{j=0}^{\tau-1} \left(\frac{\beta}{z}\right)^j$$

$$= \frac{\delta q^I}{z-\beta} \left(1-\left(\frac{\beta}{z}\right)^I\right) \qquad (9.1)$$

In the same way, we also obtain

$$\hat{Q}_{c'a} = \frac{\delta(1-q^I)}{z-\beta} \left(1-\left(\frac{\beta}{z}\right)^I\right) \qquad (9.2)$$

and

$$\hat{Q}_{c'b} = \left(\frac{\beta}{z}\right)^I \qquad (9.3)$$

The remaining results concerning SMC(4) can be easily derived from the above equations. In particular, see A.29.

Corollary.

SMC(4) < SMC(3)

where "<" is the filtration ordering relation.

Proof. SMC(4) is a filtration of SMC(3) by the proof of Theorem 9 and A.29.

Stage three consists in filtering out state c' in SMC(4) yielding SMC(5). The details and results are given in

Theorem 10. Filtering out c' in SMC(4) yields a new SMC, denoted by SMC(5).

Proof. Let the ordered ensemble (0,c') be denoted by c. Then, the pev for c is the vector (1,0).

First construction. Applying combinatorial analysis to the transformed pdf's in Theorem 9 (Eqs. 9.1, 9.2, and 9.3), we have

$$\hat{Q}_{ca} = \left\{\sum_{j=0}^{\infty} (\hat{Q}_{0c'}\hat{Q}_{c'0})^j\right\} \hat{Q}_{0a}$$

$$+ \left\{\sum_{j=0}^{\infty} (\hat{Q}_{0c'}\hat{Q}_{c'0})^j\right\} \hat{Q}_{0c'}\hat{Q}_{c'a}$$

$$= \frac{\hat{Q}_{0a} + \hat{Q}_{0c'}\hat{Q}_{c'a}}{1-\hat{Q}_{0c'}\hat{Q}_{c'0}}$$

$$= \frac{\overline{\delta}(z^{I+1}-\beta^{I+1})}{c(z)} \qquad (C3)$$

where

$$c(z) = z^{I+1}(z-(\beta + \delta q^I)) + \delta\beta(q\beta)^I .$$

Similarly,

$$\hat{Q}_{cb} = \left\{\sum_{j} (\hat{Q}_{0c'}\hat{Q}_{c'0})^j\right\} \hat{Q}_{0c'}\hat{Q}_{c'b}$$

$$= \frac{\hat{Q}_{0c'}\hat{Q}_{c'b}}{1-\hat{Q}_{0c'}\hat{Q}_{c'0}}$$

$$= \frac{(z-\beta)\beta^{I+1}}{c(z)} \qquad \text{(C4)}$$

Second construction. Since SMC(5) is the model to be used in deriving an expression for the constant part of IFI(t;2), t finite, we will sketch the more elaborate SMC method. The relevant absorbing SMC has transient states 0 and c'; absorbing states a and b. Using A.21, setting a = A, b = B, and c' = 1, we obtain the following transformed Backward Equations (four others, not needed, are omitted).

$$\hat{P}_{0A} = \hat{Q}_{01}\hat{P}_{1A} + \hat{Q}_{0A}\hat{P}_{AA}$$

$$\hat{P}_{1A} = \hat{Q}_{10}\hat{P}_{0A} + \hat{Q}_{1A}\hat{P}_{AA}$$

$$\hat{P}_{AA} = \hat{H}_0$$

$$\hat{P}_{0B} = \hat{Q}_{01}\hat{P}_{1B}$$

$$\hat{P}_{1B} = \hat{Q}_{10}\hat{P}_{0B} + \hat{Q}_{1B}\hat{P}_{BB}$$

$$\hat{P}_{BB} = \hat{H}_0$$

Solving for $\hat{P}_{0A}$ in the first set of three,

$$\hat{P}_{0A} = \frac{\hat{H}_0(\hat{Q}_{0A} + \hat{Q}_{01}\hat{Q}_{1A})}{1-\hat{Q}_{01}\hat{Q}_{10}} \quad .$$

Since the pev of the ordered ensemble (0,1) is (1,0), the above equation, Eq. A1, A.13, and A.22 imply

$$\hat{Q}_{ca} = \hat{F}_{0A}$$

$$= \hat{P}_{0A}/\hat{H}_0$$

$= \text{Eq. C3}$

Solving for $\hat{P}_{0B}$ in the second set of three,

$$\hat{P}_{0B} = \frac{\hat{H}_0(\hat{Q}_{01}\hat{Q}_{1B})}{1-\hat{Q}_{01}\hat{Q}_{10}} .$$

Again, since the pev = (1,0), the above equation together with Eq. A1, A.13 and A.22 imply

$$\hat{Q}_{cb} = \hat{F}_{0B}$$

$$= \hat{P}_{0B}/\hat{H}_0$$

$$= \text{Eq. C4.}$$

SMC(5) has three states: a, c, and b. The transformed pdf's for transitions of a to c, b to c, and b to a are the same as those for a to 0, b to 0, and b to a, respectively, in SMC(4).

We finish the proof of Theorem 10 by remarking that states a and c cannot be combined since a pev (from state b) does not exist.

Corollary.

SMC(5) < SMC(4).

Proof. Construction of the state c in SMC(5) is equivalent to filtering out state c' in SMC(4). SMC(5) is an irreducible SMC by A.29.

We can now derive an expression for the constant part of IFI(t;2) in

Theorem 11. Given the 3 state model, SMC(5),

$$IFI''(t;2) = \nu I\left\{\frac{N_b(t)-C_b(t)}{t}\right\} \qquad (C5)$$

Proof. $N_b(t)$ gives the number of entrances to state b by time t. The number of exits from state b is clearly $N_b(t)-C_b(t)$, the second term being the characteristic function of state b.

Corollary.

$$AIFI''(t;2) = \nu I \left\{ \frac{E_a[N_b(t)]}{t} - \frac{F_{ab}(t)}{t} \right\} \qquad (C6)$$

Proof. Apply $E_a[\cdot]$ to Eq. C5.

In order to use Eq. C6, we must be able to develop a useable expression for the mean of the renewal function. Towards this end, we prove

Proposition 6. Let N(t) be a renewal process. Then

$$E[N(t)] = H_0 * F * \left(\sum_{j=0}^{\infty} F^{(j)} \right)$$

where F is the renewal pdf.

Proof.

$$P[N(t) = n] = P[U(n+1) > t] - P[U(n) > t]$$

$$= H_0 * F^{(n+1)}(t) - H_0 * F^{(n)}(t)$$

$$= P_n(t)$$

Thus,

$$\hat{P}_n(z) = \hat{H}_0 (\hat{F})^n (1-\hat{F})$$

Therefore,

$$\sum_{n=0}^{\infty} \frac{\hat{P}_n(z)}{s^n} = \hat{H}_0 (1-\hat{F}) \left\{ 1 + \sum_n \left(\frac{\hat{F}}{s} \right)^n \right\}$$

$$= \frac{s\hat{H}_0(1-\hat{P})}{(s-\hat{P})}$$

$$= \hat{P}(z,s)$$

From the last function, we have

$$-s\,\frac{\partial \hat{P}(z,s)}{\partial s} \quad (\text{at } s = 1) = \frac{\hat{H}_0\hat{P}}{(1-\hat{P})} \quad .$$

The LHS is the transform of the mean and we are done.

Corollary 1.

$$\frac{E_a[N_b(t)]}{t} = \frac{H_0 * F_{ab} * (1-F_{bb})^{-1}(t)}{t}$$

where the inverse expression is shorthand for the summation.

Proof. Renewals of state b, starting in state a, form a delayed renewal process with initial probability function F_{ab}. Then Proposition 6 finishes the proof.

Corollary 2.

$$\lim_{t\to\infty} \frac{E_a[N_b(t)]}{t} = \delta\alpha_b$$

Proof. From Corollary 1 above, we have

$$\lim_{t\to\infty} \frac{E_a[N_b(t)]}{t} = \lim_{t\to\infty} \frac{H_0 * S(t)}{t}$$

$$= \lim_{t\to\infty} \frac{S(t)}{t} \quad ,$$

where $S(t) = H_0 * F_{ab} * (1-F_{bb})^{-1}(t)$,

$$= \lim_{z \to 1} \left(\frac{z-1}{z}\right) \frac{\hat{F}_{ab}}{(1-\hat{F}_{bb})}$$

$$= \frac{\hat{F}_{ab}(z)}{-zD_z F_{bb}(z)} \quad (\text{at } z = 1)$$

$$= \delta a_b .$$

The second equality follows from the simple argument that if $S(\cdot)$ is a sequence with limit A, then the Cesàro limit of $S(\cdot)$ also exists and is equal to A.

From the second corollary to Proposition 6, we have in addition

$$\nu I \left\{ \frac{E_a[N_b(t)]}{t} - \frac{P_{ab}(t)}{t} \right\} \longrightarrow \nu I \delta a_b$$

as t approaches infinity, since the second term goes to zero.

The main results about IFI(t;2) are summed up in

Theorem 12. For the transient case of CSP-12, we have

$$AIFI(t;2) = AIFI'(t;2) + AIFI''(t;2)$$

$$= \frac{\nu\delta \sum\sum k P_{ak}(n)}{t} + \nu I \left\{ \frac{E_a[N_b(t)] - P_{ab}(t)}{t} \right\}$$

where the first and second terms on the RHS are evaluated using SMC(h), h = 3 and 5, respectively.

Proof. Combine Eqs. C2 and C5 (taking the limit, we get ν times the result using W' and W").

When t is finite, in order to compute $E_a[N_b(t)]$, we need to know $F_{ab}(t)$ and $F_{bb}(t)$. Since SMC(5) has 3 states, we have 9 Backward Equations, only one of which is needed for the mean value of the above renewal function. The following statements sketch the results.

