UNCLASSIFIED

AD NUMBER

AD869266

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution: Further dissemination only as directed by the Naval Safety Center, Naval Air Station, Norfolk, VA 23551 or higher DoD authority.

AUTHORITY

NSC, per Dept. Navy ltr, dtd 17 Nov 1976

THIS PAGE IS UNCLASSIFIED

MDC J0698/01

AD869266

HUMAN ERROR RESEARCH AND ANALYSIS PROGRAM

SUMMARY AND FINAL REPORT

DATA HANDLING AND FATIGUE STUDIES

APRIL 1970

STATEMENT #5 UNCLASSIFIED

This document may be further distributed by any holder **only with** specific prior approval of Naval Safety Center
US Naval Air Station
Norfolk, Va 23511.

ENCLOSURE (1)

28-611U (REV 8-68)

MDC J0698/01

HUMAN ERROR RESEARCH AND ANALYSIS PROGRAM

SUMMARY AND FINAL REPORT

DATA ANALYSIS AND FATIGUE STUDIES

APRIL 1970

Prepared for the Naval Safety Center
Naval Air Station
Norfolk, Virginia

Under Contract No. N00189-70-C-0012

Requisition No. N00188-8290-2131

Prepared by:

Fatigue Studies

L. R. Creamer
Consultant
Science Research

D. E. Wheeler, Ph. D.
Life Sciences
Science Research

R. F. Gabriel, Ph. D.
Manager, Life Sciences
Science Research

Data Analysis

J. S. Tarr
Principal Investigator
Science Research

J. A. Baldwin
Science Research

H. J. Rostami
Science Research

APPROVED:
for A. A. Burrows, Ph. D.
Director, Science Research

DOUGLAS AIRCRAFT COMPANY

LONG BEACH, CALIFORNIA

MCDONNELL DOUGLAS CORPORATION

ABSTRACT

In 1965 the U. S. Naval Safety Center, Norfolk, Virginia funded the first of a series of studies of human error in aircraft accidents, entitled "Human Error Research and Analysis Program (HERAP)," because human error caused accidents still accounted for an average of 55% of the total and constituted a secondary cause of some of the remaining 45%, after twenty years of decline in accidents.

The program began with a detailed systems analysis of the Navy accident reporting program and the pilot training program. With the cooperation of several departments of the Navy a data bank was prepared, which contained personal data, training data and individual flight time data for each of 17,000 Navy pilots. It also contained all mishap data from 1962 to 1968. This data was merged into a single file and manipulated by various statistical programs developed or modified for HERAP. An Exposure Index was developed to equalize individual exposure to risk.

A study was performed to establish possible relationships between a pilot's mishap record and his education, training, demographic characteristics, and flying experience, and to test the distribution of mishaps of the pilots for randomness. Each pilot's mishap record was compared to the cumulative binomial distribution to establish a safety rating based on the probability that he should have had a worse mishap record by chance alone. A second safety rating was developed by comparing expected and actual mishaps for each pilot. Using a stepwise regression analysis, a study was made of the functional dependence of the safety ratings on the data. The highest multiple correlation coefficient was obtained by predicting the binomial ranking from the variables age, months of combat, last year of combat, rank, and type flying experience.

Fatigue, while frequently cited as a contribution factor in accidents remains an elusive element in terms of definition, recognition, measurement, prediction and alleviation.

Fatigue, defined as "The degradation in performance or affective state, resulting from previous work," was explored in 1968 - 1969 in a complex simulation experiment using an anti-submarine warfare flight trainer. Light work load over two hours and heavy work load over eight hours were imposed upon four professional crews. Each crew was in a controlled work environment 24-hours a day over a 2-week period. Psychological, physiological, biochemical, and performance measures were taken for individual and crew as a group.

Some statistically significant effects in the physiological and blood chemistry measures were found. Bartlett's hypothesis in regard to increased variability of performance and perceptual breakup, and the activation pattern hypothesis were to some extent verified, although it was indicated that heavier or longer loads would be needed for greater significance.

In a second study in 1969 - 1970, a state of fatique, induced by continuous performance of tasks over prolonged working time, was studied in a laboratory environment. The tasks were selected to resemble the airborne activities necessary to a prolonged ASW mission.

Two loading conditions were imposed such that 12 subjects in a High Work Load (HWL) group worked without rest for approximately 18 hours and 12 different subjects in a Low Work Load (LWL) group worked for the same time span, but were permitted equal time on and off work, such that their workload was approximately one-half that of the HWL group. Psychological, physiological, biochemical, and performance measures were taken for each individual subject.

Performance decrement was obtained for the HWL reflecting at least moderate fatigue. A specially designed Discrete Tracking Task was the most sensitive index of fatigue. The task was moderately difficult with almost excessive temporal demands. The HWL group also made larger time estimation errors late in the session than the LWL group. No other psychological,biochemical, or physiological variable was correlated with performance decrement (Fatigue) in this study.

TABLE OF CONTENTS

TABLES AND FIGURES

MDC J0698/01

HUMAN ERROR RESEARCH AND ANALYSIS PROGRAM

SUMMARY AND FINAL REPORT

PART I, SUMMARY

APRIL 1970

Prepared for the Naval Safety Center
Naval Air Station
Norfolk, Virginia

Under Contract No. N00189-70-C-0012

Prepared by:

J. S. Tarr
Principal Investigator
Science Research

J. A. Baldwin
Science Research

H. J. Rostámi
Science Research

Approved by: for A. A. Burrows, Ph. D.
Director
Science Research

DOUGLAS AIRCRAFT COMPANY

LONG BEACH, CALIFORNIA

MCDONNELL DOUGLAS CORPORATION

PART I

BACKGROUND SUMMARY

INTRODUCTION

In July, 1965 Douglas Aircraft Company, Long Beach, California, submitted a proposal to the Navy Safety Center, Norfolk, Virginia, for a short-term investigation of, and planning of, a long-term study of an Aircraft Accidents and Development of Accident Prevention Program. The study, entitled, "Human Error Research and Analysis Program (HERAP)", was the first of a series of HERAP studies performed since that time. This report reviews and summarizes those studies.

A BRIEF HISTORY OF NAVAL AIRCRAFT SAFETY

In the first few years of Aviation success was the criterion by which flights were judged. If it worked, it was repeated - if not, changes were made. This same criterion was dominant when the U.S. Navy formed its first Aeronautical Organization under Captain Washington Chambers in October of 1910. In November, 1910 Eugene Ely made the first take-off from the deck of a ship, and in January, 1911 landed a hydroplane beside a ship, was hoisted aboard, lowered into the water, and took off to successfully return to shore.

Safety was a conscious part of these efforts - by 1912 Captain Chambers was speaking of the need for mechanical devices to aid the pilot "which will perform . . . the things he is prone to do indifferently, especially under the strain of fatigue . . .". In that same year, Captain Chambers brought about the first catapult launching - Naval Aviation was on its way.

By 1916, efforts were well under way to standardize personnel and aircraft design, production, inspection, and testing. While numerous safety orders and regulations were issued, one essential item - thorough record keeping - was lagging. The earliest preserved accident record was a 5 x 6 card written on one side.

An Aviation Medical Research Board was formed in 1917 and immediately established a Research Laboratory at Hazelhurst Field, Mineola, Long Island. In 1939, the Navy established a School of Aviation Medicine and Research at Pensacola, Florida. Throughout the years, from the end of the World War I, to the beginning of World War II, standards for the selection and training of pilots continued to improve. However, Armstrong (1952)* states, "The increased performance of aircraft attained by about 1930 resulted in a situation wherein the human element in flight was becoming the weakest link in the chain!" A remark which appears to be no less true today.

THE BACKGROUND FOR THE HERAP STUDIES

As early as 1955, the Navy Safety Center had requested assistance in planning for personnel and equipment to improve the analysis of Navy Accident Data. The resulting report stimulated further efforts by Safety Center personnel and culminated in 1965 when additional personnel and equipment were authorized and the HERAP effort was funded.

(*) As quoted in HERAP, First Planning Study, DAC, 1966

The HERAP effort was based on the fact that while the general trend of Navy aircraft accidents was downward, the percent of those accidents attributed to human error was consistently greater than 50% of the total. More than two-thirds of the Navy aircraft accidents have human factors listed as a contributing cause. The effort expended to reduce this factor, however, has been historically too little in proportion to its importance. The effort to reduce accidents and their associated costs was given added impetus when the all Navy accident rate rose in Fiscal 1966 after many years of annual reduction.

Reference 2 describes the change as follows:

"The downward trend in the All Navy Accident rate was interrupted in Fiscal 1966. While flying 3,739,856 hours, 476 accidents occurred with a resulting accident rate of 1.27 compared to 457 and a rate of 1.25 in Fiscal 1965. Although the rate increase was very small, Fiscal 1966 was the first year since the Korean War years that an improvement was not experienced in the All Navy rate, making this the third time in twenty years that the accident rate has increased (the last two occurrences were in Fiscal Years 1951 and 1952)."

The primary objectives of the HERAP investigations were:

a. to identify and describe those characteristics which differentiate successful pilots from unsuccessful pilots in terms of their susceptibility to accidents and their contribution to fleet readiness;

b. once identified, to order these characteristics into meaningful relationships for the purpose of prediction and control.

The research objectives included short, intermediate and long-term investigations leading toward the ideal goal of eliminating aircraft accidents.

The HERAP Planning Study

The first HERAP Planning Study investigated information sources and requirements within the Naval establishment. A preliminary mission-system-task analysis was performed on an A-4 aircraft ferry mission to demonstrate its applicability to human error accident investigations. Data reporting and analysis procedures were reviewed and the literature on accidents was reviewed. A program for the development of improved data gathering was proposed, as well as a computer system adequate to the task facing the Naval Safety Center (NSC).

One of the critical elements of the Planning study was an analysis of resource requirements sufficient to predict measurable gains in understanding and alleviating human error as a cause of accidents. The generalized model of the plan in shown in Figure 1. Also recommended were five critical areas of concern about which sufficient knowledge is lacking, but which are generally believed to be closely related to human error. Among these are fatigue, disorientation, carrier landing, collision avoidance and an index of exposure for flying in the U.S. Navy.

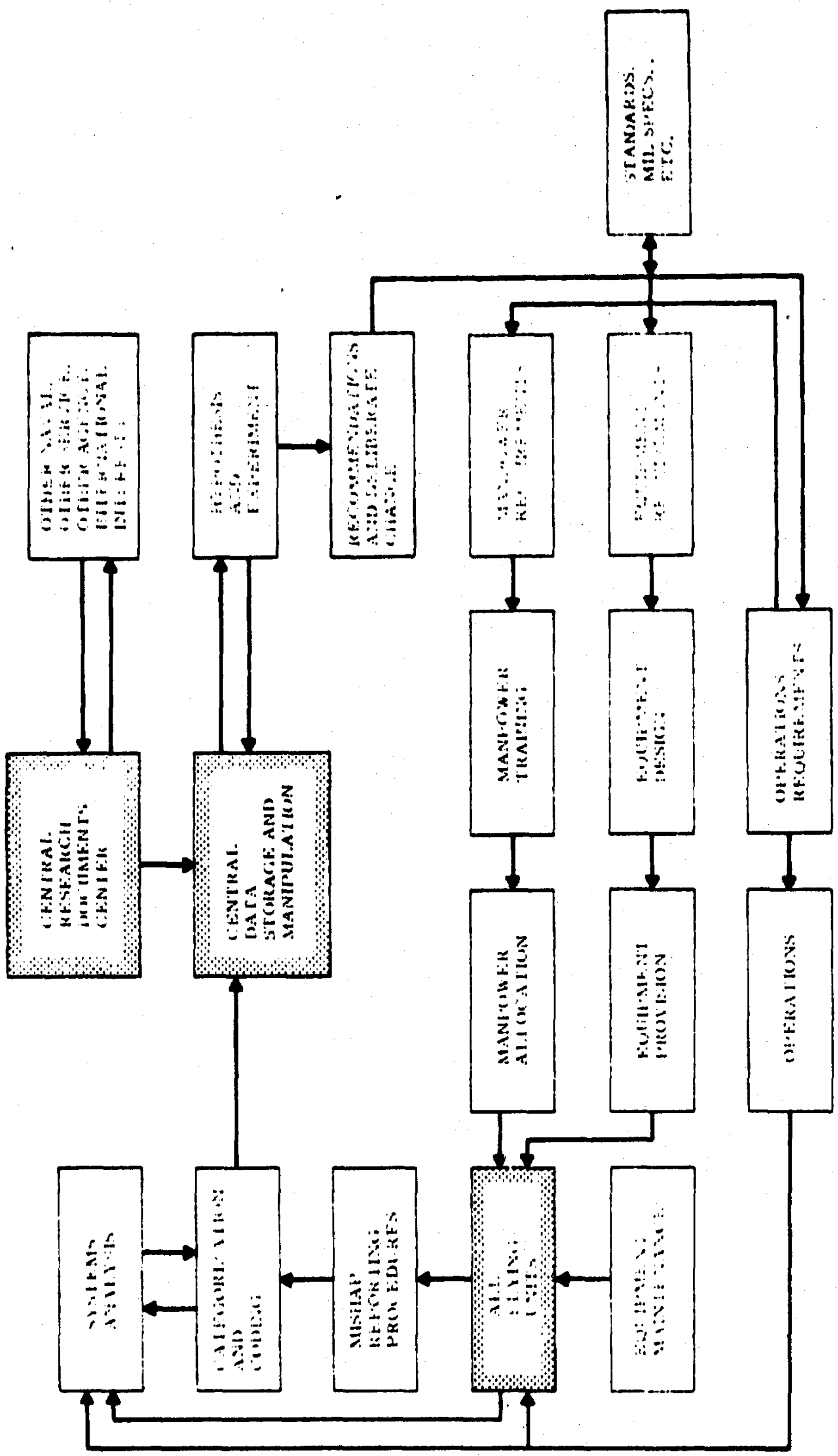


Figure 1. General Model of Human Error Research

The Philosophical Foundations of HERAP Research

During the initial HERAP investigations, in which many activities interested in the Naval Aviation safety were visited, viewpoints on accident reduction due to human error could be summarized in three philosophical approaches.

The first was one which states that a method to reduce the frequency of mishaps is to detect and remove personnel who for one reason or another are not proficient enough for the demanded flying tasks. A second was to detect and remove the equipment and procedural stressors which in some way caused personnel to fail in their tasks. A third was that problems causing mishaps would show up in data in an empirical fashion and special studies could then be instigated to deal with them. Each of these points of view has a place in any systematic program, but none stands on its own feet.

In the first case, proficiency is a function of innate characteristics, training and task difficulty. If selection chooses personnel whose capabilities are adequate to be modified by training procedures to a level sufficient for the defined task, and the training is itself adquately structured with regard to the task, detection and removal of personnel for lack of proficiency will apply only to those who are not performing when they well could; that is we are confining ourselves to motivational and command deviations. On the other hand, if selection and training are not adequately matching personnel and task, the problem of urgency is not only detection and removal of personnel, but of restructuring selection and training procedures.

In the second case, the equipment factor follows the same argument. Equipment which induces operational mishap through poor design should most certainly be modified and replaced, but there will always be an operator who can use it without difficulty. How can we insure that equipment is designed for the capabilities of operators without knowing something about expected performance of our population over the range of equipment?

The third is the most common approach. It is reflected in the wide variety of data categories found in the files of organizations concerned with safety. Categories are based on detected hazards and not on predicted ones. Where this is true, the range of variables connected with a particular item might be large in real life, but the range used for study is inevitably small.

The approach in the HERAP studies was one which stated that we should establish knowledge of the capabilities of the individual and grouped personnel in regard to the current and predicted tasks which they must perform. For personnel already in the system, data existed from selection, training, readiness and operational sources as to capabilities of certain sorts. It was possible that by statistical study we might relate mishaps to these measured capabilities. Some of this work had been done and more begun, notably at Pensacola. But knowledge of capability in regard to future and predicted specific tasks had not been carried out except in isolated studies. Thus we attempted to isolate the tasks in some systematic way, update these over the long term, and establish levels of expected performance for them.

HUMAN ERROR

It is of little use to attempt the description of a long range accident prevention program related to human error without a working definition of what human error is. Like many terms, this one has a stigmatic connotation related to "making a mistake" and this is unfortunate, for "making a mistake" is somewhat a dichotomous notion assuming that there are only two conditions of consideration, right and wrong.

Because, until now, the primary effort in human error research has been an effort to aid in the design of reliable systems, it has been concentrated in the areas of system design and development. Studies by Cooper, 1961; Meister, 1962; Willis, 1962 gave impetus to the consideration of the impact of human error on system performance. It thus became necessary to measure the operators potentiality for failure (that is, his potentiality for error) so that this could be combined with those of other system components, to give a true indication of total system reliability. Consequently, human error has been defined within the framework of system design and development, due to these utility goals.

Human error in the usual context of the operator, is necessarily a wide all inclusive term which emphasizes the frailty of the operator without recognizing that there is a "human system" which can fail. The human system failure is pinpointed when the operator does not complete, successfully, the required operation. The "human system" complexity has stymied many studies, all generalized solutions, and made significant progress in corrective actions very slow.

Human error in this context, has been defined in terms of four categories:

1. Incorrect performance of a required action

2. Failure to perform the required action

3. Out-of-sequence performance of a required action

4. Performance of a non-required action.

These are actually categories of mishaps and are constrained by one qualification: the error must have some significant effect upon the system as a whole or upon some behavioral component of the system or it must be considered as not significant. Further, "reversible" errors corrected prior to system damage are not significant. Human Error, in this context, then, has been defined as any deviation from a system performance standard which is caused directly or indirectly by an operator and which has significant consequences to the system operation in which it was made.

While it is worthwhile to take the categorical definitions of human error used in system design to work with in HERAP so that a continuity of definition exists, it would be remembered that HERAP is concerned with operational systems in being and both "reversible" and "active" error must be considered. A reversible error today may be the seed of tomorrow's accident, as in a near miss. Human error research in systems design is concerned with quantification in a manner similar to equipment, so that reliability estimates for design can be made. In the prevention of accidents in the operational phase of a system, some other categories must also be considered.

Some current Methods of Human Error Analysis and Predictability have been presented by:

1. Altman et al, 1964
2. Swain and Rook (1962 and 1963)
3. Miller et al, 1964, and
4. Irwin et al, 1964.

All but Irwin's are based upon Altman's "A.I.R. Data Store." This store was probably the first attempt to manipulate performance data for use in the development of an operability index for prediction of human performance. As Altman himself points out, it is generally highly limited, but for those specific tasks listed, it is considered to be excellent. The operability index was built with three objectives:

1. It would be task oriented.
2. Scores from the operability index would be directly meaningful in terms of time and error of performance (speed and reliability).
3. Data on speed and reliability of operator performance would be obtained, insofar as possible, from available experimental and field data.

The basic units for evaluation are the task elements, which are defined as inputs, mediating processes, and outputs.

The method of using the data store was as follows:

1. Equipment design information and data obtained from task analyses must be analyzed into behavioral steps and sequenced by mission phases.

2. Each operation must be subdivided to its relevant components, parameters, and dimensions for each behavioral step.

3. Each of these dimensions must be assigned a reliability estimate from the data store. These estimates are then used to derive scores for units of behavior, such as steps, tasks, phases and entire missions.