From Theorem 10, A2.1, and A1.4, we have

$$\hat{P}_{bb} = \hat{Q}_{bc}\hat{P}_{cb} + \hat{Q}_{ba}\hat{P}_{ab} + \hat{J}_b$$

This equation is equivalent to

$$1 = \hat{Q}_{bc}\left(\frac{\hat{P}_{cb}}{\hat{P}_{bb}}\right) + \hat{Q}_{ba}\left(\frac{\hat{P}_{ab}}{\hat{P}_{bb}}\right) + \frac{\hat{J}_b}{\hat{P}_{bb}}$$

or

$$1 - \frac{\hat{J}_b}{\hat{P}_{bb}} = \hat{Q}_{ba}\hat{F}_{ab} + \hat{Q}_{bc}\hat{F}_{cb}$$

But, LHS = $\hat{F}_{bb}$. Therefore,

$$\hat{F}_{bb} = \hat{Q}_{ba}\hat{F}_{ab} + \hat{Q}_{bc}\hat{F}_{cb}.$$

From Theorem 10,

$$\hat{Q}_{bc} = \frac{\delta q^I}{z-\beta} \text{ and } \hat{Q}_{ba} = \frac{\delta(1-q^I)}{z-\beta}.$$

Applying combinatorial analysis to the transformed pdf's of SMC(5), we have

$$\hat{F}_{cb} = \left\{\sum_j (\hat{Q}_{ca}\hat{Q}_{ac})^j\right\}\hat{Q}_{cb}$$

$$= \hat{Q}_{cb}/(1-\hat{Q}_{ac}\hat{Q}_{ca})$$

and

$$\hat{\vartheta}_{ab} = \left\{\sum_j (\hat{Q}_{ac}\hat{Q}_{ca})^j\right\}\hat{Q}_{ac}\hat{Q}_{cb}$$

$$= \hat{Q}_{ac}\hat{Q}_{cb}/(1-\hat{Q}_{ac}\hat{Q}_{ca}) .$$

From these equations, $E[N_b(t)]/t$ can be computed [cf., 6.1].

The use of SMC(3) suggests the following alternative treatment of CSP-12. Instead of splitting $\overline{uIs}^*$ into I + 2 states, we split it into an infinite number by splitting state b into the states b(j), $1 \le j \le \infty$. The resulting model, SMC(6), consists of two nontrivial SMC states (a and 0) and an infinite number of trivial SMC(ie, MC) states (1 through I and the b(j)'s). For the long run case, we can obtain AIFI(∞;2) via the transient case as shown in

Proposition 7. SMC(6) is an infinite state, irreducible, and positive recurrent SMC. The result for IFI(t;2) for SMC(6) is the same as previous results.

Proof. For b(j), $1 \le j \le \infty$, we have

$$\alpha_{b(j)} = \delta\beta^{I+j} P_2(\infty;2) \qquad (7.1)$$

and

$$\mu_{b(j)} = 1.$$

Thus $\mu_{b(j)b(j)} = 1/\alpha_{b(j)}$ which is finite, proving the chain positive recurrent.

For the functional, it suffices to deal with the part defined on the b(j)'s, W'''.

$$\frac{W'''(t)}{t} = \frac{\sum_{n=0}^{t-1}\sum_{j=0}^{\infty} C_{b(j)}(n)(1-C_{b(j+1)}(n+1))}{t}$$

Taking the mean value, conditioned by an initial entrance from state a,

$$\frac{E_a[W'''(t)]}{t} = \frac{\sum_n \sum_j \delta P_{ab(j)}(n)}{t}$$

which, as t approaches infinity, approaches

$$\delta^2 \beta^{I+1} P_2(\infty;2) \sum_{j=0}^{\infty} \beta^j$$

$$= \delta a_b$$

by Eq. 7.1 and the Lebesque Dominated Convergence Theorem (for sequences).

Proposition 8. The models used for CSP-12 are ordered, w.o. filtration, as follows.

$$SMC(2) < SMC(5) < SMC(4) < SMC(3)$$

and

$$SMC(5) < SMC(6)$$

Proof. Corollaries to Theorems 9 and 10 imply the first ordering. By filtering out states b(j), $j \geq 2$, we get the second ordering.

If we split state a into its component MC states and state 0 into a MC state in SMC(6), we get (S)MC(7) > SMC(5), SMC(3). If we instead split a and 0 as before but now split b by treating it as a MC state, we get (S)MC(8) > SMC(5), SMC(3). Clearly, (S)MC(7)>(S)MC(8). MC(8) can be thought of as a finite state MC model which fills the role of the initial MC model described in the introduction to Chapter 1, though the construction is backwards from that description.

3.3 Liberal DSI. To obtain a more liberal DSI, we alter the following transformed pdf's for states 0 through I-1 in

Theorem 13. The DSI sampling plan CSP-13 is obtained from the SMC(3) model of CSP-12. The result, SMC(3), is an irreducible SMC.

Proof. The appropriate quantities and properties are given below.

$$\hat{Q}_{0a} = 0,\ q_{0a} = 0$$

$$\hat{Q}_{01} = \beta/(z-\delta),\ q_{01} = 1$$

$$\hat{Q}_{ka} = \overline{\delta}/z,\ q_{ka} = \overline{\delta}$$

$$\hat{Q}_{k0} = \delta q^k/z,\ q_{k0} = \delta q^k$$

where $1 \le k \le I-1$

The other transformed pdf's remain the same as those for SMC(3).

Ordering the states a, 0, 1, ---, I, and b, we obtain, from the stationary vector equation, the system of equations now given.

$$\delta \sum_{j=1}^{I} (1-q^j)e_j + (1-q^I)e_b = e_a \qquad (13.1)$$

$$e_a + \delta \sum_{j} q^j e_j + q^I e_b = e_0 \qquad (13.2)$$

$$\beta^{k-1} e_0 = e_k \qquad (13.3)$$

where $1 \le k \le I$, and

$$\beta^I e_0 = e_b \qquad (13.4)$$

From Eqs. 13.2, 13.3, and 13.4, we get

$$e_a = \frac{p(1-(\beta q)^I)}{1-(\beta q)} e_0 \qquad (13.5)$$

Since the components of $\underline{e}$ are normalized, we obtain, together with Eq. 13.5,

$$e_0 = \frac{\delta(1-\beta q)}{G} \qquad (13.6)$$

where

$$G = \delta p(1-(\beta q)^I) + (1-\beta q)(1+\delta-\beta^{I+1})$$

Eqs. 13.4, 13.5, and 13.6 imply

$$e_a = \frac{\delta p(1-(\beta q)^I)}{G}$$

$$e_k = \frac{\delta \beta^{k-1}(1-\beta q)}{G} \quad , \quad 1 \le k \le I$$

and

$$e_b = \frac{\delta \beta^I (1-\beta q)}{G} \ .$$

Similarly, from the derivatives of the transformed pdf's, we obtain

$$\mu_a = \frac{1-q^I}{\gamma} \ , \ \mu_0 = \frac{1}{\beta} \ , \ \mu_b = \frac{1}{\delta}, \text{ and } \mu_k = 1$$

where $1 \le k \le I$.

3.4 Comparison of CSP-12 and CSP-13. In the equations to be derived in this section, $P_2(\infty;3)$ is the long run percentage of time spent in state b=2 in the three stage reduction of SMC($\bar{3}$) to SMC($\bar{2}$) which is the analogue of SMC(2) for CSP-13. $P_2(\infty;3)$ can also be directly obtained

from SMC($\overline{3}$) by filtering out the states 0 through I and b, again yielding SMC($\overline{2}$). This latter filtration is equivalent to the SMC method applied to the ordered ensemble (0, 1, ···, I, b), with pev = (1, 0, ···, 0), I+1 zeros, to obtain the two state model for CSP-13.

Given the stationary vector components and the state mean time values, from Theorem 13, we get the α's for CSP-13.

$$\alpha_k = \delta\beta^k P_2(\infty;3),\ 1 \le k \le I \tag{13.6}$$

and

$$\alpha_I = \beta^{I+1} P_2(\infty;3) \tag{13.7}$$

where

$$P_2(\infty;3) = \frac{\mu_2'}{\frac{p\beta(1-(\beta q)^I}{(1-\beta q)}\mu_1' + \mu_2'} \tag{C7}$$

(1 = a and 2 = $\overline{b}$).

Applying the Ergodic Theorem and Eqs. 13.6 and 13.7 to the functional $\overline{W}(t)$, defined as SMC($\overline{2}$), yields

$$\lim_{t\to\infty} \frac{E[\overline{W}(t)]}{t} = \delta \sum_1^I k\alpha_k + \delta I\alpha_b$$

$$= \beta\delta^2 P_2(\infty;3) D_\beta \left(\frac{1-\beta^{I+1}}{1-\beta}\right)$$

$$+ \delta I\beta^{I+1} P_2(\infty;3)$$

$$= \beta(1-\beta^I) P_2(\infty;3),$$

Upon taking the limit, the definition of IFI(t;3), analogous to Definition 4, gives

$$AIFI(\infty;3) = \nu\beta(1-\beta^I)P_2(\infty;3).$$

Adding AFI(∞;3) to the above leads to the final equation

$$ATFI(\infty;3) = 1-\nu P_2(\infty;3)(\delta+\beta^{I+1}) \qquad (C8)$$

With regard to the last equation, we have

Theorem 14. For p in the open unit interval,

$$ATFI(\infty;3) < ATFI(\infty;2).$$

Proof. The statement is equivalent to

$$P_2(\infty;3) > P_2(\infty;2)$$

which is implied by

$$\frac{p\beta(1-(\beta q)^I)}{1-\beta q} < 1-q^I.$$

Dividing both sides by p and using the theorem on geometric sums, the above inequality is equivalent to

$$\beta\left(1+\sum_{j=1}^{I-1}(\beta q)^j\right) < \left(1+\sum_{j=1}^{I-1}q^j\right)$$

or

$$\beta[1+s_1] < [1+s_2]$$

But $\beta < 1$ and $\beta s_1 < s_2$, for p between zero and one. The cases for $p = 0$ and $p = 1$ lead trivially to the same formulas.

To handle the transient case of CSP-13, SMC($\overline{3}$) is used for the increasing part of $\overline{W}$. The constant part of $\overline{W}$ is handled in the same way as the corresponding constant part of W is handled for CSP-12. That is,

SMC($\overline{3}$) is collapsed (or filtered) to SMC($\overline{4}$) which in turn is collapsed to SMC($\overline{5}$). This analogous two stage process for CSP-13 is briefly given in

Theorem 15. For CSP-13, filtration gives the following ordered set of models:

$$SMC(\overline{5}) < SMC(\overline{4}) < SMC(\overline{3}).$$

Proof. Combining states 1 through I, in SMC($\overline{3}$), into state c' as is done with SMC(3), we have

$$\hat{Q}_{c'b} = \hat{Q}_{12}\hat{Q}_{23}\cdot\cdots\cdot\hat{Q}_{Ib}$$

$$= (\beta/z)^I$$

Similarly,

$$\hat{Q}_{c'0} = \hat{Q}_{10}+\hat{Q}_{12}\hat{Q}_{20}+\cdots+\hat{Q}_{12}\hat{Q}_{23}\cdot\cdots\cdot\hat{Q}_{I0}$$

$$= \frac{\delta q}{z-\beta q}\left(1-\left(\frac{\beta q}{z}\right)^I\right)$$

$$\hat{Q}_{0c'} = \beta/(z-\delta)$$

$$\hat{Q}_{c'a} = \frac{\delta}{z-\beta}\left(1-\left(\frac{\beta}{z}\right)^I\right) - \hat{Q}_{c'0}\ .$$

Secondly, combining states 0 and c' into the new state c is similarly accomplished and yields

$$\hat{Q}_{cb} = \frac{\hat{Q}_{0c'}\hat{Q}_{c'b}}{1-\hat{Q}_{0c'}\hat{Q}_{c'0}}$$

$$\hat{q}_{ca} = \frac{\hat{Q}_{0c}\hat{Q}_{c'a}}{1-\hat{Q}_{0c}\hat{Q}_{c'0}}$$

The corresponding q's are given by

$$q_{cb} = \frac{(1-\beta q)\beta^I}{A}$$

and

$$q_{ca} = \frac{(1-\beta q)(1-\beta^I)-\delta q(1-(\beta q)^I)}{A}$$

where

$$A = p + \delta q(\beta q)^I.$$

Once again, the constant part of the functional $\bar{\bar{W}}(t)$ is given by

$$I\left\{\frac{\bar{N}_b(t)}{t} - \frac{\bar{P}_{ab}(t)}{t}\right\} .$$

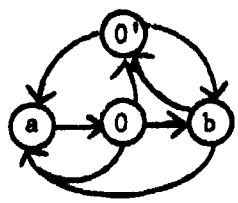
4.0 DSI AND OTHER FUNCTIONALS.

4.1 Introduction. The TFI functional makes a distinction between the two plans treated in Chapter 3 in terms of the "pseudophase" transitional probabilities. However, because of its very definition, TFI does not explicitly take account of multiple inspections of a given production unit. That is, TFI is defined in terms of an operational time which is measured by a flow of successive and nonrepeating production units. In this chapter, a new functional, along with a variation, is introduced to augment TFI as a measure of plan performance. The functional is Fraction of Repetitions (FR). It will be analyzed only for the first type of plan (CSP-12). Furthermore, FR is chosen as the principal functional because 1.) it is naturally normalized and 2.) its long run moments can be naturally derived from those of the transient case with a certain amount of ease. Short run higher moments for its variant cannot be obtained so readily; indeed, appeal must be made to the Strong Ergodic Theorem (or Renewal Theorem) for even the long run (expected) value.

4.2 SMC(9) and FR(N;2). The model which will be used, SMC(9), is a modification of SMC(2) and is portrayed in Figure 5.

Figure 5

CSP-12 and SMC(9)



The transitional matrix of the embedded MC is

$$\begin{array}{c} \\ a \\ 0 \\ 0' \\ b \end{array} \begin{array}{c} \begin{array}{cccc} a & 0 & 0' & b \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ \overline{\delta} & 0 & \delta q^I & \beta \\ \overline{\delta}/r & 0 & 0 & \beta/r \\ 1-q^I & 0 & q^I & 0 \end{bmatrix} \end{array}$$

where $\overline{\delta} = \delta(1-q^I)$ and $r = 1-\delta q^I$.

The matrix entries are obtained from the transformed pdf's given in Theorem 16. SMC(9) is an irreducible SMC

Proof. The transformed pdf's are

$\hat{Q}_{a0} = q^I(z-q)/\phi(z)$

$\hat{Q}_{0a} = \overline{\delta}/z$, $\hat{Q}_{00'} = \delta q^I/z$, and $\hat{Q}_{0b} = \beta/z$

$\hat{Q}_{0'a} = \overline{\delta}/(z-\delta q^I)$ and $\hat{Q}_{0'b} = \beta/(z-\delta q^I)$

$\hat{Q}_{ba} = \overline{\delta}/(z-\beta)$ and $\hat{Q}_{b0'} = \delta q^I/(z-\beta)$.

The mean holding times, obtained from the derivatives of the transformed pdf's, are

$$\mu_a = \frac{1-q^I}{\gamma}\ ,\ \mu_0 = 1,\ \mu_{0'} = \frac{1}{1-\delta q^I}\ ,\ \text{and}\ \mu_b = \frac{1}{\delta}\ .$$

Using the matrix given after Figure 5 to solve the usual eigen value equation, for the stationary vector $\underline{e}$, yields the system of equations given below.

$$\bar{\delta} e_0 + \frac{\bar{\delta}}{r}\ e_{0'} + (1-q^I) e_b = e_a$$

$$e_a = e_0$$

$$\delta q^I e_0 + q^I e_b = e_{0'}$$

$$\beta e_0 + \frac{\beta}{r}\ e_{0'} = e_b$$

(where $r = 1-\delta q^I$).

Solving the system gives

$$e_a = e_0$$

$$e_b = \left(\frac{\beta}{1-q^I}\right)\ e_0$$

$$e_{0'} = \frac{q^I(1-\delta q^I)}{1-q^I}\ e_0\ .$$

Again we use the fact that the components of the stationary vector add to one. Using the equation which expresses this fact, together with the last three, gives

$$e_a = e_0$$

$$= (1-q^I)/G$$

$$e_{0'} = q^I(1-\delta q^I)/G$$

and

$$e_b = \beta/G$$

where

$$G = (1-q^I) + (1-\delta q^I) + \beta.$$

We finish the proof by translating, into English text, what the transitions mean in SMC(9); we will write "state x goes to state y" as "x to y". 0 to b if no defect, 0 to 0' if a defect is found but DSI finds none, 0 to a if a defect is found and DSI finds one or more, 0' to a if a defect is found and DSI finds one or more, 0' to b if unit is either not inspected or is, and found non-defective, and 0' to 0' (remaining in 0') if a defect is found but DSI finds no defects. The transition 0' to 0' is "internal" - that is, 0' has no self transitions and is consequently a non-trivial SMC state (see its pdf above and Chapter 1, section 5).

We are now ready to define the principal functional in

Definition 6. Given the model SMC(9) for CSP-12, the functional Fraction of Repetitions is

$$FR(t) = \left(\frac{N_a(t)-1}{t}\right) + \frac{\sum_{k=0}^{t}{}' C_{0'}(k)}{t} .$$

The definition of FR(t) is motivated by the comments made at the end of the proof to Theorem 16. In addition, we remark that minus one appears since the inspection process begins in state a and the summation appears for 0' since self transitions are not allowed. For infinite t, FR has the value given in

Theorem 17.

$$\lim_{t\to\infty} FR(t) = \frac{1}{(1-q^I)\mu'_1 + \mu'_2} , \text{ [a.e.]}$$

where μ'_1 and μ'_2 are defined in Theorem 1.

Proof. From Definition 6,

$$\lim_{t\to\infty} FR(t) = \lim_{t\to\infty}\left(\frac{N_a(t)-1}{t}\right) + \lim_{t\to\infty}\left(\frac{\sum C_0'(k)}{t}\right)$$

$$= \frac{\alpha_a}{\mu_a} + \alpha_0', \text{ [a.e.]}$$

by the Strong Ergodic Theorem. From Theorem 16 and A.25, we have

$$\frac{\alpha_a}{\mu_a} = \frac{(1-q^I)}{\mu_1'(1-q^I)+(1-q^I)+q^I+\beta/\delta}$$

$$= \frac{(1-q^I)}{\mu_1'(1-q^I)+\mu_2'}$$

and

$$\alpha_0' = \frac{q^I}{\mu_1'(1-q^I)+\mu_2'} .$$

Adding the two expressions finishes the proof.

Since (I)·(tFR(t)) can be regarded as the degree of inspection overlap, we are led to define a variant of FR(t) in

Definition 7.

$$FR'(t) = \frac{t}{I(tFR(t)) + t} .$$

Concerning this functional, we have

Theorem 18.

$$\lim_{t\to\infty} FR'(t) = \frac{(1-q^I)\mu_1'+\mu_2'}{I + (1-q^I)\mu_1' + \mu_2'} , \text{ [a.e.]}$$

Proof. From Definition 7,

$$\lim_{t\to\infty} FR'(t) = \lim_{t\to\infty} \frac{1}{(I)(FR(t))+1}$$

$$= \frac{1}{I\left(\frac{a_a}{\mu_a} + a_{0'}\right) + 1}$$

by Theorem 17,

= the result.

4.3 Expansions and Extensions. Another possible treatment of DSI is the expansion of MRP (and SMC) models to "transition state" models. We will work here only with MRP's.