In general, time scores were the sum of individual time estimates for each dimension, while reliability scores were the product of individual reliability estimates for each dimension. The total output was a reliability score which was an estimate of the probability that serious operator error would not occur within the span of performance encompassed by the score.

THERP (Technique for Human Error Rate Predictions), was an extended utilization of Altman's Data Store by Swain and Rook in the form of an iterative procedure of 5 steps which were repeated, (though not always in the same order), until the system degradation resulting from human error was at an acceptable level. These five steps were:

1. Define the system or subsystem failure which is to be evaluated.

2. Identify and list all the human operations performed and their relationships to system tasks and functions.

3. Predict error rates for each human operation or group of operations pertinent to the evaluation.

4. Determine the effect of human errors on the system

5. Recommend changes as necessary to reduce the system or subsystem failure rate as a consequence of the estimated effects of the recommended changes.

As Swain points out, the steps are typical of the usual system reliability study, "if one substitutes Humans for Hardware."

A third method, developed by Miller employed a computer technique to generate distributions of operator performance times. Task performance is simulated by sampling the distribution of times, using Monte Carlo methods, during 10 computer runs.

Miller's approach to operator simulation was based on four assumptions:

1. The human transfer function depends upon the individual's capability, his operating environment, and the nature of the task requirements. Therefore, to describe the operational

behavior of a man-machine system req res the simulation of groups of operators and maintenance personnel of varying capability, degree of training and level of motivation, operating under the range of variability of conditions and stresses encountered in the operation environment.

2. Individuals tend to perform in a nonlinear fashion. In deriving the model varying paths of task accomplishment were provided, the number of paths being a function of all possible alternatives of task execution available to the operator.

3. The performance of a subtask by operators classified by skill level can be described by time and success probability distributions, and these distributions approach normality. The assumption of normality is not a limiting factor but merely one of convenience.

4. The path of task accomplishment, the performance times, and the success probabilities were dependent on psychological stress induced by the operating environment, task accomplishment, and work load, and this stress could be an organizing or disorganizing agent on behavior, depending on stress level and operator stress threshold.

The method developed by Irwin was one in which judges ratings were used to develop human reliability data estimates when other data (such as the data store), were not available. His method was as follows:

1. Specify tasks that must be performed.

2. Identify the task elements necessary to accomplish the total task.

3. Determine the performance reliability for each task element.

4. Perform a preliminary rating study.

5. Perform a man-rating study.

Though the use of a rating scale may seem to be the crudest form of elemental psychophysics, work by Whitlock indicates that a proper choice of judges (more than 6) utilized, and no more than 40 items to be rated results in surprisingly high validity when tested.

SOME PROBLEM AREAS

One of the major problem areas for HERAP, which is less for system design predictability, is the consideration of additional data that may be pertinent to human error prevention where mishaps are concerned. A partial listing from Kock and from Sells of some of the stimulus variables to be considered are:

1. Natural aspects of the environment:

Weather-temperature, humidity, seasons, climate
Terrain-altitude, erosion
Food
Clothing
Natural resources

2. Description of the task situation:

   Area and level of knowledge and skills required
   Hazards and risks involved
   Novelty of situation in relation to prior experience
   Procedures permitted
   Information required and available
   Number of participants present and available
   Material and facilities required and available
   Degree of personal contact involved
   Role expectations of other persons concerning the individual

3. Definition of the individuals relation to the situation:

   Degree of freedom vs. restriction in group activities
   Degree of competion vs. cooperation required
   Degree of friendliness vs. hostility required
   Status hierarchy position required

4. Description of other persons in the situation:

   Age and sex of participants
   Social and economic background of participants
   Skills, abilities, and experience
   Training and motivation
   New or previous acquaintances
   Pre-existing relationships

5. External reference characteristics of the individual:

   Biological - sex, age, physique
   Physical abnormalities or injuries
   Race
   Education
   Marital status
   Citizenship, legal restraints
   Geographic position-rural, urban, national
   Social status
   Income and occupation
   Residence and transportation
   Debts and savings
   Status and role in family group
   Parents - religion, age, health, language

Siblings - age, sex, type
Primary group and children
Group memberships - number, type structure
Status and role in group

6. Individuals performing relative to others:

Formal group structure - goals, objectives, and facilities
Membership requirements - age, sex, race
Requirements concerning experience, training, achievement
Group's control of members - SOP's, regulations
Group procedures - meetings, staffing
Dependency/cooperation of other groups
Formal role in group - prestige, power, privilege
Site requirement of group - location, space facilities
Informal group structure - goals and objectives
Group member ship requirements - social class, demography
Group control of membership
Regulations of group procedures - activities, participation
The group's significance to its members
Cohesion of the group, functional relationships
Collective situations - crowds, mobs, audience

This partial list indicates the great number and range of stimuli that are suggested to be relevant.

If the term "human error" is to survive over a long term program, then a more precise definition is required which reduces value judgments to their proper place as adjuncts to objective measures which can effectively be used.

The definition proposed for this purpose was "Human error is the specified deviation of a human activity or decision from an operationally defined norm."

This says, in effect, that an error cannot occur unless limits have been set before a mishap occurs. This is a necessary condition for setting up any methodology for recording, manipulating and acting upon data in a meaningful way. In this way, too, it can be argued that although the operational definitions may be modified and their number increased with time, no groundless judgments can be made.

In Naval aviation, there are many circumstances where mishaps occur in conditions where it is perfectly clear that the human operator has had no chance to mitigate or even influence the result. In these cases, there is no question of immediate mistakes, although relationship to a remote, prior influence may be traceable. It is a truism that in any system failure the failure may be ultimately traced to a human source. If there is a wing root failure or an equivalent physiological catastrophe, we may consider these conditions unmodifiable by the human operator.

Between these limits are circumstances which are more difficult to describe and define. Considerably more data, definition and manpower are, and have been, involved in the study of equipment failure compared with human "failure." This is not surprising, for methods of measurement, definable variables and access, each provide greater ease of manipulation in this area. Yet, as we probe deeper into the equipment failure concepts, we find that here, too, dichotomous judgments are not acceptable in most cases. The concept of "tolerance" states that an item has been designed and should have been produced and tested, so that its performance is related to its function in such a way that failure should not occur unless the tolerance limits are exceeded. If the item fails to perform as it is specified and required to do, it may have done so because either it does not meet the tolerance limits in its quality or that these limits have been exceeded in operation in such a way that it is probable that it would fail. This, again, may be related to the conditions that the tolerances have not been set correctly for the function required to be performed, or that the item has not been tested to determine if it does fall within the limits set. It is clear that here there is a range of performance capability set up which allows some operational freedom in the use of the item. One may also expect that this range may be described in a distributive manner, that is, the specified quality or qualities can be shown, over large samples, to fit some descriptive transfer function, such as a Gaussian or similar relationship.

Furthermore, in equipment parlance, there also occurs the concept of expected failure, related to use over time. In this, a historic or empirical approach will result in probability judgments that there is such and such a "mean-time-to-failure" for this item, or some similar predictive statement. Here, items may be replaced before some value of the distribution of this parameter is reached on the assumption that actual failure of the item will be reduced. Alternatively, inspections may be made at specific points along this distribution in order to search for indications that an item is approaching tolerance limits with time and use.

If we are to consider the effects of the human operator within the system, it is necessary to interface with the methods and descriptions used in dealing with the equipment he uses, for few mishaps occur within the "unmodifiable" category described earlier and there is very little that can be done about them within our terms of reference. On the other hand, the greater part of our study must fall within the interface area where both operators and equipment are contributing to activity and decision within a system.

In summarizing the equipment "approach", however, there is little to disagree with, and nothing which contradicts our proposed definition of human error. Qualities in operators can be shown to be distributed across samples, to have tolerances which can be exceeded and result in failure, can be tested and inspected before and during use, can be "mis-matched" where the use does not correspond to the quality selected and specified. This, of course, does not imply that we should treat human operators in the same way that we do equipment, there are other descriptors and qualities which will not fall within these contexts (such as motivational factors), but we can assure ourselves that commonality over much of our study area can be assured for the purposes of prediction, description and diagnosis, with equipment methods, allowing our interface to occur without too much difficulty.

In some recent studies of "human error", notably in highly structured weapon systems, it has become common to attach criticality values to error incidence. That is, it is recognized that an "error" (in the loose sense) can occur in any of n circumstances, but that in some circumstances it does not matter and in others it will matter a great deal. Thus, error probability is, in these approaches, linked to a level of criticality. In the context of everyday military aviation, we may illustrate this by stating that there is a known and finite probability of misreading a specific altimeter, by a specific population of pilots, with an error of 1000 feet. At higher altitudes it is unlikely that this will be "critical" (except under special circumstances), but at 1000 feet altitude it certainly will. From this point of view, it is necessary to relate the error to the activity in the operational context; that is in the "equipment" context, the "tolerance" of the quality of instrument reading capability could be variable. It is, however, easy to be confused with the worth of this concept. In a highly structured weapon system where there are probably many operators, and these can be selected and trained and utilized in different ways, it may well be valuable to introduce a criticality factor. In military aviation, where numbers are fewer, where training and selection are common to the population, and where the pilot, for example, has considerable freedom in the operation of his vehicle system, there may be less value. In this case, "error" must be reduced to a minimum, for all occasions, because it is not always known when it might or might not be critical. A further point, especially in this example, is that the pilot may not (and more often than not, does not) know that he has misread unless something results from the misreading. The attachment of criticality measures then, should be considered with care in the context of aviation.

Our approach then was to attempt to specify what we expected of our human operators in precise qualities which they must possess to carry out specific tasks in specific systems. Only in this way could be determine when something unusual had happened. There are various ways of setting about this. Unfortunately, we were not beginning from the start of an organization, we were forced to deal with an existing, highly formalized structure. Thus, we were necessarily empirical in much the same way that equipment interests are. What are the distributions of qualities among the existing population of aviators and other associated personnel in regard to the tasks they now have to perform?

"Error" was examined then, from the point of view of a deviation from what could be expected, but blame was not necessarily attached to the individual for the deviation.

Once again, commonality between equipment failure analyses and human error analyses techniques should be borne in mind. This commonality should permit relatively easy integration of a larger human error and existing equipment failure and safety program.

THE HERAP STUDIES

As a result of the HERAP Planning Study several areas of safety research related to human error were recommended as especially suitable for immediate investigation. They included fatigue, and accident exposure index and carrier landings. The results of those studies are described elsewhere in this report.

The first planning study (HERAP I)* included a representative task analysis with comments and recommendations for further uses of this method as an analytical tool. HERAP II included an extensive task analysis of an A-4 mission and beginning development of data handling and analysis.

HERAP III included an extensive study of the relationship of flight phase to accidents. Task analyses of representative U. S. Navy/U. S. Marine Corps A-4 aircraft missions were made. The missions were divided into static-incident to flight, taxi, takeoff, inflight and landing phases. The landing and takeoff phases each have separate analyses made for carrier and field operations.

Computer tabulations were made of A-4 accidents and incidents for fiscal year 1967. The mishaps were tabulated to graphically indicate the number of accidents and incidents which occurred during each phase of flight. Analyses were made for the A-4 A/B/C models and for the A-4 E/F models. Comparisons were made of the numbers of mishaps occurring for the various flight phases. This work is summarized in Appendix A, Figures 1 through 8.

The causes of mishaps are summarized under field and carrier operations and under flight phases. Further analyses were made of pilot error causes of mishaps for both carrier and field operations during landing.

A task analyses form was developed to show graphically the phases of flight which have high mishap potential. Results of the analyses show that the greatest number of mishaps occur during the landing phase. Carrier operations show a higher rate of mishaps than field operations which is a generally accepted fact. A-4 aircraft accidents are summarized in Tables I and II.

*The HERAP studies have arbitrarily been assigned number I, II, III, and IV to differentiate between contracts as follows:

HERAP I	Contract N189-(188)-59791A	8/65 to 2/66
HERAP II	Contract N00189-67-D0329	9/66 to 9/67
HERAP III	Contract N00189-68-G0565	10/68 to 3/69
HERAP IV	Contract N00189-70-C-1321	10/69 to 4/70

TABLE I

COMPARISON OF PILOT ERROR CAUSE TO OTHER CAUSES OF MISHAPS

FOR THE OVERALL FLIGHT PHASES AND FOR

THE LANDING PHASE SEPARATELY

A4 AIRCRAFT

	ACCIDENTS		INCIDENTS	
A. FIELD OPERATIONS:				
PILOT ERROR	24	36.9%	110	25%
OTHER CAUSES *	41	63.1%	330	75%
TOTALS:	65	100%	440	100%
B. CARRIER OPERATIONS:				
PILOT ERROR	27	20.45%	27	18.49%
OTHER CAUSES	105	79.55%	119	81.51%
TOTALS:	132	100%	146	100%
C. FIELD PLUS CARRIER OPERATIONS:				
PILOT ERROR	51	25.89%	137	23.38%
OTHER CAUSES	146	74.11%	449	76.62%
TOTALS:	197	100%	586	100%
D. LANDING PHASE - FIELD OPERATIONS				
PILOT ERROR	7	50%	25	39.84%
OTHER CAUSES	7	50%	38	60.16%
TOTALS:	14	100%	63	100%
E. LANDING PHASE - CARRIER OPERATIONS				
PILOT ERROR	19	54.29%	9	34.61%
OTHER CAUSES	16	45.71%	17	65.39%
TOTALS:	35	100%	26	100%

* Other Cause Factors include combat, airbase/carrier facilities, other personnel (maintenance/supervisory), material failure/malfunction, weather, unavoidable and undetermined/other causes.

TABLE II

COMPARISONS OF ACCIDENTS AND INCIDENTS BY FLIGHT PHASES

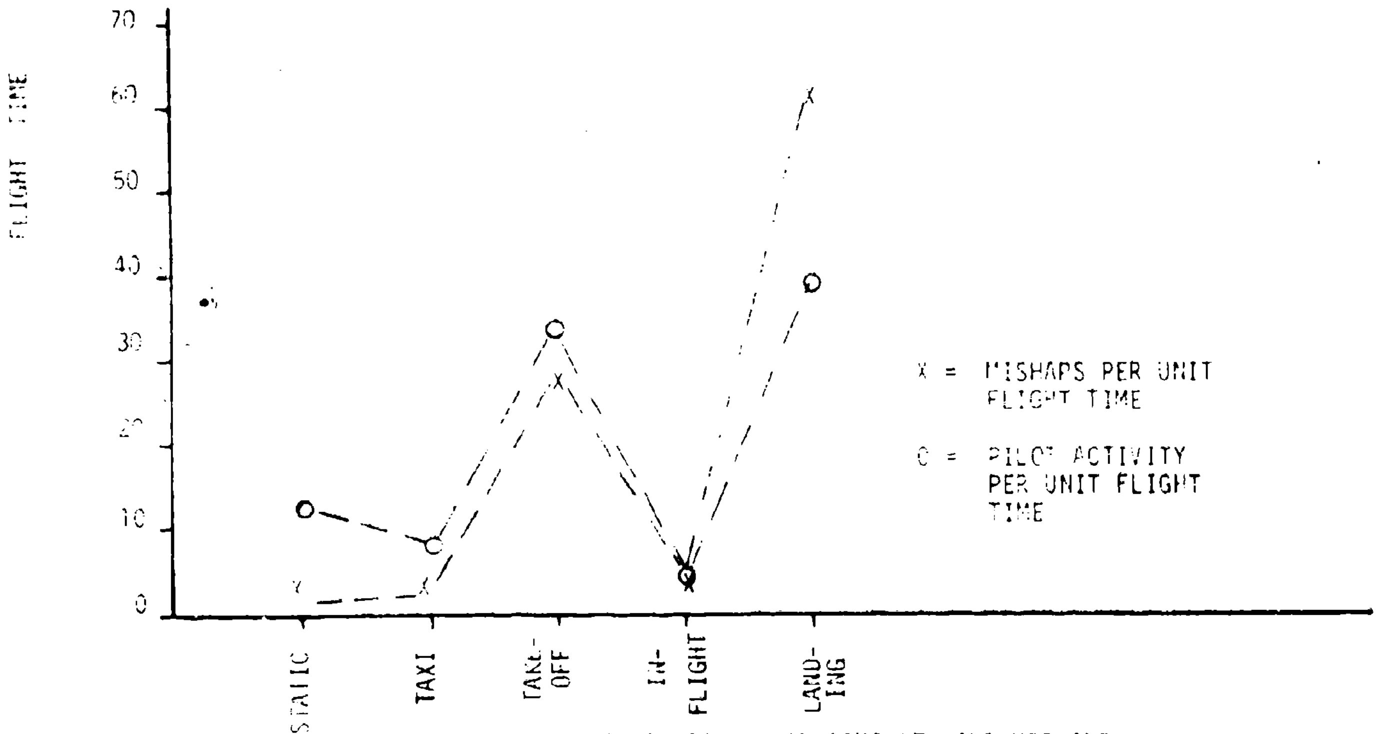
	A-4 A/B/C/E/F AIRCRAFT							
	FIELD OPERATIONS				CARRIER OPERATIONS			
FLIGHT PHASES	ACCID	%	INCID	%	ACCID	%	INCID	%
Part A:								
1. STATIC	0	0	9	2%	0	0	3	2%
2. TAXI	1	1.5%	11	2.5%	2	1.5%	3	2%
3. TAKE OFF	5	8%	36	8%	8	6.02%	15	10%
4. INFLIGHT *	42	65.5%	316	72%	87	65.41%	99	68%
5. LANDING	14	22%	63	14.3%	35	26.32%	26	18%
6. WAVE OFF	0	0	3	0.7%	0	0	0	0
8. LET DOWN	2	3%	2	0.5%	1	0.75%	0	0
TOTALS FLIGHT PHASES:	64	100%	440	100%	133	100%	146	100%
Part B: Not Incident To Flight and Undetermined Phases Added:								
9. NOT INCIDENT TO FLIGHT	57	48%	4	0.88%	23	15%	1	0.66%
10. UNDETERMINED	1	0.8%	12	2.6%	1	0.6%	5	3.3%
TOTALS ALL PHASES:	122	100%	456	100%	157	100%	152	100%
* **Part C:** INFLIGHT DIVIDED BETWEEN:								
A. ENEMY ACTION, AND,	15	33.7%	87	27%	68	78%	36	36%
B. NOT ENEMY ACTION	27	66.3%	229	73%	19	22%	63	64%
TOTALS INFLIGHT:	42	100%	316	100%	87	100%	99	100%

* INFLIGHT MISHAPS, PART A-4, HAS BEEN FURTHER SUBDIVIDED IN PART C, A, AND B.

The study showed that a significant number of landing mishaps are caused by pilot errors in the use of engine power, use of the brakes, control of sink rate, and in landing with the wheels up.

The results of the study substantiate the general belief that the work load is heaviest during critical phases of flight. Reduction of pilot work load during

WORKLOAD VERSUS ACCIDENT RATE



THE CORRESPONDENCE BETWEEN PILOT ACTIVITY AND MISHAPS

Figure 2.

critical phases of flight, i. e., reduction of pilot workload during landing, takeoff, and attack phases might result in a reduced number of pilot-induced mishaps, but deeper study based on more information, is needed.

Fatigue Study

Fatigue is frequently cited as a major contributor to accidents. The wide acceptance of this assumed relationship is not surprising. Most people have experienced "fatigue" more than once, and subjective feelings of drowsiness, apathy, weakness, etc., certainly "ought" to be precursors of accidents.