Given a MRP (Y, U) as in A.19, we can easily prove that

$$P[T_n = t \mid Y_{n-1} = i \text{ and } Y_n = j] = \frac{Q_{ij}(t)}{q_{ij}} \qquad \text{(D1)}$$

where $T_n = U_n - U_{n-1}$. From A.19 and Eq. D1, we can also easily show that

$$\{((Y_n, Y_{n+1}), U_n)/n \text{ varies over the nat'l nos.}\} \qquad \text{(D2)}$$

is a (derived) MRP whose pdf's are given by

$$P[(Y_n, Y_{n+1}) = (\ell,k), T_n = t \mid (Y_{n-1}, Y_n) = (i,j)]$$

$$= \frac{Q_{ij}(t)}{q_{ij}} q_{\ell k}\delta_{j\ell} \qquad \text{(D3)}$$

We name the MRP given by expression D2 and simplify notation in

Definition 8. The MRP given by D2 is called the Expanded MRP. Its pdf's are given by Eq. D3 and denoted by

$$Q_{(ij)(\ell k)}(t).$$

Such a derived process can automatically keep track of transitions, their number and type, in the parent process. Thus, for example, FR(=;2) could be defined (and evaluated) on "expanded" MRP(2) as given below.

Theorem 19. Expanded MRP(2) is a MRP.

Proof. From Definition 8 and Theorem 2, the transformed pdf's are (dropping the argument)

$$\hat{Q}_{(12)(22)} = q^I \hat{Q}_{12}$$

$$\hat{Q}_{(12)(21)} = (1-q^I)\hat{Q}_{12}$$

$$\hat{Q}_{(22)(22)} = \hat{Q}_{22}$$

$$\hat{Q}_{(22)(21)} = \frac{(1-q^I)\hat{Q}_{22}}{q^I}$$

$$\hat{Q}_{(21)(12)} = \frac{\hat{Q}_{21}}{(1-q^I)} \quad .$$

Letting $z = 1$ in the above equations, we get the transitonal matrix of the embedded MC

$$\begin{array}{c} \\ (12) \\ (22) \\ (21) \end{array} \begin{array}{c} \begin{array}{ccc} (12) & (22) & (21) \end{array} \\ \begin{bmatrix} 0 & q^I & 1-q^I \\ 0 & q^I & 1-q^I \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

Using the matrix to solve for the components of the stationary vector gives

$$e_{(12)} = (1-q^I)/G,\ e_{(22)} = q^I/G, \text{ and } e_{(21)} = e_{(12)}$$

where

$$G = 2-q^I.$$

Defining μ_{ij} as the mean holding time till transition to state j from state i, using Definition 8, and using the mean value property of the transformed pdf's, we get

$$\mu_{(ij)} = (\mu_{ij} \sum_k q_{jk})/q_{ij}$$

$$= \mu_{ij}/q_{ij} \qquad (19.1)$$

Applying Eq. 19.1 to the transformed pdf's yields

$$\mu_{(12)} = \mu'_1/q_{12} \qquad \mu_{(22)} = \mu_{22}/q_{22} \qquad \mu_{(21)} = \mu_{21}/q_{21}$$

$$= \mu'_1/1 \qquad = \mu_{22}/q^I \qquad = \mu_{21}/(1-q^I)$$

$$= \mu'_1 \qquad = \mu'_2 \qquad = \mu'_2$$

where μ'_1, μ'_2, and the transitional probabilities are defined (or derived from) Theorem 1.

Definition 9. For Expanded (MRP(2)),

$$FR(t;2) = (I)\left\{\frac{N_{(22)}(t) + N_{(21)}(t)}{t}\right\}.$$

Theorem 20. For FR(t;2) in Definition 9,

$$\lim_{t\to\infty} FR(t;2) = \frac{1}{(1-q^I)\mu'_1 + \mu'_2}, \qquad [a.e.]$$

$$= \frac{1}{E[T]}$$

where E[T] is given in Proposition 1.

Proof. Theorem 19 and Definition 9.

We close this chapter by showing that SMC(9) cannot be collapsed into any of the other models for CSP-12. Any collapsing would require that the ordered ensemble S = (0,0') be a macrostate as defined in Chapter 1. However, entrance from state a or b would require the pev to be (1,0) or (0,1), respectively. If we picked the former pev and formally defined $\hat{Q}_{bS}$ to be the same as $\hat{Q}_{b0'}$, the Backward Equation system, for SMC(9)', say, would not hold. For example, if S were a macrostate, then, letting S = d, the equations

$$P_{ab}(t) = Q_{ad} * P_{db}(t)$$

and

$$P_{bb}(t) = Q_{bd} * P_{db}(t) + Q_{ba} * P_{ab}(t) + J_b(t)$$

would have to hold. However, entrance to d from state a results in a greater probability for a given holding time in d than an entrance from state b. Consequently, $P_{db}(t)$ is not well defined.

Another way of stating this inconsistency is provided by

Definition 10. Let $P_{xy}(t;w)$ be the Fundamental Probability Function, from x to y, given that entrance into x is from w.

Then consistency requires that $P_{xy}(t;w)$ be independent of state w. However, for SMC(9)',

$$P_{db}(t;b) \neq P_{db}(t;a)$$

Similar results are obtained if we pick (0,1) as the pev and define Q_{ad} formally.

Under certain conditions, we can still reduce a MC to a SMC in the case that the relevant probability functions are indexed by ensembles of MC states as occurs in SMC(9)'. The dependence of the probability functions on the entrance ensemble is equivalent to the dependency of the pev's. We

therefore drop the restriction of pev independence by using $\underline{v}(x;y)$ to devote the pev of the ensemble x given an entrance from y. Furthermore, since $v_j(x;y)$ being zero, for a given MC state j, can imply that j cannot be reached from any other states in x, x itself becomes a function of y: $x = x(y)$. Further dependence is handled by dropping the inner parenthesis: for example, $x(y(w)) = x(yw)$. Letting a, b, c, d, ... be (disjoint) ensembles of MC states which we wish to transform into macrostates, we make a provisional definition for the holding time pdf's in

Definition. Given a, b, c, and $\underline{v}(a;c)$

$$Q_{ab}(t;c) = \sum_j v_j(a;c) f^t_{j,B}$$

where j varies over the set a and B is the absorbing "state" corresponding to b.

Given the underlying MC, $M(\cdot)$, the above Definition will yield a SMC iff (letting $R_n = M(U_n)$, U_n being the elapsed time)

$$P[R_{n+1} \text{ in } b(ac\cdots) | R_n \text{ in } a(c\cdots),\ R_{n-1} \text{ in } c(d\cdots),\ \cdots R_0 \text{ in } y]$$

$$= P[R_{n+1} \text{ in } b(a) | R_n \text{ in } a(c)]$$

$$= P[(R_n, R_{n+1}) = (a,b),\ T_{n+1} = t | (R_{n-1}, R_n) = (c, a)]$$

where $T_{n+1} = U_{n+1} - U_n$. Thus $\underline{v}(b, a(c\cdots)) = \underline{v}(b;a)$ and T_{n+1} depends only on $Q_{ab}(\ ;c)$. Therefore, it is necessary and sufficient to require that a(c) include all the states of a which communicate with the states of all other ensembles, (for all a, c) since $\underline{v}(b;a)$ depends only on the one step MC transitional probabilities. In particular, it is sufficient that a(c) = a, for all sets a and c.

Under the above necessary and sufficient condition, we can now write

$$P_{ad}(t;c) = \sum_b Q_{ab}(\ ;c) * P_{bd}(\ ;a)(t)$$

$$+ \delta_{a,d} J_a(t;c).$$

From another point of view, we can also let a(c) denote the state $(a;\hat{Q}_{ax}(\ ;c))$, x varying over the exit states. Using this latter notation, we can set

$$Q_{ab}(t;c) = Q_{a(c),b}(t).$$

For a given MC, the resultant number of states may be small enough to warrant SMC reduction, in the above case of dependent pev's, if the reduction in complexity is substantial enough. This extended SMC reduction can be applied to SMC(9); S(a) = the ordered set (0,0') and S(b) = (0'). However, nothing is gained here since we still have 4 states.

In closing this chapter, we point out yet another deviation from the conditions of a state independent, stationary pev. The deviant condition can be found in [6.2, Chp. 5]. The type of pev found there is an initial pev used in the arbitrary entry case of CSP's. It is shown that the existence of these pev's is equivalent to that of initial (or delayed) holding time pdf's in the stationary (or random entry) case for ergodic SMC's. Thus, this special type of pev is handled in a manner analogous to that used for state dependent pev's - as an "index" (given, in the paper cited, by a prime over the Q's).

5.0 CONCLUSION.

5.1 Summary. Two approaches to the DSI modification of CSP-11 are considered in Chapters 2 through 4. The first approach, found in Chapters 2 and 3, ignores any overlap in the inspection process by using a functional, defined on a new DSI model, to count only the additional units which are inspected from sampling phase segments - units which would otherwise not be inspected under CSP-11. Since the functional TFI is not sufficient to deal with all the important aspects of CSP-12, a second approach, found in Chapter 4, uses a new functional, defined on a slightly different DSI model, to take account of inspection overlaps. In either treatment, there is no explicit backtracking in operational time itself; both approaches incorporate the time shift into the transitional changes, induced by DSI, which are, in turn, incorporated in the pdf's of the underlying models. Throughout the paper, variations in functionals and sampling plans, together with comparisons of them with the primary objects of study are also considered.

5.2 Methods Used. Two principal tools are used in the analysis of DSI: SMC (and MRP) reduction and the z-transform. Since the SMC's constructed for the analysis are modifications of the SMC model of CSP-11, the process

of constructing a SMC class from a MC model, described in Chapter 1, is turned around. In Chapter 4, the importance of the probability entrance vector (pev) is brought out by the incompatibility of SMC(9) with the other CSP-12 models. Also in Chapter 4, the use of an Expanded MRP in the analysis of DSI is illustrated; this kind of analysis could be elaborated on for further investigation of functionals dependent on a sequence of transitions.

We conclude this paper with the observation that DSI can be used to modify the more complex CSP's described in Chapter 1.