Operational requirements frequently make it necessary or desirable to program extended and/or very frequent activities. There is undoubtedly a limit to the workload that men can assume without seriously influencing the safety and/or probability of successful completion of the assigned task.

It would be desirable to supply medical personnel and leaders with valid and reliable guidelines of performance changes, as well as predictors or symptoms of these changes accruing for a given workload. Attempts to define the effects of fatigue under controlled conditions, however, have been generally unsuccessful.

Among the factors that have probably contributed to the lack of definitive results obtained in studies of fatigue are the following:

1. Complexity of the human organism, with its ability to draw upon extensive resources.

2. Vagaries of motivation and physical condition.

3. Effects of age and training.

4. Similarity of the symptoms of fatigue and those of illness.

5. Difficulty of separating the effects of other types of stress from fatigue.

In order to delimit the scope of the problem, fatigue, for the purpose of this study, was defined operationally as "the degradation in performance or affective state resulting from previous work." The purpose of this study was to initiate a systematic program intended to ultimately achieve the following goals:

1. Determine performance parameters (in tasks representative of those required of Navy personnel), influenced by continued heavy workload.

2. Determine the workload which results in the onset of a degradation in performance.

3. Determine the magnitude of impairment for two selected workloads.

4. Define physiological and/or behavioral indices of the onset of fatigue.

5. Develop an understanding of the underlying mechanisms involved in performance degradation.

6. Discover methods of mitigating or alleviating degradation without adverse side effects.

Conclusions

1. The subjects were fatigued only slightly by the conditions encountered in the high WL/treatment.

2. No mean performance changes resulted in any performance measure.

3. Pilot tracking data supported the increased variability and perceptual breakup hypothese.

4. Physiological measures generally reflected changes in workload. Particularly promising are respiration rate and sleep recovery data.

5. The pattern hypothesis was supported but cannot be used indiscriminately. The most affected parameters must be determined for the combined data treatment to be effective.

6. The phorias are worth investigating in greater depth as indices of fatigue.

Recommendations

1. A parametric study of work load should be conducted to determine the amount or the duration of work that can be performed before serious performance degradation results.

2. Control over the complete situation for basic studies would be easier in a non-operational situation. Therefore, serious consideration should be given to conducting additional exploratory studies using non-operational or, at least less complex, tasks.

3. An inflight study should be performed to verify the 8-hour mission duration acceptability for tactical ASW crews in aircraft such as the S-2.

MDC J0698/01

HUMAN ERROR RESEARCH AND ANALYSIS PROGRAM

SUMMARY AND FINAL REPORT

PART II, DATA ANALYSIS

APRIL 1970

Prepared for the Naval Safety Center
Naval Air Station
Norfolk, Virginia

Under Contract No. N00189-70-C-0012

Prepared by:

J. S. Tarr
Principal Investigator
Science Research

J. A. Baldwin
Science Research

H. J. Rostami
Science Research

Approved by:
A. A. Burrows, Ph. D.
Director
Science Research

DOUGLAS AIRCRAFT COMPANY

LONG BEACH, CALIFORNIA

MCDONNELL DOUGLAS CORPORATION

TABLE OF CONTENTS

PARTS I & II

TABLES AND FIGURES

PART II

MISHAP DATA ANALYSIS

INTRODUCTION

A considerable body of data describing the personal histories and flying experience of Navy pilots has been built up over the past years. These data range over a wide variety of factors which could conceivably be related to how safely pilots perform. The main thrust of the HERAP data analysis program for the 1969 - 1970 contract was to systematically test the existing data for significant relationships with the criterion of "pilot safety."

Linear regression (linear least-squares curve fitting) was used to test for these relationships. The variables on the data bank were used to try to predict a numerical measure of flying safety for each pilot. The error in the predictions were then analyzed; that is, the discrepencies between predicted and observed pilot safety were analyzed. If there were no error, the predictions would be perfect and it could be confidently said that there was a relationship between pilot safety and at least some of the variables on the data bank. If the error in the predicitions turned out to be as great as the intrinsic scatter in the observed pilot safety measure, it could be said with equal confidence that no such relationship existed. The actual situation is almost always somewhere between these two extremes, but statistical techniques can be employed to allow some statement about the probability of a relationship existing between the measure of pilot safety and the data on individual pilots.

This analysis proceeded in three successive steps: development of a data bank, development of a measure of pilot safety, and the actual regression analyses. These are discussed in the following sections.

THE DATA BANK SYSTEM

In the years of its existence, the Naval Safety Center has collected an enormous amount of data. These data have served as the source of numerous studies. As the data have grown, so have the problems of using them. The sheer quantity had made comparative studies a near impossibility unless the study utilized a sampling technique to limit the data management problem.

Douglas Aircraft Company began the data analysis for HERAP by investigating and listing all the data sources available during 1967 (HERAP, 2nd Quarterly Progress Report). From this, a number of sources were selected. They were:

1.	Naval Safety Center	Mishap Master File
2.	Naval Safety Center	Individual Flying Time Record
3.	Naval Training Center	Training Records
4.	Bureau of Naval Personnel	Personal Data

Data files on computer tape were obtained from these sources, translated to make them compatible with Douglas' computers and merged into a single data bank after editing.

The data system is used on an IBM 360/65 computer. The statistical programs are basically the University of California at Los Angeles Bio-medical Programs (BMD). Cobol is the programming language for file maintenance and report generator programs, and Fortran IV is the language used for callable selected mathematical subroutines. Details of the operation of the HERAP data bank and statistical programs are fully reported in: Human Error Research and Analysis Program (HERAP II) Final Report. Douglas Aircraft Company, September 1967.

Data gathering, translation, and editing has continued during all the HERAP contracts. However, the major task in development of the data bank has been the selection, translation, modification, and testing of suitable statistical programs. This was completed in 1967. The merged master data bank was completed in 1968 and tested against a dummy data file.

During the current contract, the whole system has been fully utilized. The system was used to study the backgrounds, flying records, training records, and mishap reports for 17,436 Navy pilots active during 1967. The data were obtained from four files: the Officer Master Tape, the Individual Flight Time Records, the Training Record, and the Mishap Master File.

The data bank (Figure 3) was constructed by matching the other files to the records on the 1967 Officer Master Tape by means of pilot file numbers. The Officer Master Tape (OMT) contains demographic data, records of advancement through rank and of past and present duty stations, information on promotional status, and summaries of flying experience.

HERAP DATA BANK

Total Records	17,436
IFTR Records	10,598
Training Records	4,374
Mishap Records	895

Figure 3.

The Individual Flight Time Records (IFTR) report the number of flights and flying hours in 1967 for each pilot, broken down by type of activity and type of aircraft. These were available for only 10,598 of the pilots on the OMT.

The Training Record describes in terms of scores and grades the pilot's progress through the Pilot Training Curriculum from start to finish.

The Mishap Master File (MMF) is a data bank on which is coded all of the data collected in the course of a mishap investigation. The term "mishap" includes comparatively minor accidents occurring on the ground. When the data is keypunched, Navy Safety Center Personnel do not keypunch pilot identification unless the mishap is judged to be at least partially attributable to the pilot.

Therefore, since the HERAP data bank was built up by matching pilot file numbers, it was impossible to use most of the records on the MMF. From approximately 5,000 mishap reports, it was possible to match only 895 "pilot error" mishaps. All mishap rates used in this study were calculated on the basis of these 895 mishaps. Some tabulations were made to further determine the type of mishaps on the HERAP data bank, as shown below, Figure 4.

TABULATION OF MISHAP PARAMETERS

Aircraft Damage	% of Total
Total Loss	30
Substantial	9
Minor	14
Limited	17
None	26
Not Reported	5
	101

Extent of Injury	% of Total
Fatal, Missing	10
Major	5
Minor	12
None	74
	101

Due to Enemy Action	% of Total
Yes	23
No	77
	100

First Accident Type	% of Total
Collision - Aircraft	4
Collision - Ground/Water	10
Collision - Other	38
Stall/Spin	1
Airframe Failure	7
Fire/Explosion	3
Abandoned Aircraft	4
Landing Mishap	22
Other	11
	100

Type Operation	% of Total
FCLP/FMLP	1
CV Operation/LPH	47
Field Operation	36
Sea Operation	1
GCA/ILS	3
GCA	5
Ship Operation	4
Carrier Qualifications	5
	102

Figure 4.

Since the data bank occupied three reels of computer tape and required eight minutes to read, a smaller data bank was developed. It consisted of 2,054 randomly chosen records. The records on this tape were shortened by one half by dropping fields which were usually blank or otherwise unusable. Most of the statistical analysis was done with this data bank.

The Exposure Index and a Measure of Pilot Safety

It was necessary to move from vague intuitive concepts of what constitutes a "safe" or an "unsafe" pilot to a numerical definition of pilot safety. To be useful in the search for relationships with the wide range of data bank variables, the measure of pilot safety had to vary continuously from "very safe" through "average" to "very unsafe". Assuming that a "safe" pilot is one who doesn't have an excessive number of mishaps for the amount and circumstances of the flying he does, two different techniques of rating pilots were developed. Either of these rating techniques is just a mathematical statement of an intuitive definition of pilot safety.

Before discussing the pilot rating techniques, it is necessary to consider the interaction between pilot safety and the immediate environment of the pilot during his flight. Parameters such as weather, phase of flight, type of aircraft, day or night, etc., will obviously influence the probability that a pilot will suffer a mishap. The purpose of the human error research program is to isolate safety-related factors intrinsic to the pilot; factors which will cause some pilots to consistently have more mishaps than other pilots regardless of the type of flying they do. With the goal of making the pilot safety rating free of the effects of the environmental variables, an exposure index was developed. The exposure index was originally conceived as a means of ranking squadrons in terms of safety, but it became clear during HERAP II that the study of independent pilot safety would require some adjustment for exposure.

The basic concept behind the exposure index is that pilots flying under the same conditions should be expected to have mishaps at the average rate for all pilots. The rating systems compare each pilot's actual performance to this expected performance. In order to calculate expected performance as accurately as possible, it is necessary to find the mean mishap rates within each possible set of flying conditions. To find these rates requires accurate data on the number of mishaps and the number of flights under each set of conditions. This data is certainly not available, since no two flights are under the exact same conditions, so lumping into categories is required.

Data was available to calculate the mishap rates only when broken down by aircraft type, type of operation, and day or night. The Mishap Master File was also supposed to contain complete information on several other environmental factors for which allowance might have been made, but these data fields as shown below were found to be over fifty percent blanks and could not be used. There is no data on the number of flights under the various conditions described by these fields, so that at best the data would have had to come from some inaccurate guesses. The exposure index could be improved if the missing data were available.

Sample of data fields from the Mishap Master File which contained over fifty percent blanks:

Data	Percent Blank
Kind of flight	85
Clearance	91
Wind velocity	95
Altitude	70
Maneuver prior to emergency	97

The final exposure index divided the types of flying into only twelve categories broken down on the basis of three variables (Figure 5). The aircraft types were grouped into three categories of hazardousness on the basis of published mishap rates[1]. Type of operation was divided into "carrier activity" and "non-carrier activity". These combinations were further broken down into "day" and "night" flying. The overall mishap rate was calculated for each of these categories, and the pilots were rated on the basis of their performance relative to these standards.

FLYING CATEGORIES AND MISHAP PROBABILITIES
(PER 10^4 LANDINGS)

		Day Non-Carrier Flights	Night Non-Carrier Flights	Day Carrier Flights	Night Carrier Flights	All Types Flights
Aircraft Risk Category	Low Risk	1.590	1.63	6.65	5.39	1.80
	Medium Risk	3.87	5.42	33.7	46.6	9.05
	High Risk	7.96	8.50	65.8	98.5	18.8
	All Aircraft Types	2.99	3.02	31.7	41.6	6.04

Figure 5.

1. U. S. Navy Aircraft Accident Statistics, Fiscal Year 1967 (Confidential)

RANKING OF PILOTS BY SAFETY RECORDS

The first ranking method was to obtain a quantitative answer to the question, "What is the probability that, by chance alone, a pilot should have had more mishaps than he actually had?" Within each category of flying, the value of the cumulative binomial distribution was calculated, where

$$CB(x) = \sum_{x+1}^{\infty} B(x) = \sum_{x+1}^{\infty} \frac{N!}{x!(N-x)!} \rho^{x} q^{(n-x)}$$

given x = number of mishaps the pilot had within the category

N = number of landings the pilot had within the category

ρ = average mishap probability for the category (from Figure 3)

$q = 1-\rho$

A value of CB(x) was calculated for each category in which the pilot flew, and these were then averaged to give his "binomial ranking". Figure 6 shows the frequency distribution of the binomial rankings. Most binomial rankings are nearly zero, meaning that for most pilots, the probability was very low that they should have had more than the number of mishaps which they actually had. This is to be expected; most pilots had no mishaps but didn't fly enough to really be expected to have a mishap. The highest binomial ranking value was only .28.

The second approach to ranking pilots was based on the question, "How does a pilot's mishap rate (mishaps per 10^4 landings) compare to the mishap rates of all other pilots. An arbitrary formula was devised which has the properties:

(1) Increases in value with an increasing number of expected mishaps,

(2) Decreases in value with an increasing number of mishaps,

(3) Equals zero if a pilot's actual number of mishaps is exactly equal to his expected number of mishaps.

"Expected number of mishaps" is defined as the sum over all flying categories of the pilot's actual number of landings multiplied by the average mishap rate from Figure 6, i.e.:

$$\text{Expected Mishaps} = \sum_{L=1}^{12} L_i \times \rho_i$$

where, for each pilot,

L_i = number of landings in flying category i

ρ_i = mishap probability for flying category i

The arbitrary function comparing expected to actual mishaps, called the "x-ranking", was calculated for each pilot (Figure 6). Each pilot's percentile ranking on the distribution of x-rankings was then calculated (Figure 7). This percentile ranking was used as the variable indicating how a pilot's safety record compared to those of other pilots.

These two ranking procedures were applied to the condensed data bank of 2,054 randomly chosen pilots. There were sufficient data to rank only 1,133 of these pilots.

Regression Analysis

An extensive analysis was conducted to determine if the variables in the condensed data bank could be used to predict the binomial rankings or the percentile rankings. A stepwise regression program was used for most of this work. This program builds up a linear regression equation from a set of candidate variables by successively inserting the variable with the highest probability of significantly improving the fit of the equation. Variables from the Officer Master Tape and the Training Record were used as candidate variables. Predictive equations were developed for the binomial ranking and the percentile ranking (Figure 8). Neither of these equations is really satisfactory. The equation predicting binomial ranking has a multiple correlation coefficient of only .58; the other equation has an even lower correlation coefficient.

In the equation predicting the binomial rankings the last year of combat code has a correlation coefficient with the binomial rankings of .52. This one variable accounts for almost all of the predictive ability of the equation. Similarly, date of birth and percentile ranking have a correlation coefficient of .37, accounting for most of the predictive ability in the second equation.

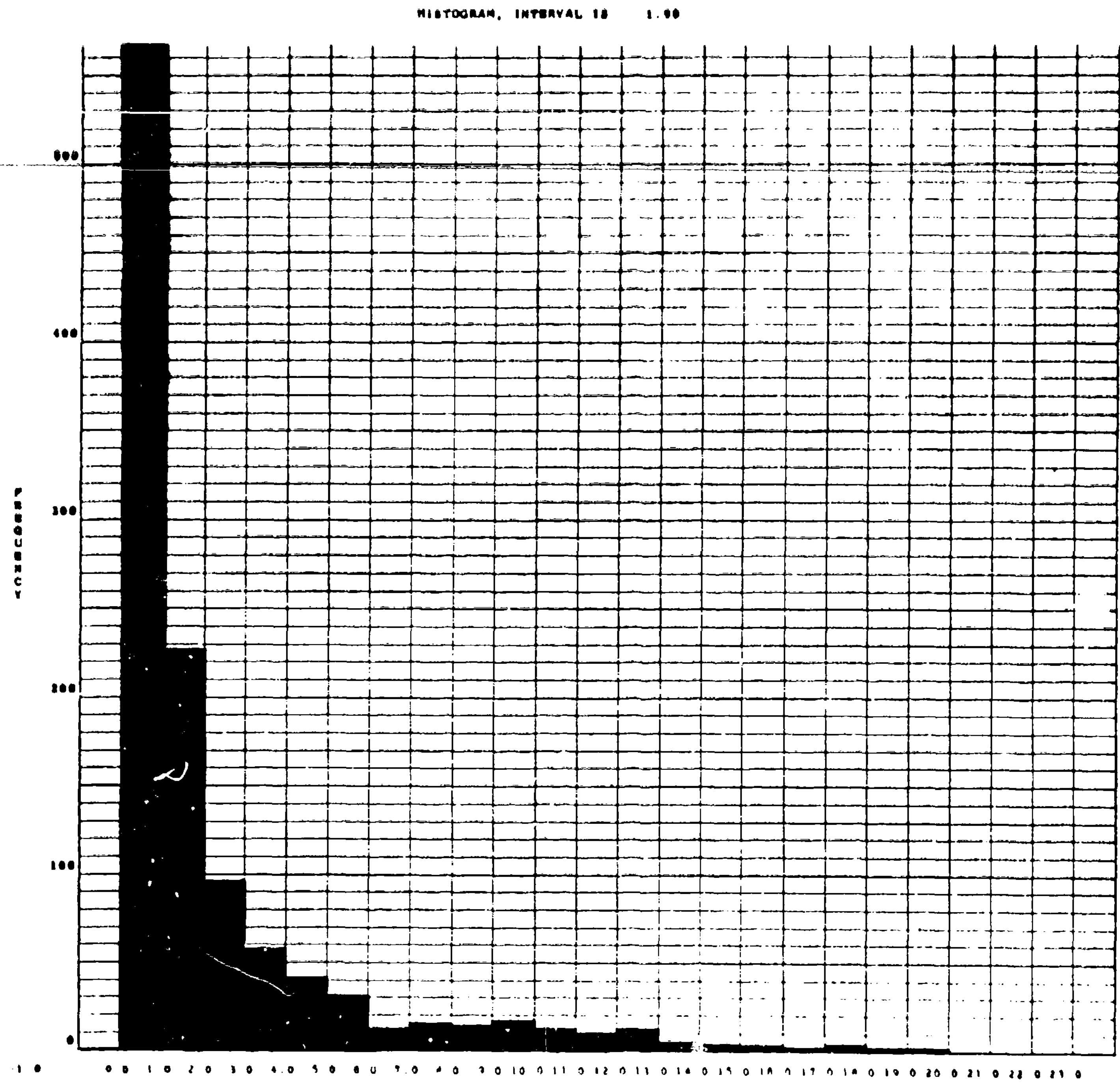
In both equations, a group of variables appear which relate to age or experience. This group includes birth year, rank, total flight hours, date of birth, and flight hours for the past five years. In each equation, the overall relationship is a slight decrease in a pilot's ranking for an increase in his age or experience.

The variable "last year of combat" was set to zero for pilots with no combat experience, so that it owes some of its predictive ability to being able to distinguish between pilots who had or had not been assigned to combat. Its correlation with the rankings both with and without these zero values (Figure 9) was calculated in an attempt to gauge how important the zeros are. No significant change was noted.

Comparison of Mishap Distribution to a Poisson Distribution

The distribution of mishaps among the population of pilots was tested for randomness. It is often stipulated[1] that the distribution of frequencies

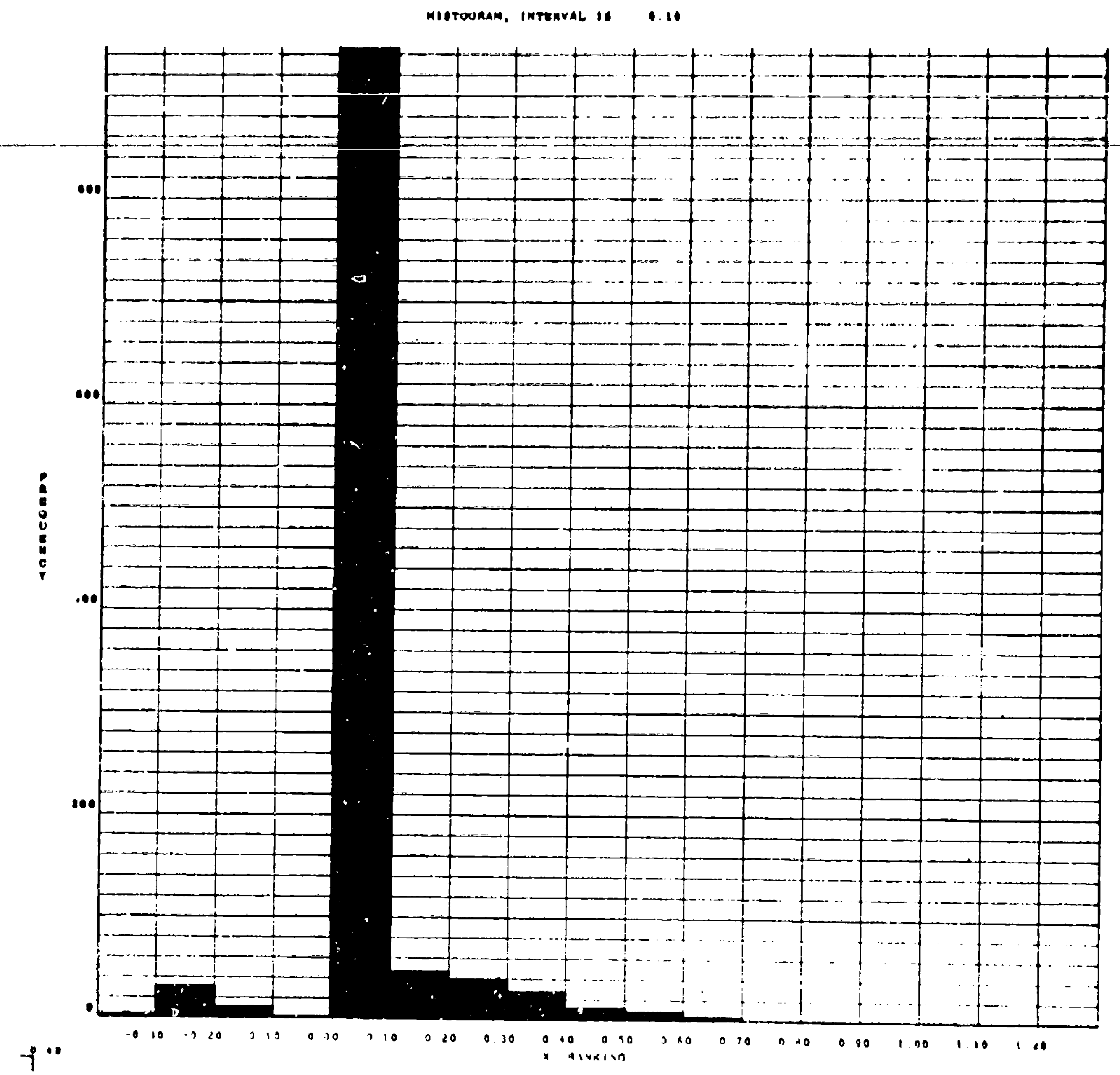
1. Stinson, P. J. and Walsh, J. E. (1965) "An Application of the Poisson Approximation to Naval Aviation Accident Data", presented at the 35th Session of the International Statistical Institute, Belgrad.



BINOMIAL RANKING

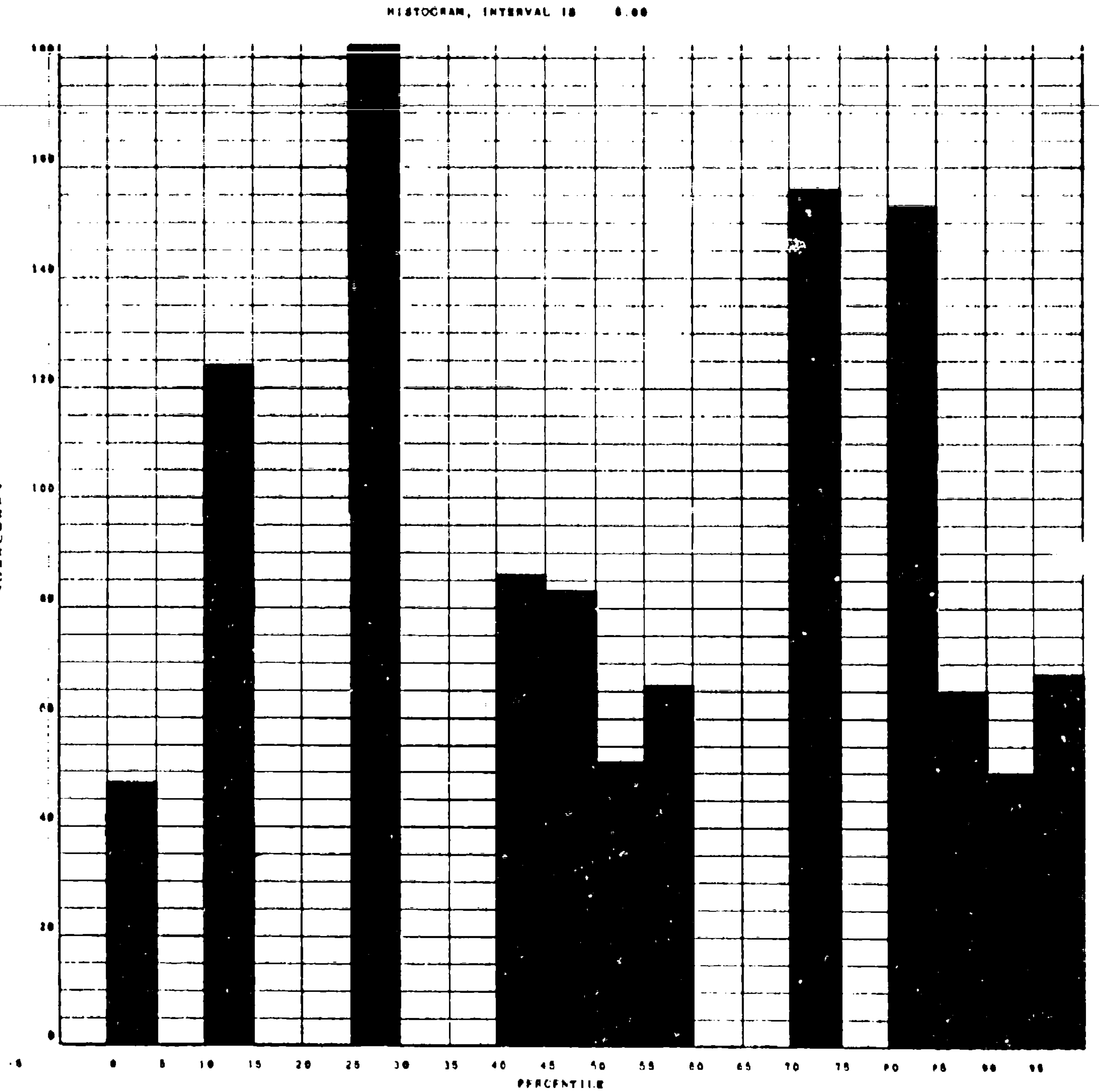
Random Sample of Pilots

Figure 6.



Random Sample of Pilots

Figure 7.



Random Sample of Pilots

Figure

[illegible]

PREDICTIVE EQUATIONS

1. Binomial Ranking = -4.703
 + .105 x (birth year, last two digits)
 + .005 x (VTA grade)
 - .006 x (VT3 grade)
 + .332 x (rank code)
 - .261 x (months of combat)
 +1.027 x (last year combat, last digits)
 + .079 x (college level code)
 - .166 x (highest award code)
 - .0001 x(total flight hours)

$R = .58,\ R^2 = .34$

2. Percentile Ranking = -1.808
 + .00005 x (date of birth, days since 1900)
 - .024 x (months of combat)
 + .044 x (last year of combat, last digits)
 + .008 x (college level code)
 + .00004 x (flight hours last five years)

$R = .42,\ R^2 = .17$

FIGURE 9

CORRELATION COEFFICIENTS FOR
"LAST-YEAR- OF-COMBAT"

	WITH ZEROES (1133 cases)	WITHOUT ZEROES (156 cases)
Correlation with Binomial Rankings	.52	.47
Correlation with Percentile Rankings	.20	.27

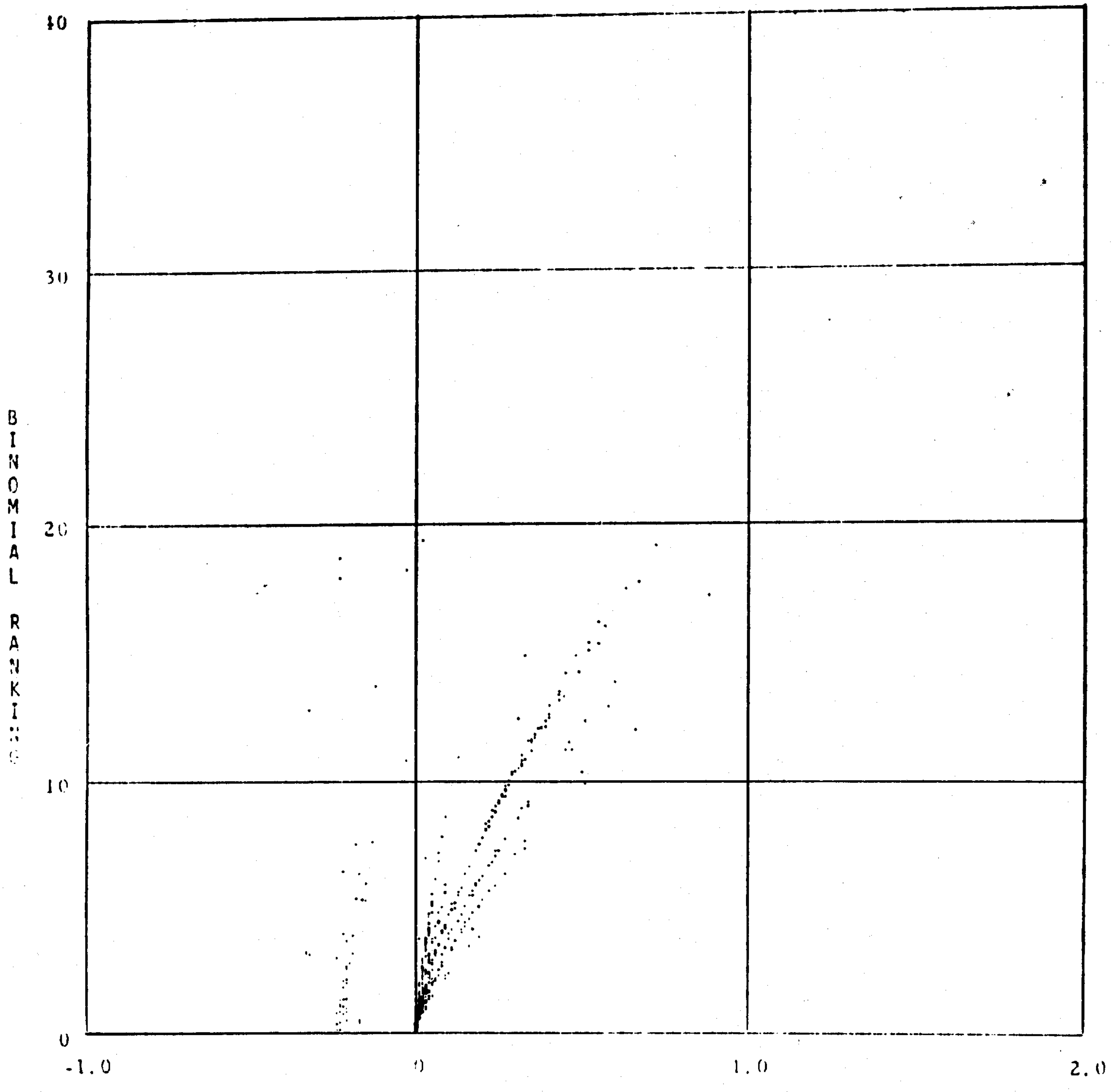
FIGURE 10.

of randomly occurring accidents is approximated by the Poisson distribution. The observed distribution of numbers of mishaps for the total pilot population was used to first calculate the mean number of mishaps for all pilots and was then compared to the Poisson distribution having this same mean. The theoretical and observed frequencies for each number of mishaps are presented in the following table.

No. of Mishaps	Observed Frequency	Expected Frequency
0	16541	16456
1	703	850
2	67	22
3	14	0.4
4	2	0
5	0	0
6	0	0
7	0	0
8	1	0

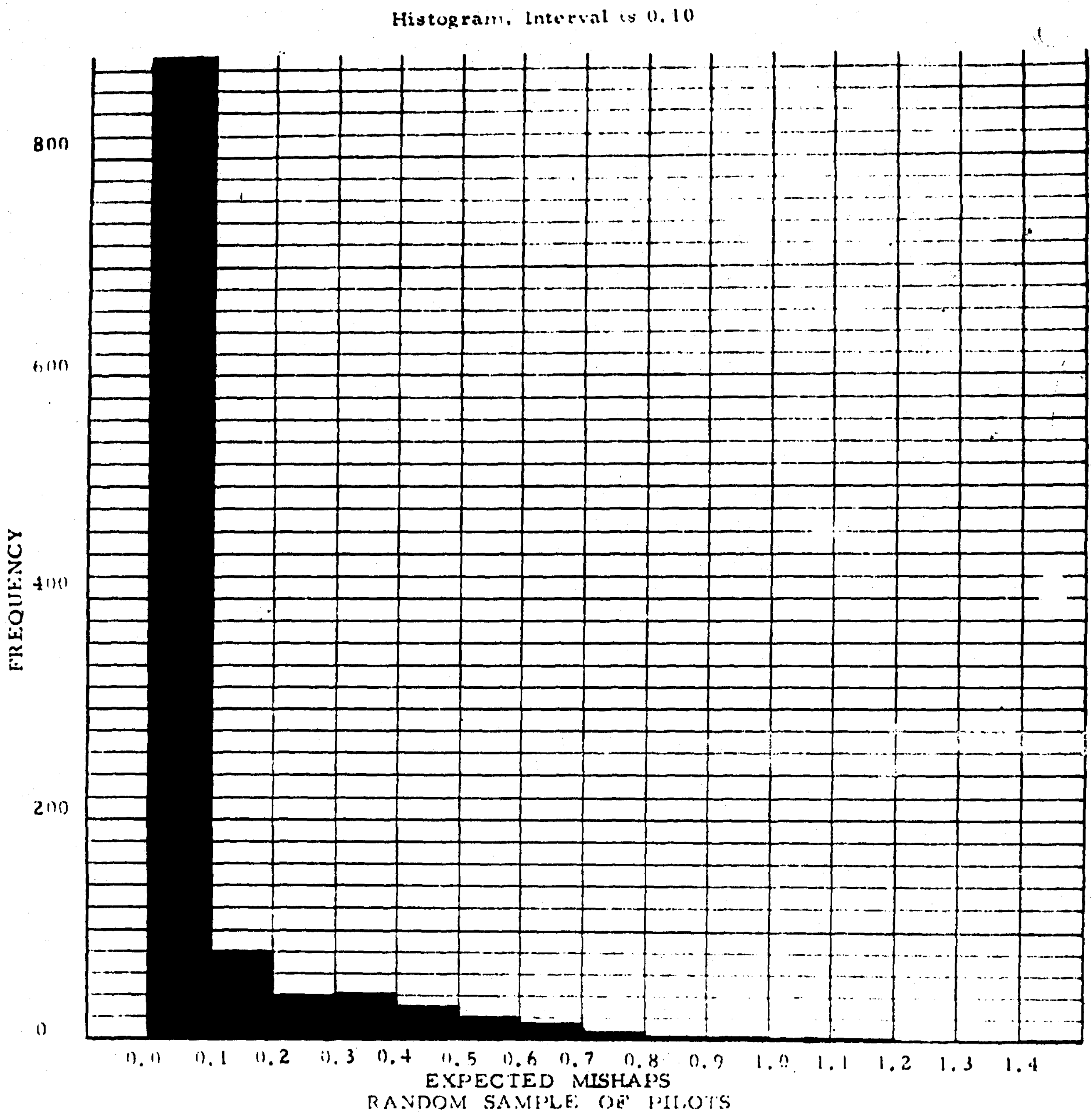
It is obvious that the fit is not a good one. The chi-squared test of the discrepancies cannot be accounted for by sampling error.

Histograms giving results of analyses of combat experience, recency of combat flying, flight hours, college level, birth date, military rank and awards are shown by Figures 14 through 20.



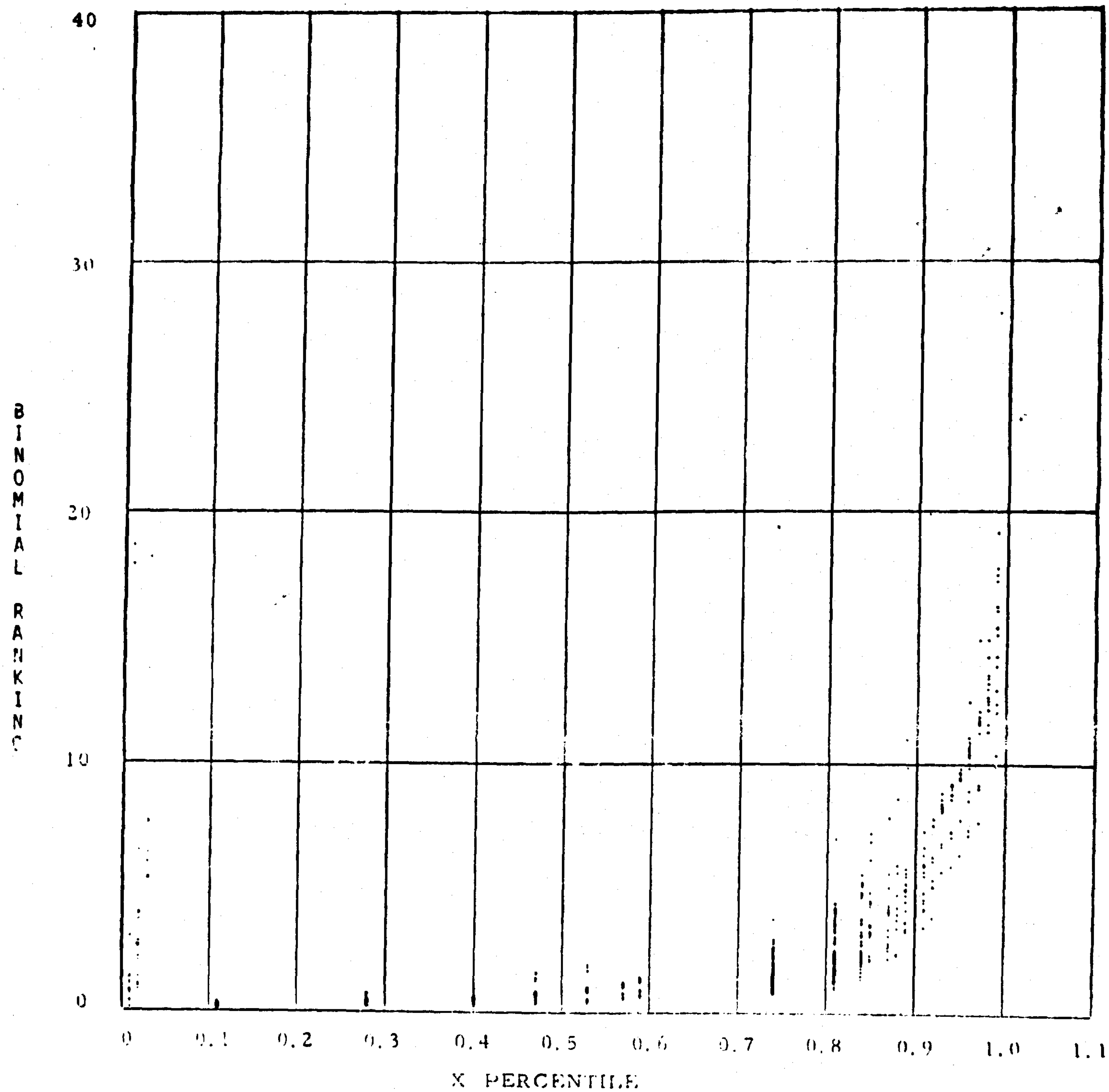
RANDOM SAMPLE OF PILOTS

FIGURE 11.