APPENDIX

A.0 SEMI MARKOV CHAINS. Given that $X(\cdot)$ is a time homogeneous, aperiodic, irreducible or absorbing, and finite state Semi Markov Chain (SMC) with state space S, the following notation and statements are used in the body of the text [cf., 6.7, 6.10, 6.14, and 6.15].

A.1 Notation and Definitions. For i, j, k, ℓ in S:

1. $Q_{ik}(t) = P[X(t)=k,\ X(t')=i,\ 0 < t' < t \mid X(0)=i]$.*

This function is the (defective) pdf of the time of sojourn in state i until a transition is made to state k (for discrete t and $i \neq k$).

2. $P_{ik}(t) = P[X(t)=k \mid X(0)=i]$.

This function is the fundamental probability function of the SMC for (i to k).

3. $F_{ik}(t) = P[X(t)=k;\ X(t') \neq k,\ 0 < t' < t \mid X(0)=i]$.

This function is the first entrance probability function for (i to k).

4. $J_k(t) = H_0 * (\delta_0 - \sum_{\ell} Q_{k\ell})(t)$.

This function is the probability of not leaving state k by time t.

5. $U_n(k)$ is the time of n^{th} entry into k.

6. $N_k(t) = \text{Max} \{ n / U_n(k) \leq t \}$

This random variable is the renewal function for state k.

7. U_n is the time of n^{th} entry.

8. $Y(n) = X(U_n)$ is the embedded Markov Chain associated with the SMC.

*This definition corrects statement 3, definition 5 in [6.2, p. 664].

For the case where self transitions are allowed, we can use the symbols above to define a Markov Renewal Process (MRP).

9. A MRP is the ordered pair (Y, U) such that, for states i, k in S,

$$P[Y_n=k,\ T_n=t \mid Y_{n-1}=i,\ Y_{n-2},\ \cdots,\ Y_0;\ T_{n-1},\ T_{n-2},\ \cdots,\ T_0]$$

$$= P[Y_n=k,\ T_n=t \mid Y_{n-1}=i],\ T_n = U_n - U_{n-1}$$

$$= Q_{ik}(t).$$

(Note that this pdf is, in general, different from that defined in A.11.)*

10. The SMC X(t) associated with a MRP is defined by

$$X(t) = Y(t)$$

$$= Y_{N(t)}$$

where $N(t) = \sum_j N_j(t)$, j in S.

A.2 Statements.

1. By time homogeneity and the method of first entrance, we have the Backward Equations:

$$P_{ik}(t) = \sum_j Q_{ij} * P_{jk}(t) + (\delta_{ik})J_k(t).$$

2. $P_{ik}(t) = F_{ik} * P_{kk}(t) + (\delta_{ik})J_k(t).$

3. If $q_{ik} = H_0 * Q_{ik}(+\infty)$,

$$T = [q_{ik}]$$

is the transitional matrix of Y.

*This definition corrects that given in [6.2, p. 695].

4. If X is irreducible, the equation

$$\underline{e}T = \underline{e}$$

has a unique normalized solution called the stationary vector of the SMC.

5. $$\lim_{t\to\infty} P_{ik}(t) = \frac{e_k\mu_k}{\sum_\ell e_\ell \mu_\ell}$$

$$= a_k \text{ (or } P_k(\infty))$$

where μ_k is the mean time of sojourn in state k and the e_i's are the components of $\underline{e}$.

6. $$\lim_{z\to 1}\left(\frac{z-1}{z}\right)\hat{P}_{ik}(z) = a_k$$

7. (Strong Ergodic Theorem.) If W is a functional defined on the SMC, we have, as N approaches infinity,

$$\frac{1}{N}\sum_s W(X(s)) \text{ approaches } E_{\underline{a}}[W], \quad \text{[a.e.]}$$

$$= \sum_k W(k)a_k.$$

In the case of self transitions, we have

8. If (Y, U) is a MRP such that $q_{ii} < 1$, the unique SMC induced by the MRP has its pdf's given via ($i \neq j$)

$$\hat{Q}^*_{ij} = \frac{\hat{Q}_{ij}}{1-\hat{Q}_{ii}}, \text{ if } q_{ii} > 0$$

$$= \hat{Q}_{ij}, \text{ otherwise}$$

where the Q's are given by A.19. It is equivalent, almost everywhere, to the associated SMC.

9. The properties of time homogeneity, irreducibility, and aperiodicity are preserved under filtration.

6.0 REFERENCES.

6.1 Arp, D. L. "Expected values from a Markov chain model of CSP-F and the z-transform," Proc. 19th Conf. on Design of Experiments in Army R, D and T (ARO Rpt. 74-1), pp. 263-296.

6.2 Arp, D. L. "Semi Markov chains applied to Markov chain models of continuous sampling plans," Proc. 20th Conf. on Design of Experiments in Army R, D and T (ARO Rpt. 75-2, part 2), pp. 657-739.

6.3 Barrière, R. Optimal Control Theory, W. B. Saunders Co., Philadelphia, PA, 1967.

6.4 Brugger, R. M. "A simplification of the Markov chain approach to continuous sampling plan formulation," QEM 21-230-12, Ammunition Procurement and Supply Agency, 1972 (transferred to U.S. Army Armament Command, Rock Island, IL 61201).

6.5 Brugger, R. M. "Responsiveness properties of continuous sampling plans," Proc. 19th Conf. on Design of Experiments in Army R, D and T (ARO Rpt. 74-1), pp. 233-262.

6.6 Brugger, R. M. "Skip-lot procedure formulation using the simplified Markov chain method," Proc. 20th Conf. on Design of Experiments in Army R, D and T (ARO Rpt. 75-2, part 2), pp. 647-656.

6.7 Cinlar, E. "Markov renewal theory," Adv. Appl. Prob., Vol 1 (1969), pp. 123-187.

6.8 Cinlar, E. "Periodicity in Markov renewal theory," AD 738-469.

6.9 Cox, D. R. Renewal Theory, Methuen, London, 1962.

6.10 Feller, W. "On semi Markov chains," Proc. of Nat'l Acad. of Sci., Vol. 54 (1964), pp. 653-659.

6.11 Feller, W. An Introduction to Probability Theory and Its Applications, Vol. I (3rd ed.), Wiley, New York, 1968.

6.12 Jury, E. Theory and Application of the z-Transform Method, Krieger Publ. Co., Huntington, N. Y., 1973.

6.13 Kemeny, J. and Snell, J. Finite Markov Chains, van Nostrand, Princeton, 1969.

6.14 Pyke, R. "Markov renewal processes: definitions and preliminary properties," Ann. Math. Stat., Vol. 32 (1961), pp. 1231-1242.

6.15 Pyke, R. "Markov renewal processes with finitely many states," Ann. Math. Stat., Vol. 32 (1961), pp. 1243-1259.

6.16 Pyke, R. and Schaufele, R. "Limit theorems for Markov renewal processes," Ann. Math. Stat., Vol. 35 (1964), pp. 1746-1764.

PROGRESSIVELY CENSORED SAMPLING IN THE THREE PARAMETER LOG-NORMAL DISTRIBUTION*

A. Clifford Cohen

The University of Georgia
Athens, Georgia

SUMMARY

This paper is an extension of previous work by the writer concerning progressively censored sampling in the normal distribution [4] and in the Weibull distribution [6]. Here local maximum likelihood estimators and estimators which utilize the first order statistic are derived for the three-parameter log-normal distribution when samples are progressively censored. An illustrative example involving life test data is included. Various properties of the proposed estimators are investigated.

KEY WORDS

Log-normal Distribution
Progressively Censored Samples
Life Testing

1. INTRODUCTION

Progressively censored samples frequently occur in life and fatigue tests, where individual observations are time ordered and where at various times during a test, some of the survivors are removed (i.e. censored) from further observation. Samples of this type from

---

*Research supported by the National Science Foundation through Grant GP-34318. Presented to the Institute of Mathematical Statistics, Montreal, Quebec, Canada, August 1972.

This paper appeared in Technometrics 18(Feb. 1976) pp 99-103. Permission of the editors of that journal to reproduce it here is appreciated.

the normal and from the exponential distribution have received previous attention from Herd [10], Roberts [18], and the writer [4]. Progressively censored samples from the two-parameter Weibull distribution were considered by the writer [5] and by Ringer and Sprinkle [17]. More recent work by the writer [6] deals with progressive censoring in the three-parameter Weibull distribution. The present paper is concerned with progressive censoring in the three-parameter log-normal distribution.

2. THE SAMPLE

Let N designate the total sample size, and n the number which fail and therefore result in completely determined life spans. Suppose that censoring occurs in k stages at times $T_j > T_{j-1}$, $j=1, 2, \ldots, k$, and that r_j surviving items are removed (censored) from further observation at the jth stage. Thus

$$N = n + \Sigma_1^k r_j. \tag{1}$$

Two types of censoring are generally recognized. In Type I censoring, which is of primary interest here, the T_j are fixed, and the number of survivors at these times are random variables. In Type II censoring, the number of survivors are fixed and the T_j are random variables. In both types, the r_j are either fixed or determined independently of the life span X. The observations x_i are ordered according to magnitude.

The likelihood function L(S), where S signifies a k-stage Type I progressively censored sample of the type described, is

$$L(S) = C \prod_{i=1}^{n} f(x_i) \prod_{j=1}^{k} [1 - F(T_j)]^{r_j}, \tag{2}$$

in which C is a constant while f(x) and F(x) are density and distribution functions respectively.

3. THE LOG-NORMAL DISTRIBUTION

We write the density function for the three-parameter log-normal distribution as

$$f(x;\ \mu,\ \sigma,\ \gamma) = \frac{1}{\sigma\sqrt{2\pi}\ (x-\gamma)}\, e^{-[\ln(x-\gamma)-\mu]^2/2\sigma^2},\ \gamma < x < \infty, \qquad (3)$$

$$= 0, \text{ elsewhere.}$$

This distribution derives its name from the fact that when the random variable X is lognormal (μ, σ^2, γ), then $Y = \ln(X-\gamma)$ is normal (μ, σ^2). The mean, median, mode, variance, coefficient of variation, β_1 and β_2 (Pearson's Betas) for this distribution (c.f. Yuan [23]) are

$$\begin{aligned} \mu_x &= \gamma + e^{\mu}\sqrt{\omega}, \\ Me &= \gamma + e^{\mu}, \\ Mo &= \gamma + e^{\mu}/\omega, \\ V(x) &= e^{2\mu}\ \omega(\omega-1), \\ CV &= \sqrt{\omega-1}, \\ \beta_1 &= \alpha_3^2 = (\omega+2)^2\ (\omega-1), \\ \beta_2 &= \alpha_4 = \omega^4+2\omega^3 + 3\omega^2-3, \end{aligned} \qquad (4)$$

where

$$\omega = e^{\sigma^2}, \qquad (5)$$

and where α_3 and α_4 denote the third and fourth standard moments. The coefficient of variation about the left terminus is defined as

$$CV = \sqrt{V(x)}\ /(\mu_x-\gamma). \qquad (6)$$

Previous investigations by the writer [3], Aitchison and Brown [1], Hill [11], Wilson and Worcester [21], and others have dealt with maximum likelihood estimation in the three parameter log-normal distribution when

samples are complete. Harter and Moore [9] considered local maximum likelihood estimation in the three parameter log-normal distribution for singly and doubly censored as well as for complete samples. Hill examined some unusual features of the likelihood function of this distribution which had apparently escaped the notice of earlier investigators. He demonstrated the existence of paths along which the likelihood function of any ordered sample $x_1, \ldots, x_n$ tends to ∞ as (γ, μ, σ^2) approach $(x_1, -\infty, \infty)$.

This global maximum of the likelihood function thereby leads to the inadmissible estimators, $\hat{\gamma} = x_1$, $\hat{\mu} = -\infty$ and $\hat{\sigma}^2 = \infty$ regardless of the sample. On the other hand, when we equate partial derivatives of the log-likelihood function to zero, solution of these equations leads to local maximum likelihood estimates which in most cases are reasonable and as noted by Harter and Moore (loc. cit.) appear to possess most of the desirable properties ordinarily associated with maximum likelihood estimators. Exceptions may occur in small samples for which the likelihood function fails to exhibit a clearly defined local maximum.

4. LOCAL MAXIMUM LIKELIHOOD ESTIMATION

With the p.d.f. as given in equation (3), the logarithm of the likelihood function (2) becomes

$$\ln L = -n\ln\sigma - \Sigma_1^n \ln(x_i-\gamma) - \frac{1}{2\sigma^2}\Sigma_1^n[\ln(x_i-\gamma)-\mu]^2 + \Sigma_1^k r_j \ln[1-F_j] + \ln C. \qquad (7)$$

Local maximum likelihood estimators (LMLE) are obtained by simultaneously solving the estimating equations

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \Sigma_1^n [\ln(x_i-\gamma)-\mu] - \Sigma_1^k \left(\frac{r_j}{1-F_j}\right) \frac{\partial F_j}{\partial \mu} = 0,$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3} \Sigma_1^n [\ln(x_i-\gamma)-\mu]^2 - \Sigma_1^k \left(\frac{r_j}{1-F_j}\right) \frac{\partial F_j}{\partial \sigma} = 0, \tag{8}$$

$$\frac{\partial \ln L}{\partial \gamma} = \Sigma_1^n \left(\frac{1}{x_i-\gamma}\right) + \frac{1}{\sigma^2} \Sigma_1^n \left(\frac{\ln(x_i-\gamma)-\mu}{x_i-\gamma}\right) - \Sigma_1^k \left(\frac{r_j}{1-F_j}\right) \frac{\partial F_j}{\partial \gamma} = 0.$$

Let $$Z_j = Z(\xi_j) = \frac{\phi(\xi_j)}{1-F(\xi_j)}, \tag{9}$$

where

$$F_j = F(T_j) = \int_{\gamma}^{T_j} f(x)dx = \int_{-\infty}^{y_j} g(y)dy = \int_{-\infty}^{\xi_j} \phi(z)dz =$$

$$F(\xi_j), \tag{10}$$

in which $f(x)$ is given by (3), $g(y)$ is the normal density (μ, σ^2), $\phi(z)$ is the standard normal density $(0,1)$, and

$$y_j = \ln(T_j-\gamma), \text{ whereas } \xi_j = (y_j-\mu)/\sigma. \tag{11}$$

It then follows from (9), (10) and (11) that

$$\left(\frac{\sigma}{1-F_j}\right) \frac{\partial F_j}{\partial \mu} = -Z_j, \left(\frac{\sigma}{1-F_j}\right) \frac{\partial F_j}{\partial \sigma} = -\xi_j Z_j, \text{ and } \left(\frac{\sigma}{1-F_j}\right) \frac{\partial F_j}{\partial \gamma}$$

$$= -\frac{Z_j}{T_j-\gamma}. \tag{12}$$

When the results of (12) are substituted into (8), the estimating equations become

$$\Sigma_1^n[\ln(x_i-\gamma)-\mu] + \sigma\Sigma_1^k r_j Z_j = 0,$$

$$\Sigma_1^n[\ln(x_i-\gamma)-\mu]^2 + \sigma^2[\Sigma_1^k r_j \xi_j Z_j - n] = 0, \qquad (13)$$

$$\Sigma_1^n\left[\frac{\ln(x_i-\gamma)-\mu}{x_i-\gamma}\right] + \sigma^2\Sigma_1^n\left(\frac{1}{x_i-\gamma}\right) + \sigma\Sigma_1^n\left(\frac{r_j Z_j}{T_j-\gamma}\right) = 0.$$

Various iterative techniques are available for simultaneously solving these three equations for the required estimates $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$. A procedure that has performed quite well for the writer involves selecting a trial value γ_i for γ, solving the first two equations with $\gamma=\gamma_i$ for μ_i and σ_i using the standard Newton technique (c.f. page 90 of reference [20]), and then substituting these values into the third equation of (13). Once two values γ_i and γ_j have been found such that the absolute difference $|\gamma_i - \gamma_j|$ is sufficiently small and such that $H(\gamma_i,\mu_i,\sigma_i) \gtrless 0 \lessgtr H(\gamma_j,\mu_j,\sigma_j)$, where $H(\gamma,\mu,\sigma)$ designates the left side of the third equation of (13), the required estimates follow by linear interpolation. The smallest sample observation, x_1, is of course an upper bound on γ and may thus be employed as a first approximation γ_1 in the iteration procedure.

In the event that the third estimating equation of (13) is not satisfied for any value of γ in the permissible interval $\gamma \leq x_1$, then the modified estimators of Section 5 are to be recommended.

Harter and Moore encountered the related problem in connection with samples that are singly and doubly censored. With r observations

censored on the left so that x_r is an upper bound on γ, their recommendation is that an additional observation be censored on the left so that x_{r+1} then becomes a new upper bound on γ.

5. MODIFIED MAXIMUM LIKELIHOOD ESTIMATION

Alternate estimators (MMLE) which have proven most satisfactory in numerous applications, can be obtained by simultaneously solving the estimating equations

$$\frac{\partial \ln L}{\partial \mu} = 0, \quad \frac{\partial \ln L}{\partial \sigma} = 0, \text{ and } E[F(x_r)] = F(x_r),$$

where X_r is the rth order statistic in a sample of size N. Only those failures which occur prior to the time at which the first stage of censoring takes place, provide observed values for order statistics, and thus the maximum value of r is limited. In most applications, we might choose to set r=1, but a larger value might be preferred if there is reason to suspect contamination of the sample data in the vicinity of the terminus. Applicable estimating equations accordingly consist of the first two equations of (13) plus a third equation involving x_r as derived below. Since

$$F(x_r) = \int_{\gamma}^{x_r} f(x)dx, \text{ and since } E[F(x_r)] = \frac{r}{N+1}, \tag{14}$$

it follows that our third estimating equation becomes

$$\gamma = x_r - e^{\mu+\sigma\xi_r}, \tag{15}$$

where ξ_r is the standard normal deviate for which

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi_r} e^{-z^2/2} dz = \frac{r}{N+1}. \tag{16}$$

The modified estimators accordingly are found by simultaneously solving the set of equations consisting of the first two equations of (13) plus equation (15). The same procedure employed in Section 4 to calculate the LMLE is also applicable here. On determining γ_i and γ_j such that $|\gamma_i-\gamma_j|$ is sufficiently small and such that $G(\gamma_i,\mu_i,\sigma_i) \gtrless x_r \lesseqgtr G(\gamma_j,\mu_j,\sigma_j)$, where $G(\gamma,\mu,\sigma) = \gamma + e^{\mu+\sigma\xi_r}$, we interpolate for the required estimates just as we did in Section 4.

6. SOME SPECIAL CASES

Various special cases in which at least one of the parameters is known, are of interest in certain applications. The following are considered to be deserving of mention at this time.

MLE with γ known.

With γ known, there is no longer any distinction to be made between a local maximum and a global maximum. The applicable estimating equations in this case are the first two equations of (13), and they may be solved iteratively for the required estimates $\hat{\mu}$ and $\hat{\sigma}$ as outlined in Section 4. As an alternate technique, we might make the transformation $y_i = \ln(x_i-\gamma)$ and then proceed as described in reference [4] for a progressively censored sample from a normal distribution. Gajjar and Khatri [7] previously considered this special case.

LMLE with σ known.

It often happens that the shape parameter σ and thus α_3 are known, leaving only μ and γ to be estimated from the sample data. In this case, the applicable estimating equations consist of the first and third equations of (13).

MMLE with σ known.

In this case, the applicable estimating equations consist of the first equation of (13) plus equation (15).