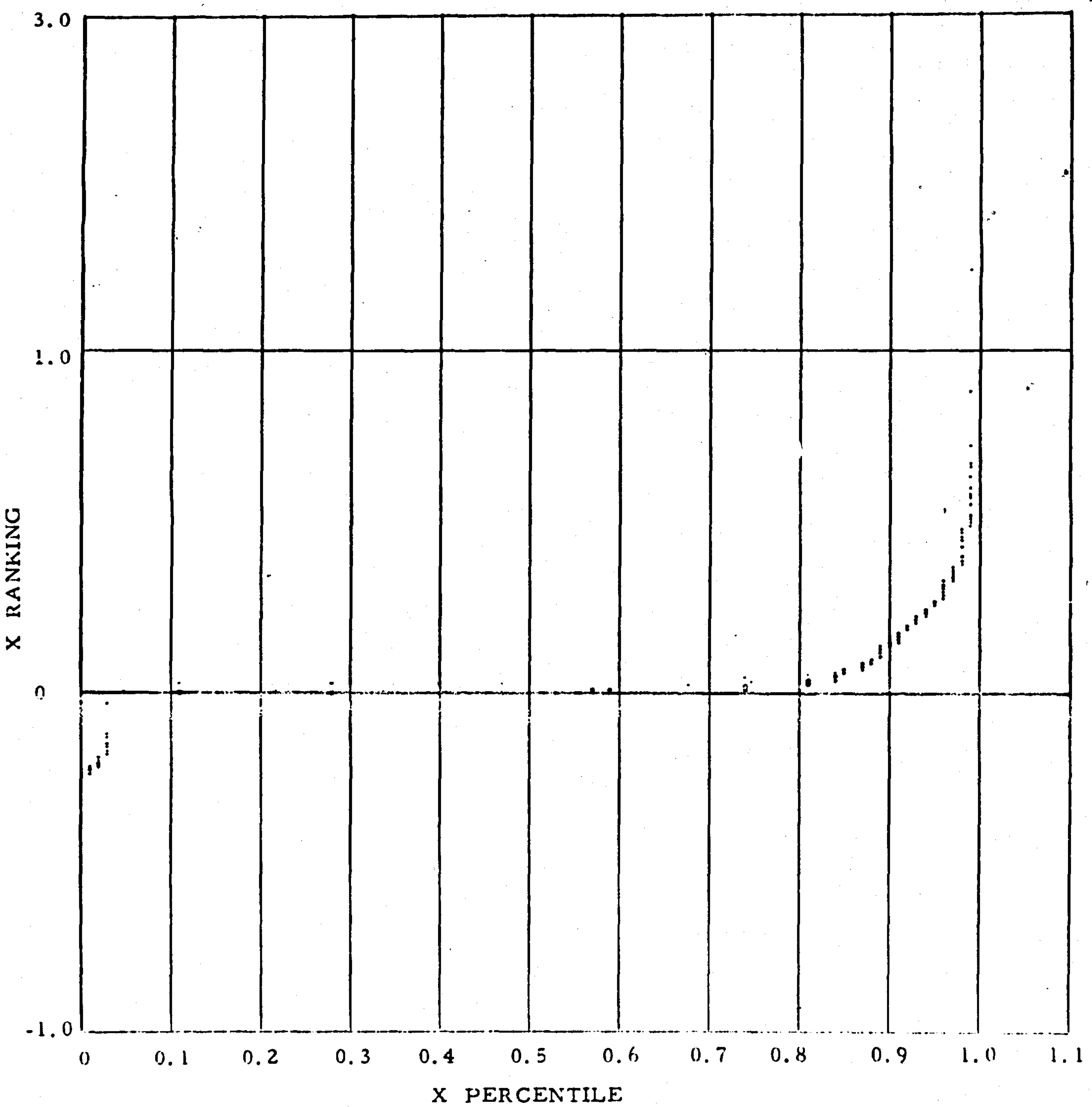
The distribution of expected mishaps, calculated by the method described in the text.

FIGURE 12.



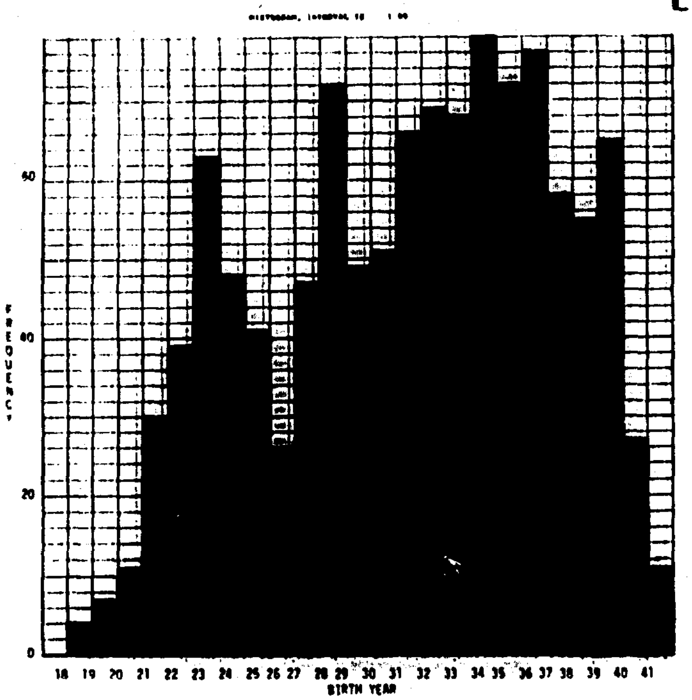
RANDOM SAMPLE OF PILOTS

FIGURE 13.

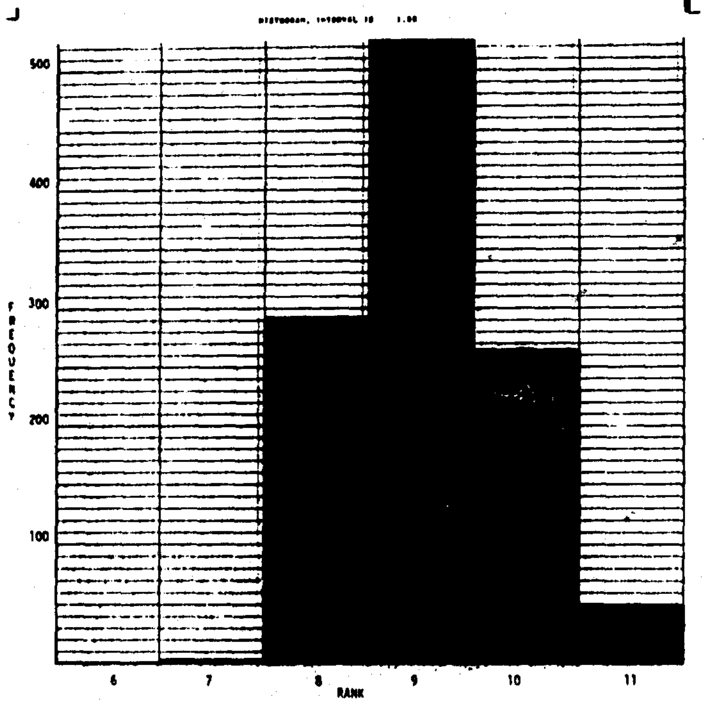


RANDOM SAMPLE OF PILOTS

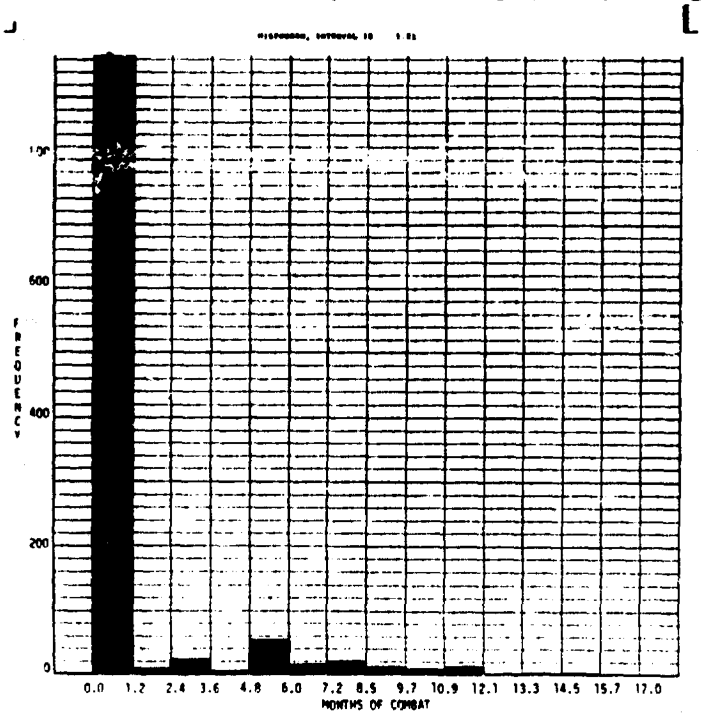
FIGURE 14.

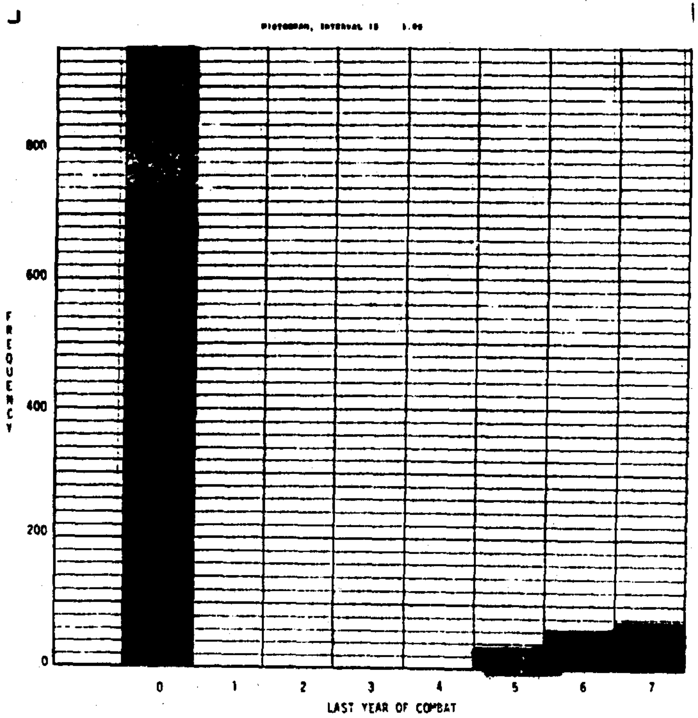


The Distribution of Birth Years, Measured From 1900.



Code for Officer's Present Work Ranges from 7 (LTJG) to 11 (CAPT).

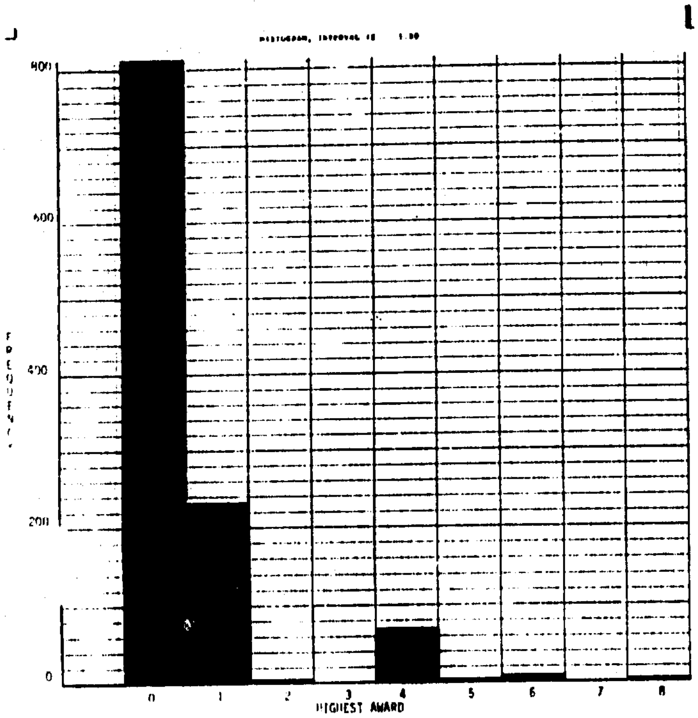




Distribution of the pilot's most recent combat experience, measured in years since 1960. A value of 0.0 indicates no combat experience.

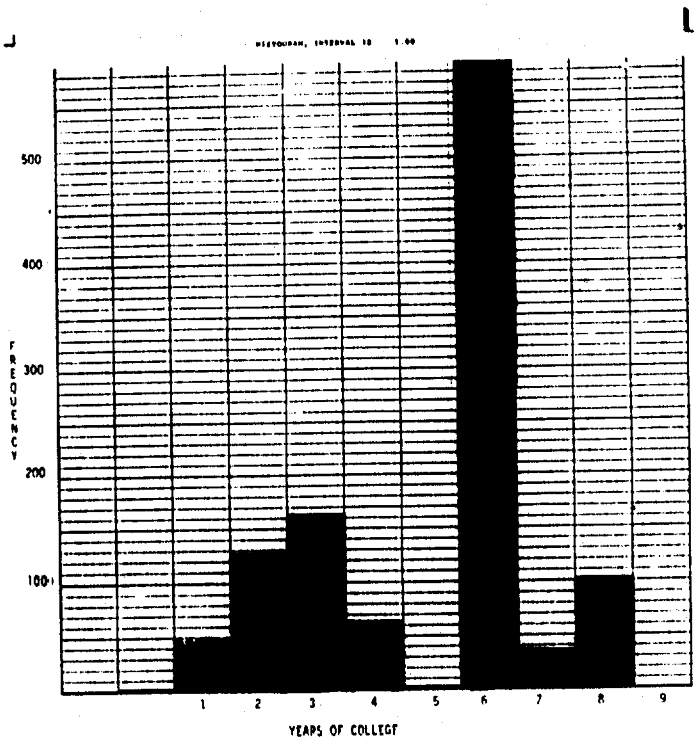
DATA DISTRIBUTIONS FOR ALL PILOTS

FIGURE 15

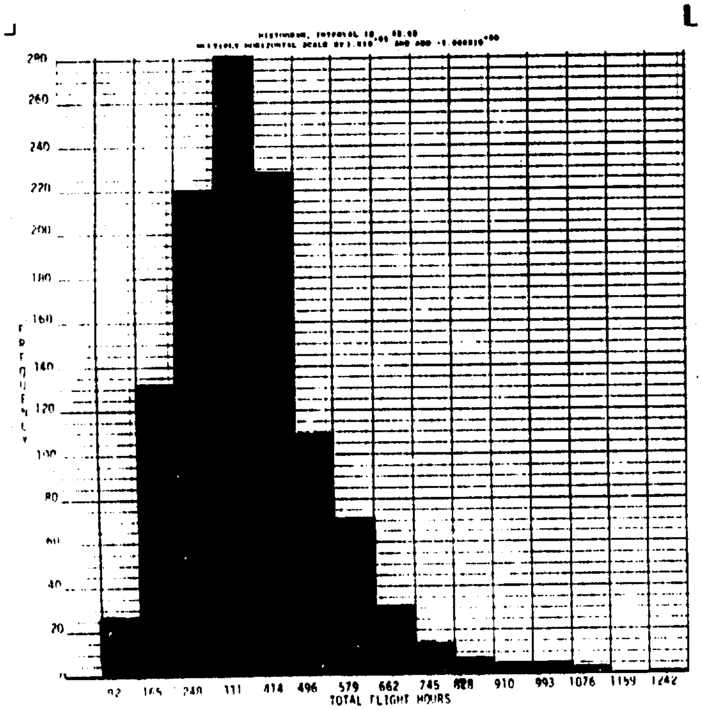


Random Sample of Pilots

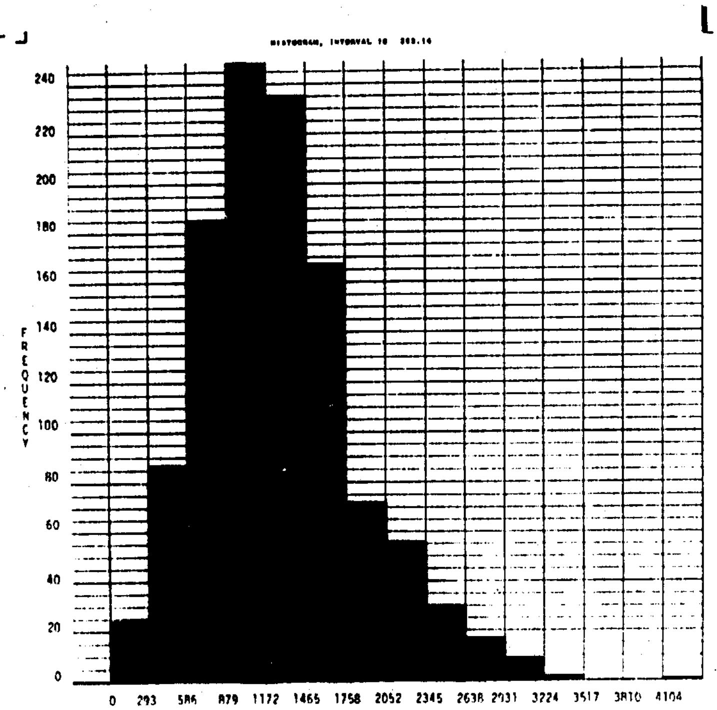
The code for the highest military decoration received, ranging from 0 for no award to 8 for the Bronze Star.



Random Sample of Pilots



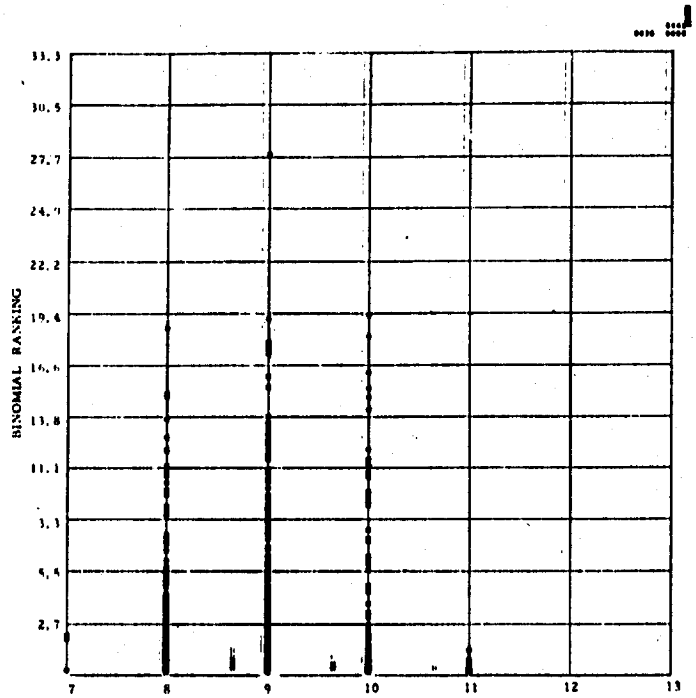
Random Sample of Pilots



Hours of Flying in the Previous 5 years.
Random Sample of Pilots

DATA DISTRIBUTIONS FOR ALL PILOTS

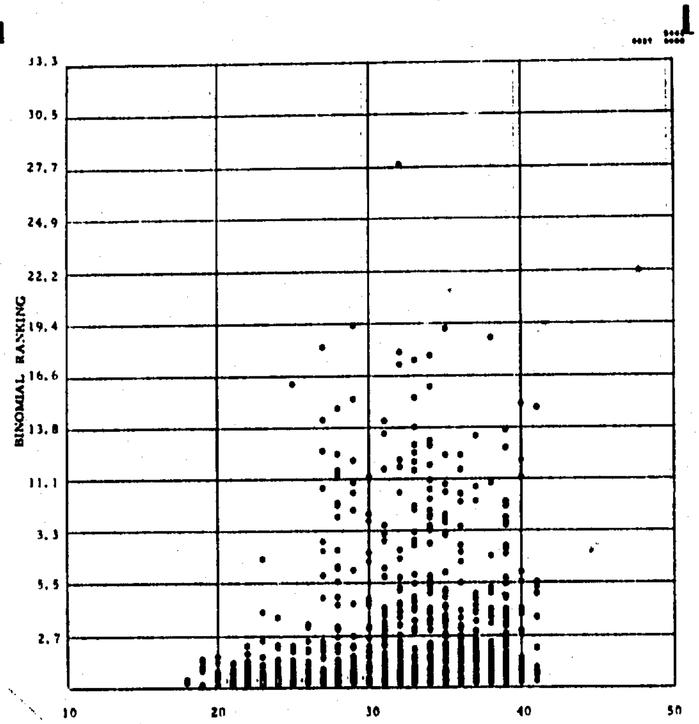
FIGURE 16



RANK

Random Sample of Pilots

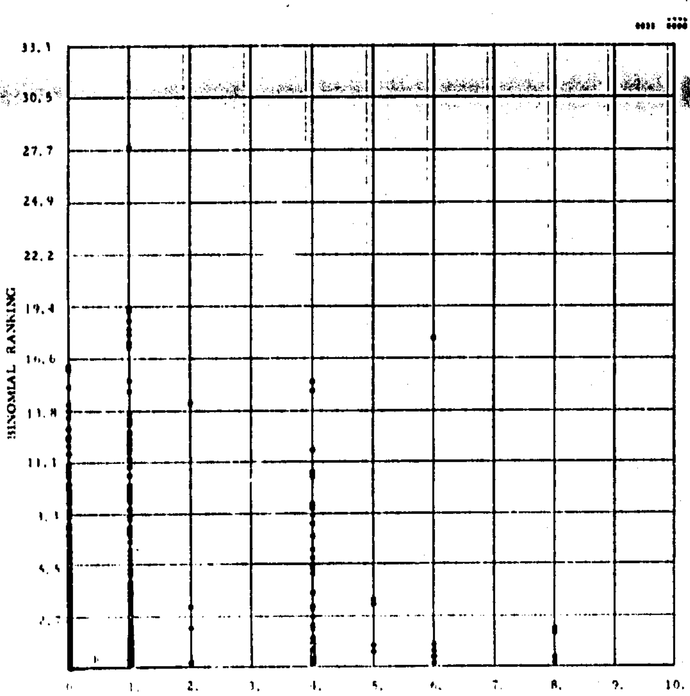
Code for officer's present rank ranges from 7 (LTJG) to 11 (CAPT).



Birth Year

Random Sample of Pilots

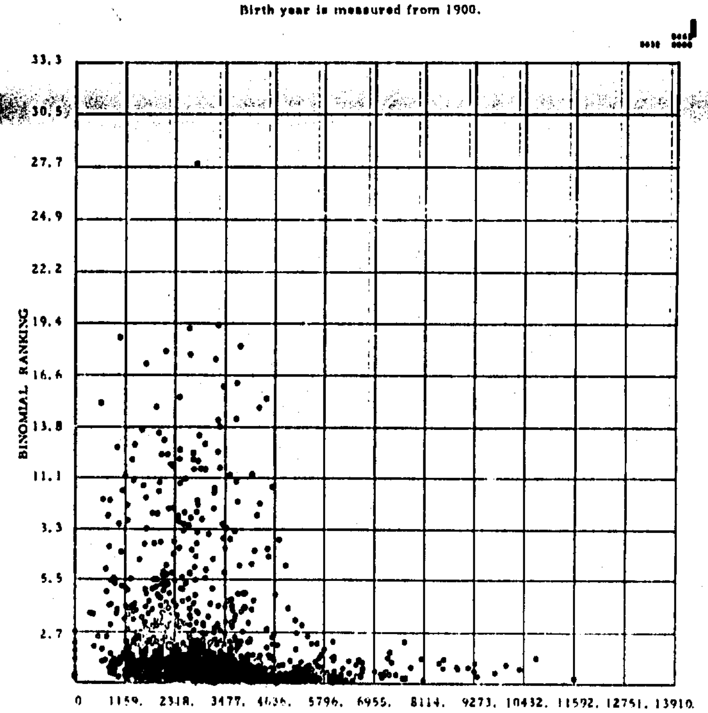
Birth year is measured from 1900.



HIGHEST AWARD

RANDOM SAMPLE OF PILOTS

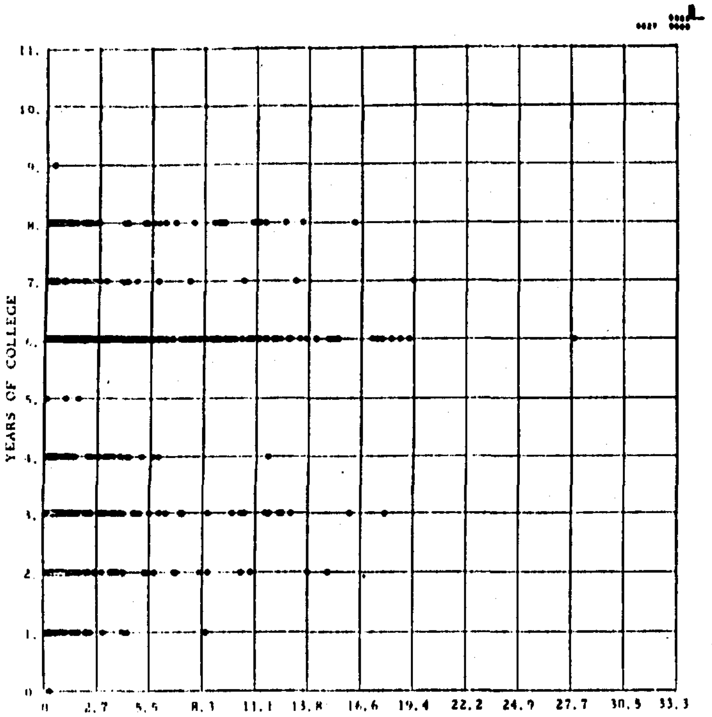
0 = no award, 1 = Bronze Star



TOTAL FLIGHT HOURS

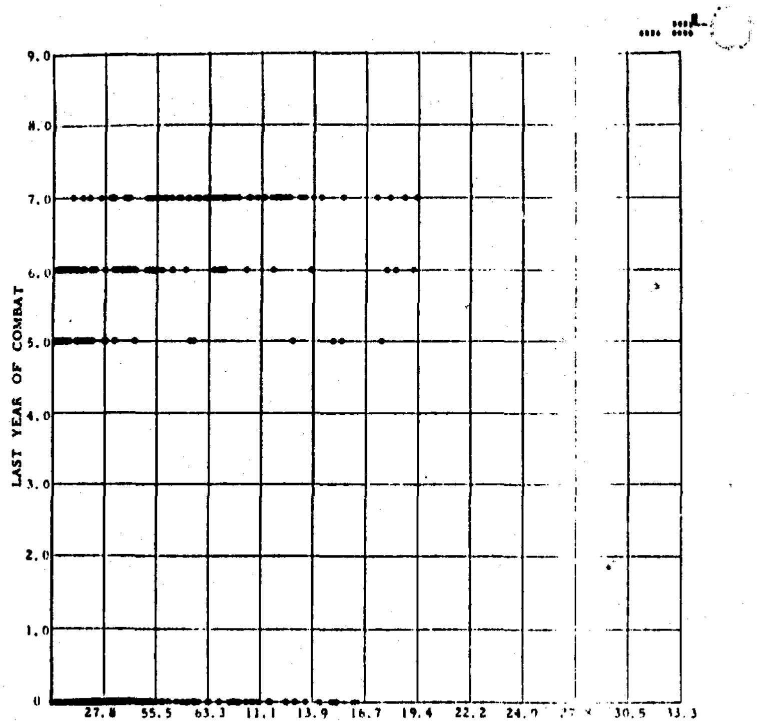
RANDOM SAMPLE OF PILOTS

FIGURE 17



BINOMIAL RANKING

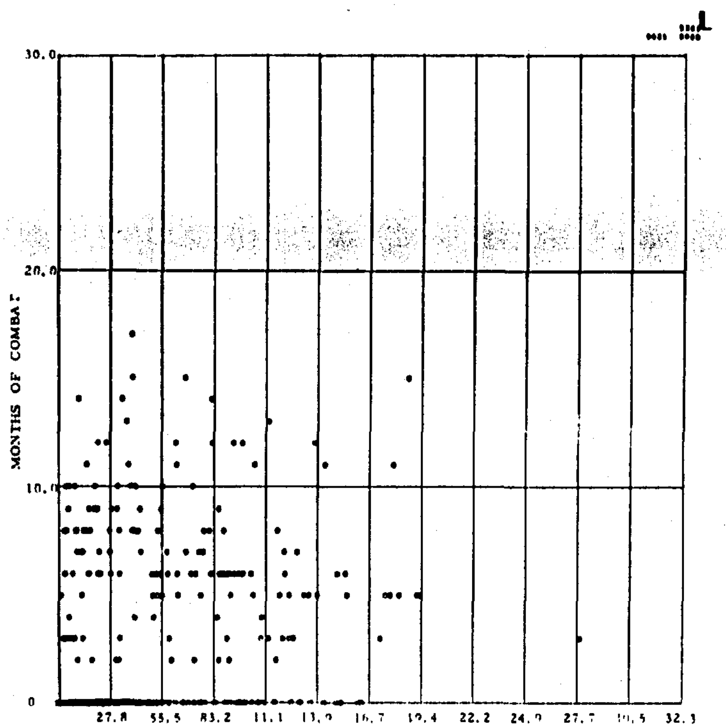
RANDOM SAMPLE OF PILOTS



BINOMIAL RANKING

RANDOM SAMPLE OF PILOTS

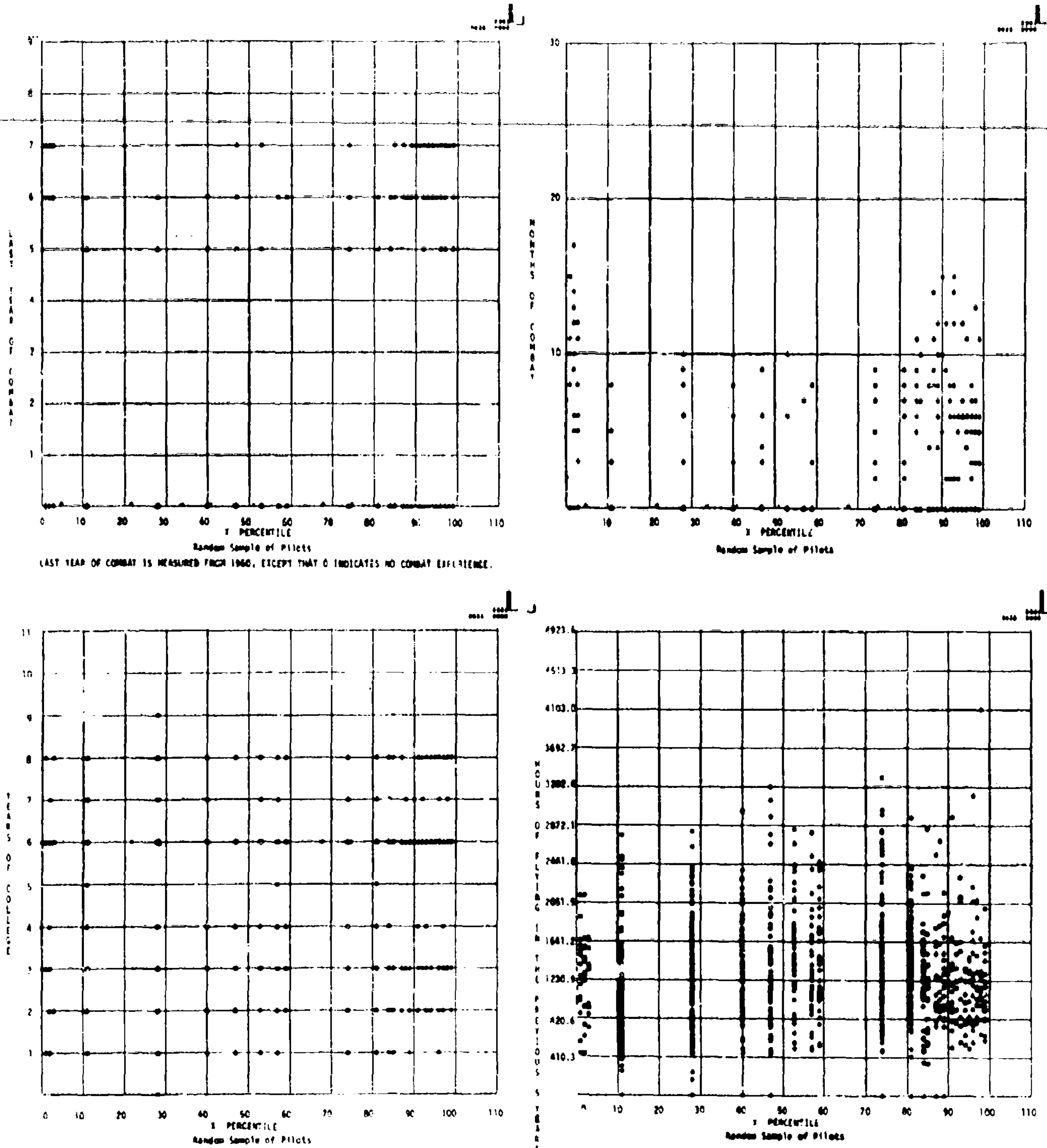
Last Year of Combat is measured from 1960, except that 0 indicates no combat experience.



BINOMIAL RANKING

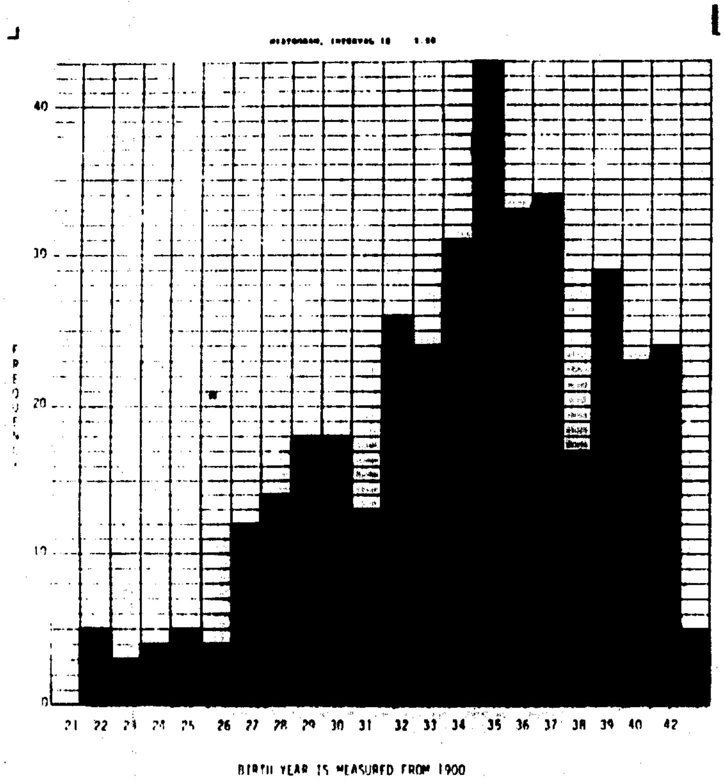
RANDOM SAMPLE OF PILOTS

FIGURE 18

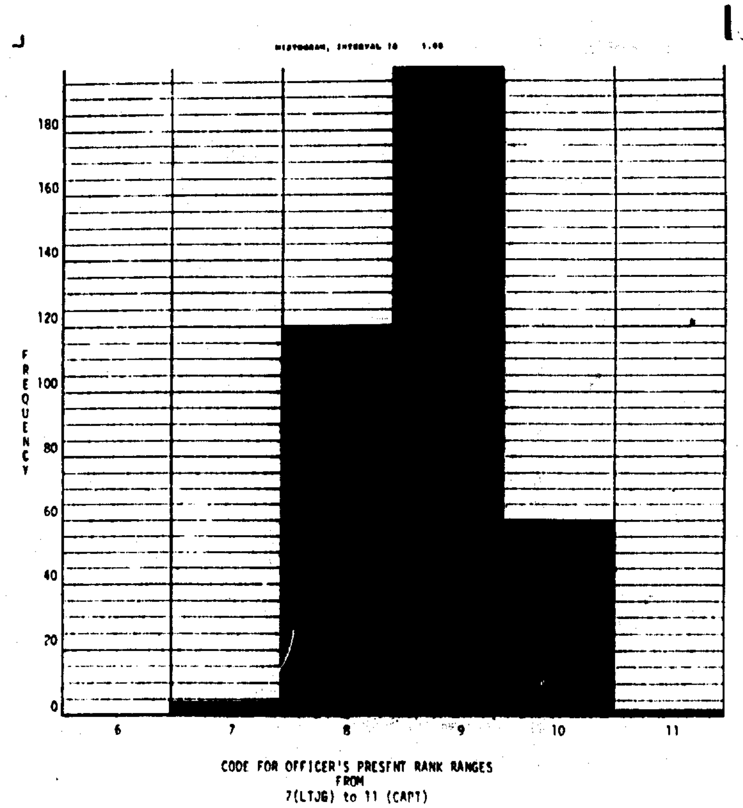


LAST YEAR OF COMBAT IS MEASURED FROM 1960, EXCEPT THAT 0 INDICATES NO COMBAT EXPERIENCE.