7. ESTIMATE VARIANCES AND COVARIANCES

The asymptotic variance-covariance matrix of the estimators $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$ is obtained by inverting the information matrix in which elements are negatives of expected values of the second partial derivatives of the logarithm of the likelihood function. For sufficiently large samples, these expected values can be approximated by substituting the estimates obtained from a given sample directly into the partial derivatives which are given below.

$$
\begin{aligned}
\frac{\partial^2 \ln L}{\partial \mu^2} &= \frac{-n}{\sigma^2} - \frac{1}{\sigma^2} \Sigma_1^k r_j Z_j (Z_j - \xi_j), \\
\frac{\partial^2 \ln L}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \Sigma_1^n [\ln(x_i-\gamma)-\mu]^2 - \frac{1}{\sigma^2} \Sigma_1^k r_j \xi_j Z_j [2 + \xi_j (Z_j - \xi_j)], \\
\frac{\partial^2 \ln L}{\partial \gamma^2} &= \frac{1}{\sigma^2} \left[\Sigma_1^n \frac{[\ln(x_i-\gamma)-\mu-1+\sigma^2]}{(x_i-\gamma)^2} + \Sigma_1^k \frac{r_j Z_j [\sigma - (Z_j - \xi_j)]}{(T_j-\gamma)^2} \right], \\
\frac{\partial^2 \ln L}{\partial \mu \partial \gamma} &= \frac{\partial^2 \ln L}{\partial \gamma \partial \mu} = -\frac{1}{\sigma^2} \Sigma_1^n \left(\frac{1}{x_i-\gamma}\right) - \frac{1}{\sigma^2} \Sigma_1^k \frac{r_j Z_j (Z_j - \xi_j)}{(T_j-\gamma)}, \\
\frac{\partial^2 \ln L}{\partial \sigma \partial \gamma} &= \frac{\partial^2 \ln L}{\partial \gamma \partial \sigma} = -\frac{2}{\sigma^3} \Sigma_1^n \frac{[\ln(x_i-\gamma)-\mu]}{(x_i-\gamma)} - \frac{1}{\sigma^2} \Sigma_1^k \frac{r_j Z_j [1+\xi_j (Z_j - \xi_j)]}{(T_j-\gamma)}, \\
\frac{\partial^2 \ln L}{\partial \mu \partial \sigma} &= \frac{\partial^2 \ln L}{\partial \sigma \partial \mu} = -\frac{2}{\sigma^3} \Sigma_1^n [\ln(x_i-\gamma)-\mu] - \frac{1}{\sigma^2} \Sigma_1^k r_j Z_j [1+\xi_j (Z_j - \xi_j)].
\end{aligned}
\tag{17}
$$

Since the estimators $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\gamma}$ are local rather than global maximum likelihood estimators, the applicability of the variance-covariance

matrix obtained here, might be open to question. However, a Monte Carlo study by Nicholas Norgaard [16] indicates that the approximate asymptotic variances and covariances obtained here should be considered satisfactory when $n \geq 50$, although they might be misleading as measures of sampling error for small samples. Norgaard's results are consistent with results of an earlier Monte Carlo study by Harter and Moore (loc. cit.) in connection with singly and doubly censored samples. It is also to be noted that Norgaard's study indicates that variances and covariances of the MMLE are approximately equal to corresponding measures of the MLE. This is an area of investigation that is continuing to receive attention both from Norgaard and the writer.

8. AN ILLUSTRATIVE EXAMPLE

A simulated life test was conducted on 100 randomly selected units of a certain electronic device having a log-normal life span with $\mu = 5.0000$, $\sigma = 0.3000$ and $\hat{\gamma} = 100$. Sixty-five complete life spans were observed, while thirty-five observations were censored in three separate stages. Following are the life spans in hours to two places of decimal, for the 65 units which failed during the test.

167.91	200.88	219.14	232.91	246.61	262.59	287.71
175.83	201.76	220.59	235.66	247.17	263.94	288.81
185.88	205.31	222.00	236.75	249.14	266.12	291.30
188.14	206.98	222.82	237.40	249.73	266.62	295.18
189.08	210.78	224.33	239.05	250.09	267.01	297.38
191.96	212.49	225.60	240.22	252.89	270.64	
195.61	213.24	226.50	240.64	253.57	271.76	
197.01	215.25	227.24	242.17	255.57	275.48	
198.76	216.75	227.24	243.03	260.60	279.62	
199.05	218.78	231.42	244.56	261.99	285.19	

When the tenth failure occured at time $T_1 = 199.05$, twelve units selected at random from the survivors were censored (i.e. removed from the test). When the forty-fifth failure occured at time $T_2 = 250.09$, ten additional randomly selected survivors were removed, and the test was terminated at time $T_3 = 297.38$ with 13 survivors. In summarizing these data, we record: $N = 100$, $n = 65$, $\sum_1^3 r_j = 35$, $x_1 = 167.91$, $T_1 = 199.05$, $r_1 = 12$, $T_2 = 250.09$, $r_2 = 10$, $T_3 = 297.38$, $r_3 = 13$, $\sum_1^{65} x_i = 15{,}327.43$, $\bar{x}_{65} = 235.8066$.

Estimates were calculated as described in Sections 4, 5 and 6 and are summarized in the following table.

In general, the estimates obtained here compare favorably with corresponding population parameters.

9. ACKNOWLEDGEMENT

The able assistance of Mr. Nicholas Norgaard who performed all of the programming and carried out all of the calculations presented here, is gratefully acknowledged.

TABLE 1 - SUMMARY OF ESTIMATES

TYPE ESTIMATORS	PARAMETERS ESTIMATED							
	γ	μ	σ	ω	μ_x	V(X)	a_3	a_4
LOCAL MLE	111.77	4.9343	0.3404	1.12284	259.04	2664.32	1.0945	5.2032
MMLE	4.9853	5.5119	0.1796	1.03278	256.62	2076.61	0.5492	3.5411

$\hat{V}(\sigma) = 0.01870$ $\hat{V}(\mu) = 0.12552$ $\hat{V}(\gamma) = 2016.68$ $\hat{Cov}(\sigma,\mu) = -0.04417$

$\hat{Cov}(\sigma,\gamma) = 5.9859$ $\hat{Cov}(\mu,\gamma) = -14.9690$

	$\gamma = 100$ (Assumed to be Known)							
MLE	(100)	5.0179	0.3089	1.10012	258.48	2515.23	0.9811	4.7589
MMLE	(100)	5.0291	0.3480	1.12874	262.34	3393.65	1.1227	5.3221

$\hat{V}(\sigma) = 0.00077$ $\hat{V}(\mu) = 0.00119$ $\hat{Cov}(\sigma,\mu) = 0.00022$

	$\sigma = 0.3000$ (Assumed to be Known)							
LOCAL MLE	96.842	5.0387	(0.3000)	(1.09417)	258.21	2452.09	(0.9495)	(4.6449)
MMLE	84.146	5.1270	(0.3000)	(1.09417)	260.42	2926.13	(0.9495)	(4.6449)

$\hat{V}(\mu) = 0.00823$ $\hat{V}(\gamma) = 134.82$ $\hat{Cov}(\mu,\gamma) = -0.98280$

POPULATION VALUES	100	5.0000	0.3000	1.09417	255.24	2269.56	0.9495	4.6449

REFERENCES

[1] AITCHISON, J. AND BROWN, J. A. C., (1957). The Log-normal Distribution. Cambridge Univ. Press, Cambridge, England.

[2] BRADU, DAN AND MUNDLAK, YAIR, (1970). Estimation in lognormal linear models. Journal of the American Statistical Association, 65, 198-211.

[3] COHEN, A. CLIFFORD, (1951). Estimating parameters of logarithmic-normal distributions by the method of maximum likelihood. Journal of the American Statistical Association, 46, 206-212.

[4] COHEN, A. CLIFFORD, (1963). Progressively censored samples in life testing. Technometrics, 5, 327-339.

[5] COHEN, A, CLIFFORD, (1965). Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. Technometrics, 7, 579-588.

[6] COHEN, A, CLIFFORD, (1973). Multi-censored sampling in the three parameter Weibull distribution. Contributed papers, 39th Session of I.S.I., Vienna, Austria. Vol. 1, 202-207. Technometrics 17(1975), 347-351.

[7] GAJJAR, A. V. AND KHATRI, C. G., (1969). Progressively censored samples from log-normal and logistic distributions. Technometrics, 11, 793-803.

[8] HAMDAN, M. A., (1971). The logarithm of the sum of two correlated log-normal variates. Journal of the American Statistical Association, 61, 842-851.

[9] HARTER, H. LEON, AND MOORE, ALBERT H., (1966). Local maximum-likelihood estimation of the parameters of three-parameter log-normal populations from complete and censored samples. Journal of the American Statistical Association, 61, 842-851.

[10] HERD, G. R., (1957). Estimation of reliability functions. Proceedings Third National Symposium On Reliability and Quality Control.

[11] HILL, BRUCE M., (1963). The three-parameter log-normal distribution and Bayesian analysis of a point-source epidemic. Journal of the American Statistical Association, 58, 72-84.

[12] LEMON, GLEN H. (1975). Maximum likelihood estimation for the three parameter Weibull distribution based on censored samples. Technometrics, 17, 247-254.

[13] MOSHMAN, JACK, (1953). Critical values of the log-normal distribution. Journal of the American Statistical Association, 48, 600-609.

[14] MUNRO, A. H. AND WIXLEY, R. A. J., (1970). Estimators based on order statistics of small samples from a three-parameter lognormal distribution. Journal of the American Statistical Association, 65 655-659.

[15] NAUS, J. I., (1969). The distribution of the logarithm of the sum of two log-normal variates. Journal of the American Statistical Association, 64, 655-659.

[16] NORGAARD, NICHOLAS, (1975). Ph.D. dissertation, University of Georgia, Athens, Georgia.

[17] RINGER, L. J. AND SPRINKLE, E. E., (1972). Estimation of the parameters of the Weibull distribution from multicensored samples. IEEE Transactions on Reliability. R-21, No. 1, 46.