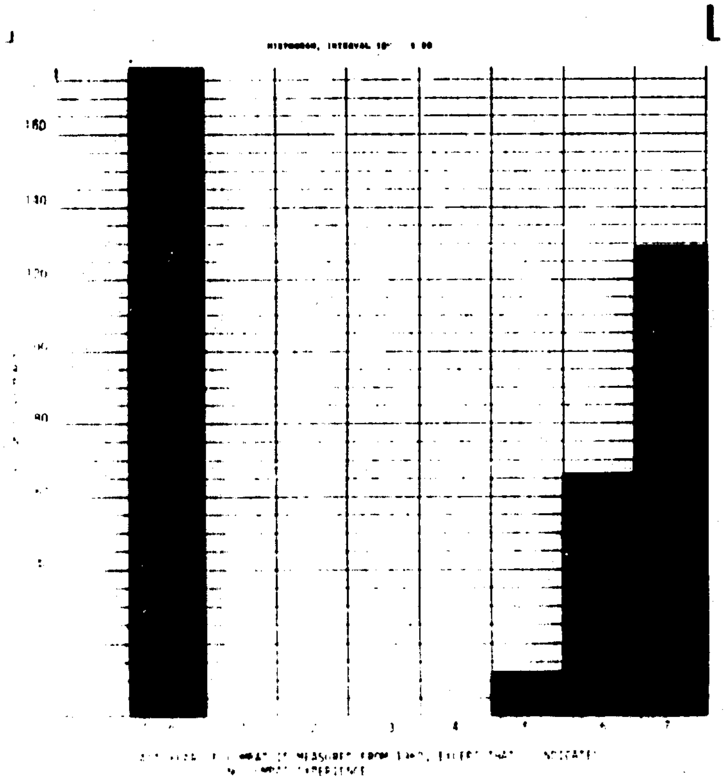
FIGURE 19

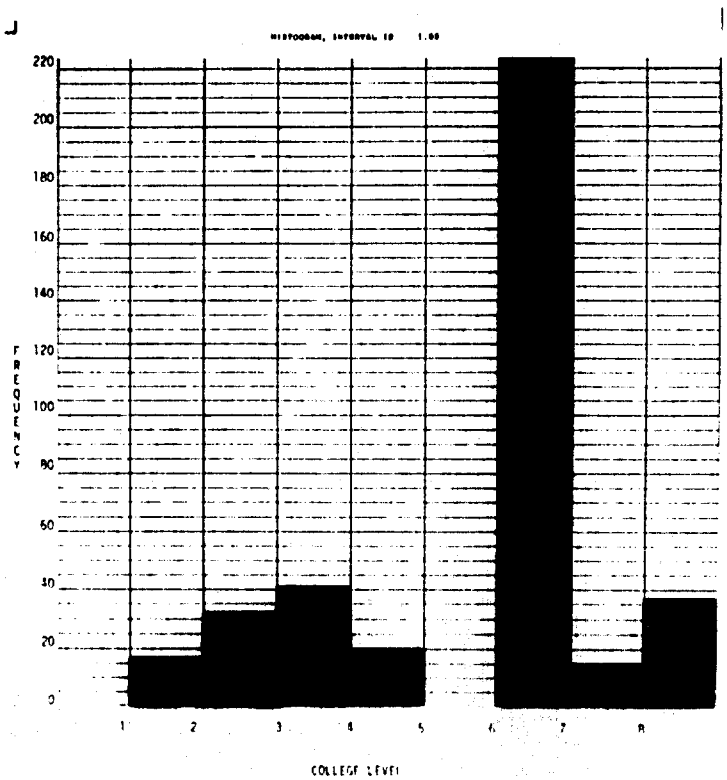


FREQUENCY
BIRTH YEAR IS MEASURED FROM 1900



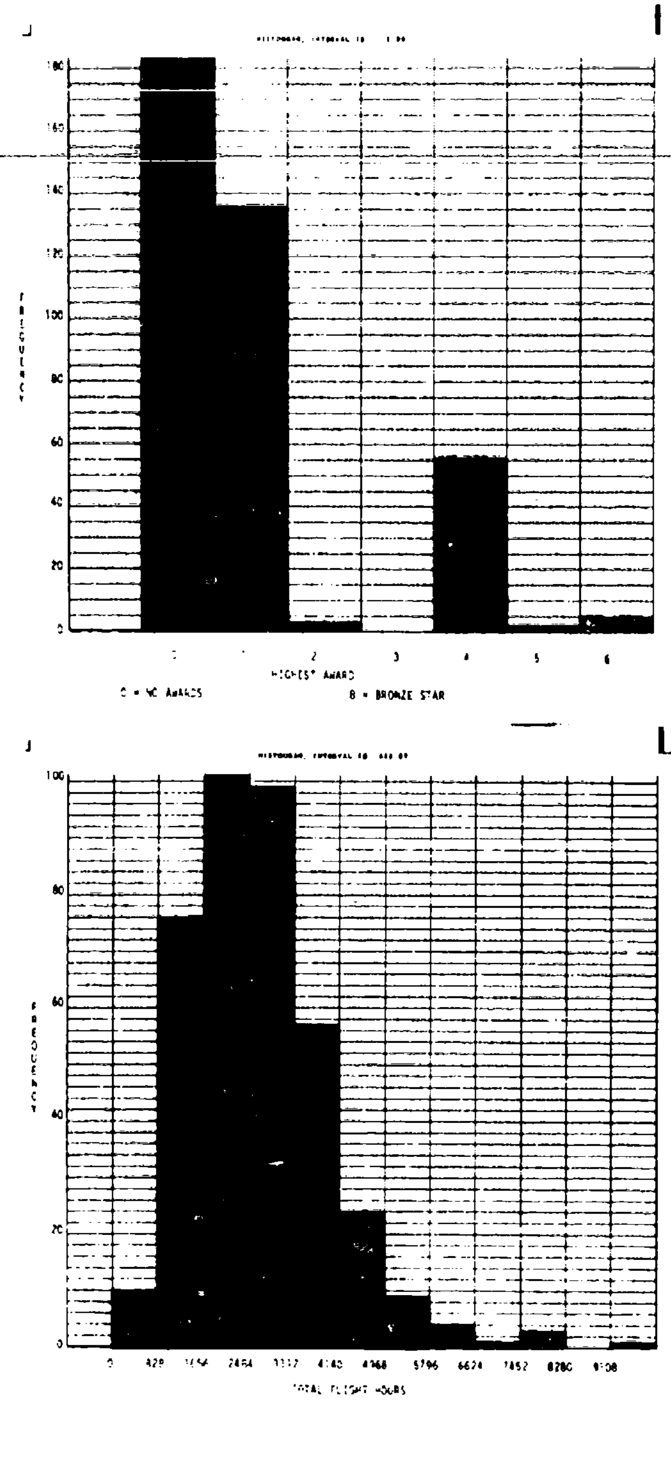
FREQUENCY
CODE FOR OFFICER'S PRESENT RANK RANGES FROM 7(LTJG) to 11 (CAPT)





FREQUENCY
COLLEGE LEVEL

[illegible]



... FOR PILOTS IN THE LOWEST QUARTILE OF THE RANKING

FIGURE 2

DISCUSSION AND CONCLUSIONS

The regression analysis demonstrated that the pilot factors described on the data bank have some influence on pilot safety. One way of interpreting the multiple correlation coefficient is to consider its square (R^2) as the proportion of the variance, or scatter, in the pilot safety rankings which is explained by the variables in the regression equation. Thus, thirty four percent of the variance in the binomial rankings can be explained by the influence of such variables as age, flying experience, and combat experience. Sixty-six percent of the variance remains unexplained.

There are several ways in which the multiple correlation coefficient obtained in this analysis might be improved. In the interests of testing a large amount of data in a systematic fashion, the data was used pretty much as it came from the Navy. In several cases we attempted to impose a linear relationship on variables which had aribtrary numbers assigned to their different levels. It would be desirable to either reorder these levels within the variable or to treat each possible level as a binary variable indicating that a pilot either does or does not have a particular characteristic. Plots of the residuals against each variable were inspected for evidence of higher order dependencies of the criterion on the predictor variables; no such evidence was found.

Another way to improve the predicition equations would be to further improve the exposure index. The fact that the safety rankings showed some dependence on intuitively meaningful variables such as pilot experience is evidence that the exposure index has successfully removed some of the effects of the different types of flying environment, but doubtlessly some of the unexplained variance is still due to this effect. Another factor contributing to the lack of a high multiple correlation is the spotty, incomplete nature of the data base. Perhaps an effort to collect fewer types of information in a more complete manner might be productive.

Figure 9 shows the final regression equation to predict the binomial rankings. This equation includes every variable with at least a ninety percent probability of actually contributing to the "fit" of the equation. Figure 22 is a summary of the increase in the multiple correlation coefficient as the computer program moved through the various steps of curve fitting. The first two or three variables entered account for most of the predictive power of the equation.

It should be remembered that an unknown portion of the correlation in this equation is due to characteristics of the sample population which are not true characteristics of the total population. If the equation were to be applied to some other sample population, it would not account for as much of the variance as it does for this population. It is difficult to estimate how much of this "shrinkage" will occur. Future investigations might well apply this equation to some other population in a cross validation to measure the actual shrinkage of the predictive power in at least one case.

Despite these criticisms, the multiple correlation found for the binomial ranking equation is in the same range as those used with success in areas such as employee selection. This equation could be used to predict the binomial ranking scores of individual pilots. The predicted binomial ranking scores reflect the probability that a pilot will have mishaps at less than the expected rate. The equation could be applied to a group of pilots to differentiate between those who were expected to perform well (high scores) and those who were not (low scores).

SUMMARY TABLE FROM STEPWISE REGRESSION

The dependent variable is Binomial Ranking

STEP NUMBER	VARIABLE ENTERED	MULTIPLE R	R-SQUARED	INCREASE IN R-SQUARED	F VALUE TO ENTER
1	Last yr. combat	.52	.27	.2748	428.7
2	Birth year	.55	.30	.0252	40.7
3	Months combat	.56	.31	.0128	21.1
4	VT3	.57	.32	.0085	14.1
5	Level college	.57	.32	.0034	5.6
6	Current source	.57	.33	.0024	4.0
7	Total flight hours	.57	.33	.0021	3.6
8	Highest award	.58	.33	.0022	3.7
9	Rank	.58	.33	.0019	3.2
10	VT4	.58	.34	.0018	3.1

Figure 22

BIBLIOGRAPHY

PARTS I AND II

Anonymous. Aircraft Accident Statistics, Fiscal Year 1966. U. S. Naval Safety Center, Norfolk, VA. (Confidential)

Anonymous. Human Error Research and Analysis Program (HERAP) 1st Quarterly Progress Report, 1 February 1967. Douglas Aircraft Company

Anonymous. Human Error Research and Analysis Program (HERAP) 2nd Quarterly Progress Report, 1 May 1967. Douglas Aircraft Company

Anonymous. Human Error Research and Analysis Program (HERAP) 3rd Quarterly Progress Report, 1 August 1967. Douglas Aircraft Company

Anonymous. U. S. Navy Aircraft Accident Statistics, Fiscal Year 1967. (Confidential).

Altman, J. W., A central store of human performance data. Proceedings, symposium on qualification of human performance, pp. 97 - 108. Albuquerque, New Mexico, August 17 - 19, 1964.

Armstrong,H. G., Principles and Practice of Aviation Medicine. The Williams and Wilkins Company, Baltimore, 1952.

Beldam, E. M., A Study to Determine an Optimum Altimeter Presentation. Royal Canadian Air Force, Report IAM, 59/2, May 1959.

Carpenter, D. L., L. R. Creamer, R. F. Gabriel, A. A. Burrows. Human Error Research and Analysis Program (HERAP III), Final Report, Part III, Douglas Aircraft Company. March 1969.

Cooper, J. I., Human-initiated failures and malfunction reporting. IRE Transactions in Human Factors in Electronics, vol. HFE, 104-109, September 1961.

Irwin, I. A., Levitz, J. J., and Freed, A. M. Human reliability in the performance of maintenance, Proceedings, symposium on qualification of human performance, pp. 143-198. Albuquerque, New Mexico, August 17 -19, 1964.

Koch, H., The Predicition of Human Performance: Theoretical and Practical Problems. Symposium and Workshop on the Quantification of Human Performance, The University of New Mexico, August 1964.

Mastropaolo, J. A., Human Error Research and Analysis Program (HERAP II) Final Report, Douglas Aircraft Company. 1 September 1967.

Mastropaolo, J. A., Human Error Research and Analysis Program (HERAP III), First Quarterly Progress Report. Douglas Aircraft Company. 10 July 1968.

McCormick, A. T. E. Human Error Research and Analysis Program (HERAP III), Final Report, Part II, Douglas Aircraft Company, March 1964.

Meister, D., The problem of human-initiated failures, Proceedings, 8th National symposium on reliability and quality control, pp. 234-239. Washington, D. C., January 9-12, 1962.

Miller, R. B., Task description and analysis. In Gagne, R. M. (Ed.) Psychological principles in system development. New York: Holt, Rinehart, & Winston, 1962.

Rook, L. W., Jr. Reduction of human error in industrial production. Albuquerque, New Mexico: Sandia Corporation, June 1962.

Sells, S. B., "Dimensions of Stimulus Situations which Account for Behavior Variance." in Sells, S. B. (Ed.) Stimulus Determinants of Behavior. New York: The Ronal Press. 1963.

Stinson, P. J. and Walsh, J. E., "An Application of the Poisson Approximation to Naval Aviation Accident Data." Presented at the 35th Session of the International Statistical Institute, Belgrad. 1965.

Swain, A. D. Technique for human error rate prediction (THERP). Paper presented at the Symposium and Workshop on the Quantification of Human Performance, University of New Mexico, 17-19 August 1964.

Tarr, J. S. Human Error Research and Analysis Program (HERAP III), Final Report, Part I, Douglas Aircraft Company. March 1969.

Willis, H. F. The human error problem, Report 62-76, Martin Marietta Corporation, Denver, Colorado, presented at American Psychological Association, September 1962.

APPENDIX A

The results of the A4 aircraft task analysis are shown on Figures 1 through 8. The figures illustrate the technique used and the relationship of work load to accidents.