[18] ROBERTS, H. R. (1962). Life test experiments with hypercensored samples. Proceedings Eighteenth Annual Quality Control Conference, Rochester Society for Quality Control.

[19] TIKU, M. L., (1968). Estimating parameters of the log-normal distribution from censored samples. Journal of the American Statistical Association, 63, 134-140.

[20] WHITTAKER, E. T. AND ROBINSON, G. (1929). The Calculus of Observations. Blackie and Son Ltd., London and Glasgow.

[21] WILSON, EDWIN B, AND WORCESTER, JANE, (1945). The normal logarithmic transform. Review of Economic Statistics, 27, 17-22.

[22] WINGO, DALLAS R., (1973). Solution of the three-parameter Weibull equations by constrained modified quasilinearization (progressively censored samples) IEEE Transactions on Reliability, R-22, No. 2, 96-102.

[23] YUAN, PAE-TSI, (1933). On the logarithmic frequency distribution and the logarithmic correlation surface. Annals of Mathematical Statistics, 4, 30-74.

LIST OF REGISTERED ATTENDEES
TWENTY-FIRST DESIGN CONFERENCE
Washington, DC

Dr. Gerald R. Andersen	USA Materiel Development & Readiness Command Alexandria, VA 22333
Mr. David L. Arp	Naval Weapons Center China Lake, CA 93555
Dr. John C. Atkinson	Food & Drug Administration Washington, DC
Mr. Carl B. Bates	USA Concepts Analysis Agency Bethesda, MD 20014
Professor Robert Bechhofer	Cornell University Ithaca, NY 14853
Dr. N. R. Bhalerao	Mathematics Research Center Madison, WI 53705
Dr. Mark Brown	CUNY Scarsdale, NY 10583
Mr. Richard M. Brugger	USA Armament Command Rock Island, IL 61201
J. Robert Burge	Walter Reed Army Institute of Research Washington, DC 20012
John C. Burke	Department of Defense Ft. Meade, MD
Joseph M. Cameron	National Bureau of Standards Washington, DC 20234
Dr. George C. Canavos	Virginia Commonwealth University Richmond, VA 23284
Dr. Aivars Celmins	USA Ballistic Research Laboratories Aberdeen Proving Ground, MD 21005
Dr. Gary A. Chase	The Johns Hopkins University Baltimore, MD 21205
Professor A. Clifford Cohen	University of Georgia Athens, GA 30602
Stanley A. Cohen	Internal Revenue Service Washington, DC 20224
Paul J. Deason	USA TRASANA White Sands Missile Range, NM 88002

John S. Doyle	USA OTEA Falls Church, VA 22041
Dr. Francis Dressel	US Army Research Office Research Triangle Park, NC 27709
Dr. Eugene Dutoit	USA Infantry School Ft. Benning, GA
Cecil Eckard	Dugway Proving Grounds Dugway, UT
Dr. Churchill Eisenhart	National Bureau of Standards Washington, DC 20234
Ken Epstein	Army Research Institute Arlington, VA 22209
Harry Feingold	Naval Ship Research & Development Center Bethesda, MD
Stanley M. Finkelstein	Polytechnic Institute of New York Brooklyn, NY 11201
Dr. Walter D. Foster	AFIP Washington, DC 20306
Dr. Fred Frishman	Internal Revenue Service Washington, DC 20224
Dr. John J. Gart	NCI Bethesda, MD
Professor Edmund A. Gehan	University of Texas System Cancer Center Houston, TX 77075
Professor Seymour Geisser	University of Minnesota Minneapolis, MN 55455
Dr. Alan E. Gelfand	University of Connecticut Storrs, CT 06268
Dr. Arthur C. F. Gilbert	USA Research Institute Arlington, VA 22209
Harold Goldberg	US Naval Sea Systems Command Washington, DC
James B. Gose	Quality Evaluation Division White Sands Missile Range, NM 88002
Dr. Thomas N. E. Greville	Mathematics Research Center Madison, WI 53706
Paul Gunther	Internal Revenue Service Washington, DC 20224
Dr. Bernard Harris	Mathematics Research Center Madison, WI 53706

Dr. Boyd Harshbarger	Virginia Polytechnic Institute and State Univ. Blacksburg, VA 24061
Dr. John R. Kennessy	USA MERCD Ft. Belvoir, VA 22060
Dr. Lars Holst	Mathematics Research Center Madison, WI 53706
Dr. Theodore W. Horner	Statistical Consultant Bethesda, MD 20014
Professor J. S. Hunter	National Academy of Science Washington, DC
CPT J. Ireland	USA Security Agency Arlington, VA 22212
William E. Jackson, III	DNA, Armed Forces Radiobiology Research Inst. Bethesda, MD 20014
Dr. Alan E. Johnsrud	Systems Review & Analysis Office, ODCSRDA Washington, DC 20310
Ralph W. Jollensten	National Security Agency Ft. Meade, MD 20755
Van O. Jones	Aberdeen Proving Grounds Aberdeen, MD 21005
Eve S. Kaplan	WRAIR - Walter Reed Washington, DC 20012
Eric A. Kevitz	USA OTEA Falls Church, VA 22041
Karen M. Kirby	University of Maryland College Park, MD 20742
Harry H. Ku	National Bureau of Standards Washington, DC 20234
D. H. Kucha	Ft. Belvoir, VA 22060
Dr. B. Kurkjian	USA Materiel Development & Readiness Command Alexandria, VA 22333
Les Lancaster	USA OTEA Falls Church, VA 22041
CPT B. Lawless	USA Security Agency Arlington, VA 22212
M. W. Layard	National Cancer Institute Bethesda, MD 20014
Y. J. Lee	University of Maryland College Park, MD 20742

Sheldon G. Levin	DNA, Armed Forces Radiobiology Research Inst. Bethesda, MD 20014
Professor R. E. Little	University of Michigan Dearborn, MI 48128
Dr. Bruce J. McDonald	Office of Naval Research Arlington, VA 22217
Dr. D. H. McElhone	USA Security Agency Arlington, VA 22212
Dr. Ian W. McLaughlin	AFIP Washington, DC 20305
Dr. James R. Maar	Department of Defense Ft. Meade, MD 20755
Dale M. Madden	FDA Rockville, MD 20852
Dr. Clifford J. Maloney	Bureau of Biologics, FDA Bethesda, MD 20014
Ben Miller	Aberdeen Proving Grounds Aberdeen, MD 21005
Dr. J. Richard Moore	USA Ballistics Research Laboratories Aberdeen Proving Ground, MD 21005
Richard L. Moore	HQ, USA Armament Command Rock Island, IL 61201
Professor Frederick Mosteller	Harvard University Cambride, MA
Dr. Ben Noble	Mathematics Research Center Madison, WI 53706
Donald M. Neal	AMMRC Watertown, MA 01762
Dale E. Olsen	The Johns Hopkins University Laurel, MD
Beatrice S. Orleans	Naval Sea Systems Command Washington, DC
Dr. Madhar S. Phadke	Mathematics Research Center Madison, WI 53706
Dr. P. V. Piserchia	Research Triangle Institute Research Triangle Park, NC 27709
CPT L. D. Proegler	USA OTEA Falls Church, VA 22041
Dr. Ronald L. Racicot	Watervliet Arsenal Watervliet, NY 12189

Dr. James L. Raney	USA Research Institute Arlington, VA 22209
Dr. Stanley S. Reisman	New Jersey Institute of Technology Newark, NJ 07102
Steve Reynolds	USA OTEA Falls Church, VA 22041
CPT Bruce Rogers	USA OTEA Falls Church, VA 22041
Mr. Carl D. Rose	USA TACOM Warren, MI 48090
Dr. Edward W. Ross	USA Natick Development Center Natick, MA 01760
Dr. Joseph Rothberg	WRAIR Washington, DC 20012
Arthur D. Salomon	Internal Revenue Service Washington, DC 20224
Professor Thomas J. Santner	Cornell University Ithaca, NY 14853
Dr. M. A. Schneiderman	National Cancer Institute Bethesda, MD 20014
Seymour M. Selig	Office of Naval Research Arlington, VA 22217
Raymond J. Shadid	Internal Revenue Service Washington, DC
John Skory	Naval Underwater Systems Center New London, CT
Professor Herbert Solomon	Stanford University Stanford, CA 94305
Dr. R. Srinivasan	Temple University Philadelphia, PA 19122
Rick Steinheiser	Army Research Institute Arlington, VA 22209
Dr. Malcolm Taylor	USA Ballistic Research Laboratories Aberdeen Proving Ground, MD 21005
Jerry Thomas	USA Concepts Analysis Agency Bethesda, MD
Professor W. A. Thompson, Jr.	University of Missouri Columbia, MO 65201
Mrs. Kiroko S. Tokieda	NBS Washington, DC 20234

Gary D. Travers	Department of Defense Ft. Meade, MD 20755
Mr. Robert H. Turner	USA White Sands Missile Range White Sands Missile Range, NM 88002
Dr. John Van Ryzin	University of Wisconsin Madison, WI 53706
Mr. M. Waller	USA OTEA Falls Church, VA 22041
Dr. M. A. Weinberger	Operational Research Analysis Establishment Ottawa, Ontario, Canada K1A OK2
Gertrude Weintraub	Picatinny Arsenal Dover, NJ 07801
Dr. John Bart Wilburn	USA Electronic Proving Ground Ft. Huachuca, AZ 85613
CPT Gary S. Williams	USA OTEA Falls Church, VA 22041
Paul E. Williams	Aberdeen Proving Grounds Aberdeen, MD 21005
Linton S. Willis	Aberdeen Proving Grounds Aberdeen, MD 21005
Langhorne P. Withers	USA OTEA Falls Church, VA 22041
Margaret R. Wrensch	Letterman Army Institute of Research Presidio of San Francisco, CA 94129
Mrs. Patricia A. Young	Department of Defense Ft. Meade, MD 20755

*U.S. GOVERNMENT PRINTING OFFICE: 1976-646-470