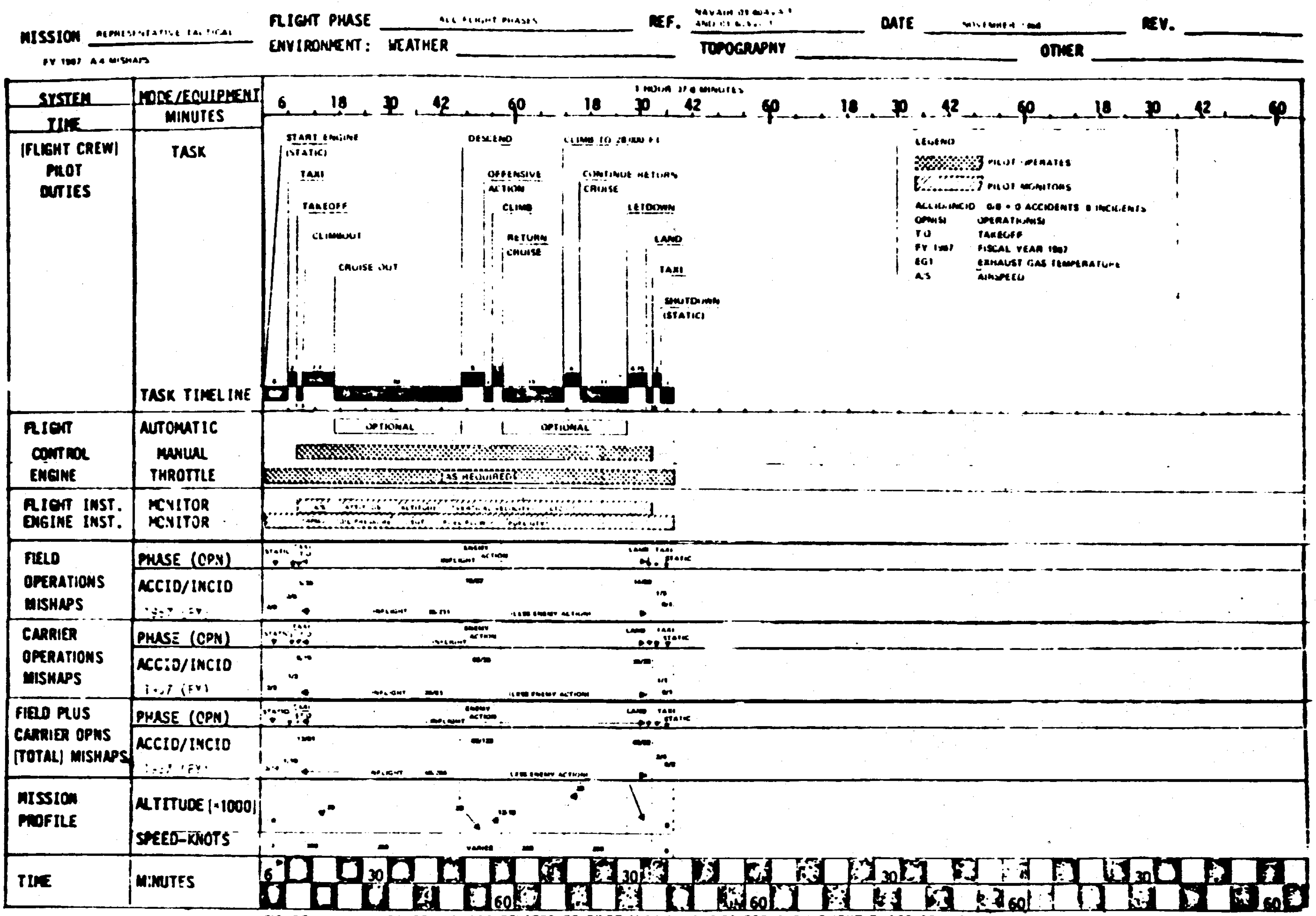


FIGURE 1. A-4 AIRCRAFT MISHAPS RELATED TO PILOT MISSION DUTIES FOR EACH FLIGHT PHASE OF A COMPLETE MISSION

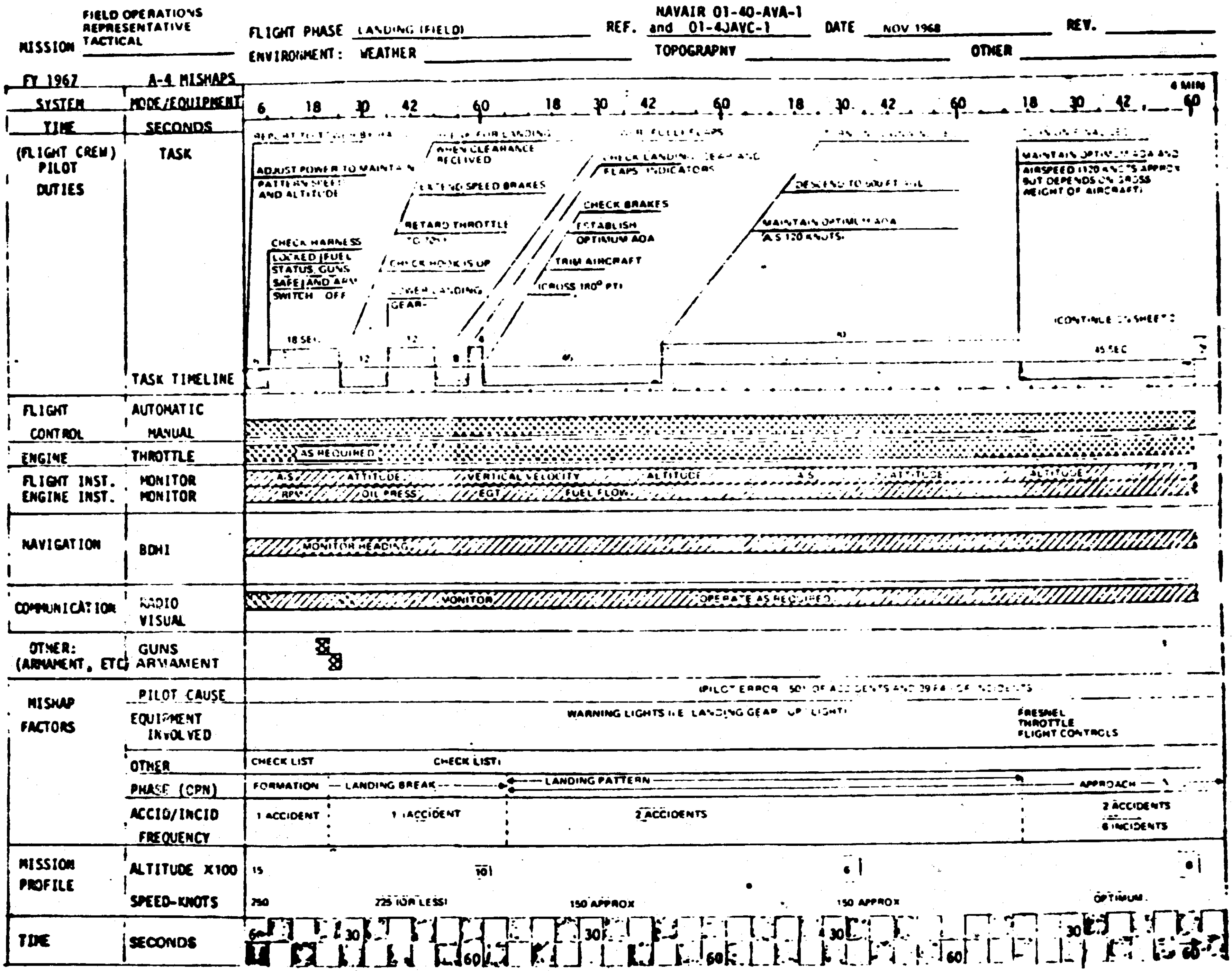


FIGURE 2. A-4 MISHAPS RELATED TO PILOT LANDING TASKS FOR FIELD OPERATIONS (1 OF 2)

MISSION FIELD OPERATIONS

FLIGHT PHASE LANDING (FIELD) REF. NAVAIR 01-40-AVA-1 and 01-40AVC-1 DATE NOV 1968 REV.

ENVIRONMENT: WEATHER TOPOGRAPHY OTHER

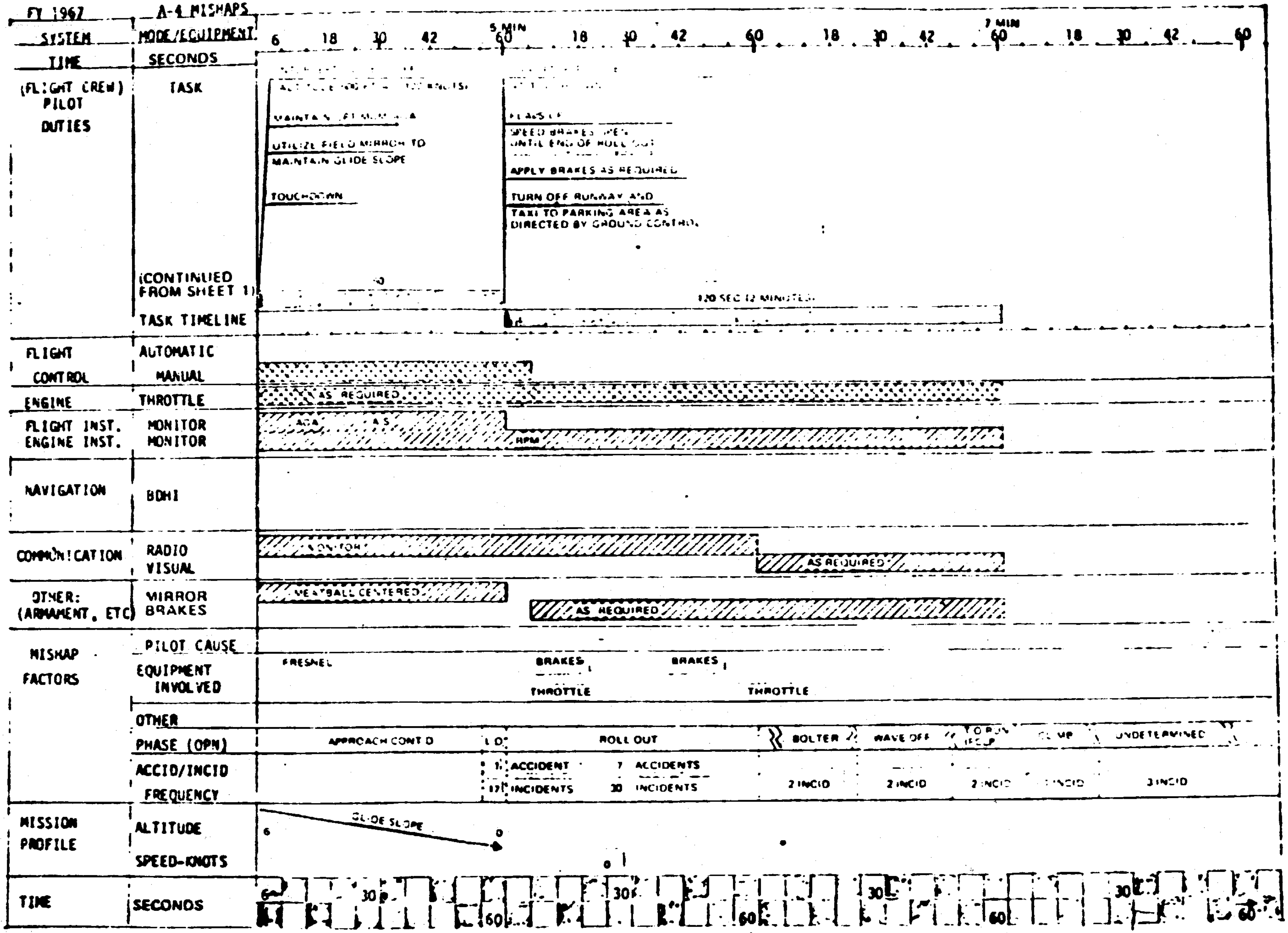


FIGURE 2. A-4 MISHAPS RELATED TO PILOT LANDING TASKS FOR FIELD OPERATIONS (2 OF 2)

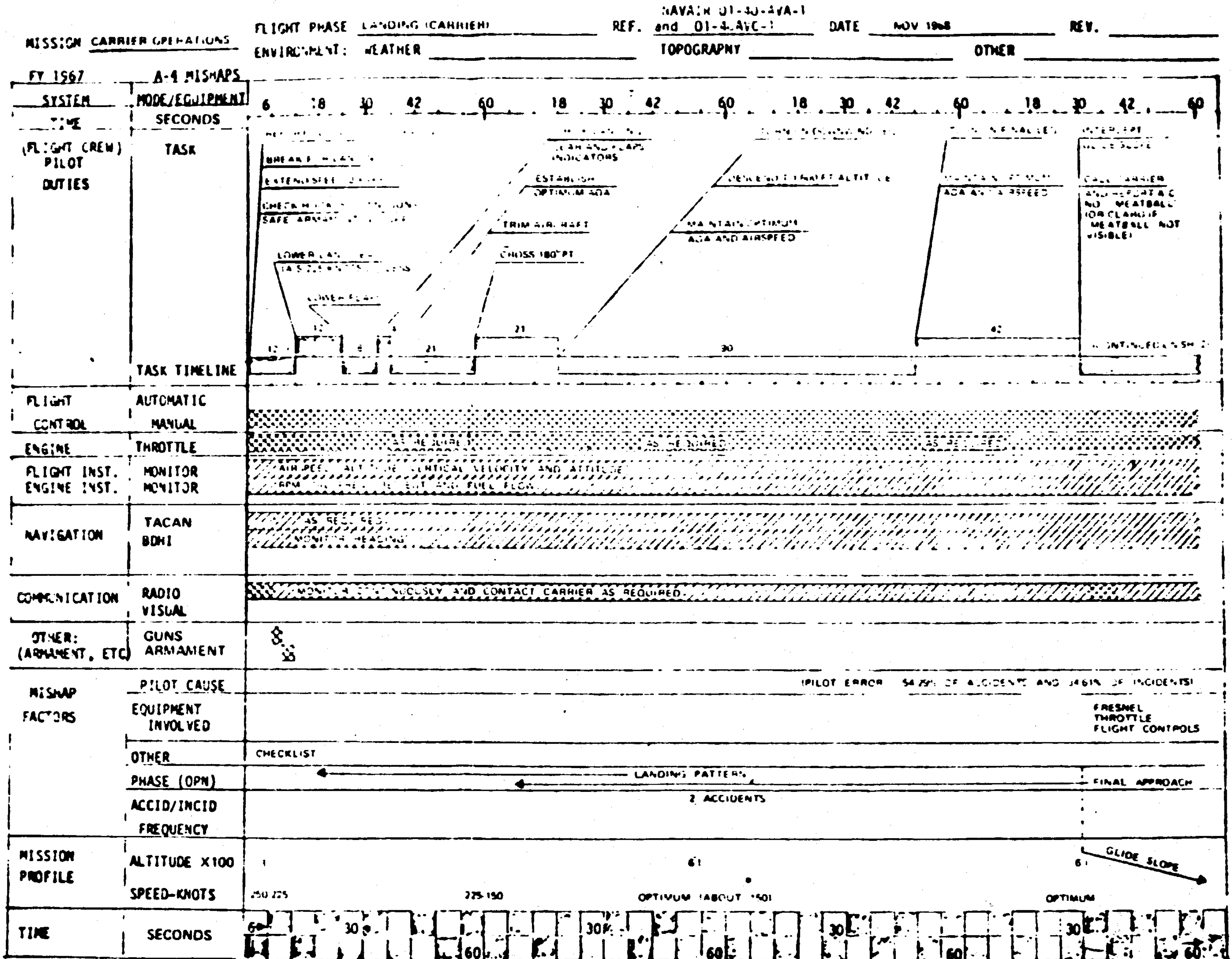


FIGURE 3. A-4 MISHAPS RELATED TO PILOT LANDING TASKS FOR CARRIER OPERATIONS (1 OF 2)

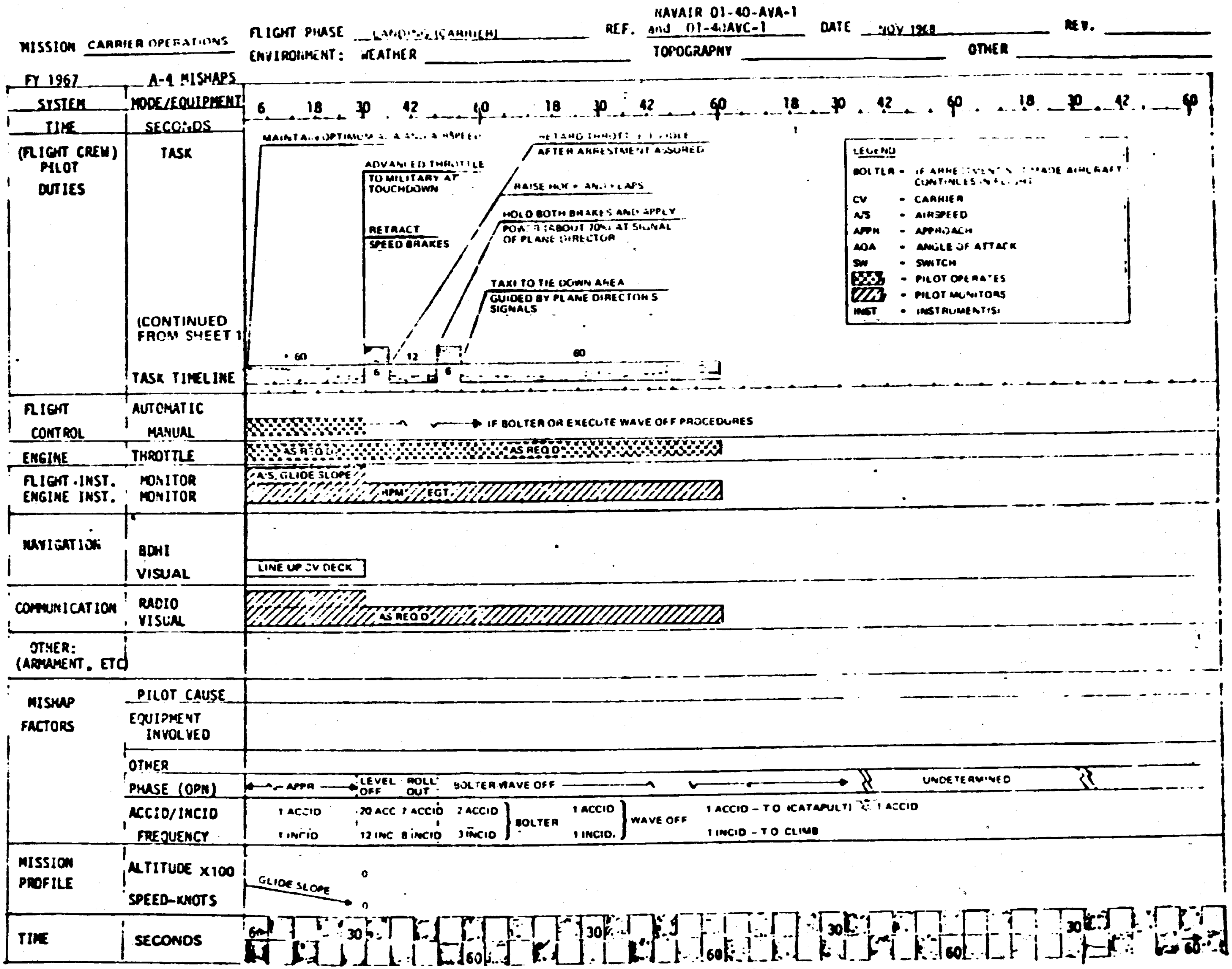


FIGURE 3. A-4 MISHAPS RELATED TO PILOT LANDING TASKS FOR CARRIER OPERATIONS (2 OF 2)

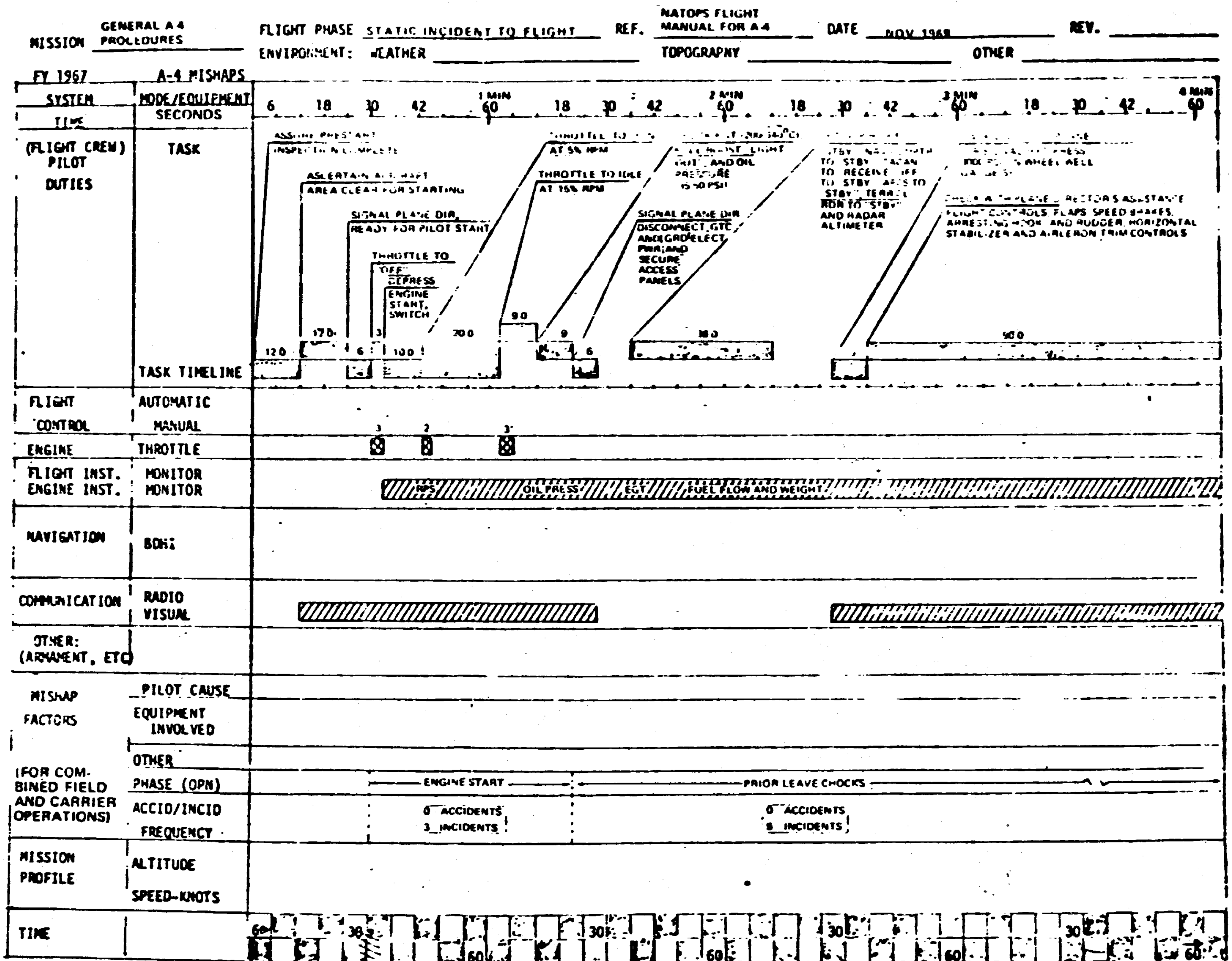


FIGURE 4. TASK ANALYSIS OF A-4 AIRCRAFT PILOT TASKS FOR STATIC INCIDENT TO FLIGHT PHASE AND RELATED MISHAPS (1 OF 2)

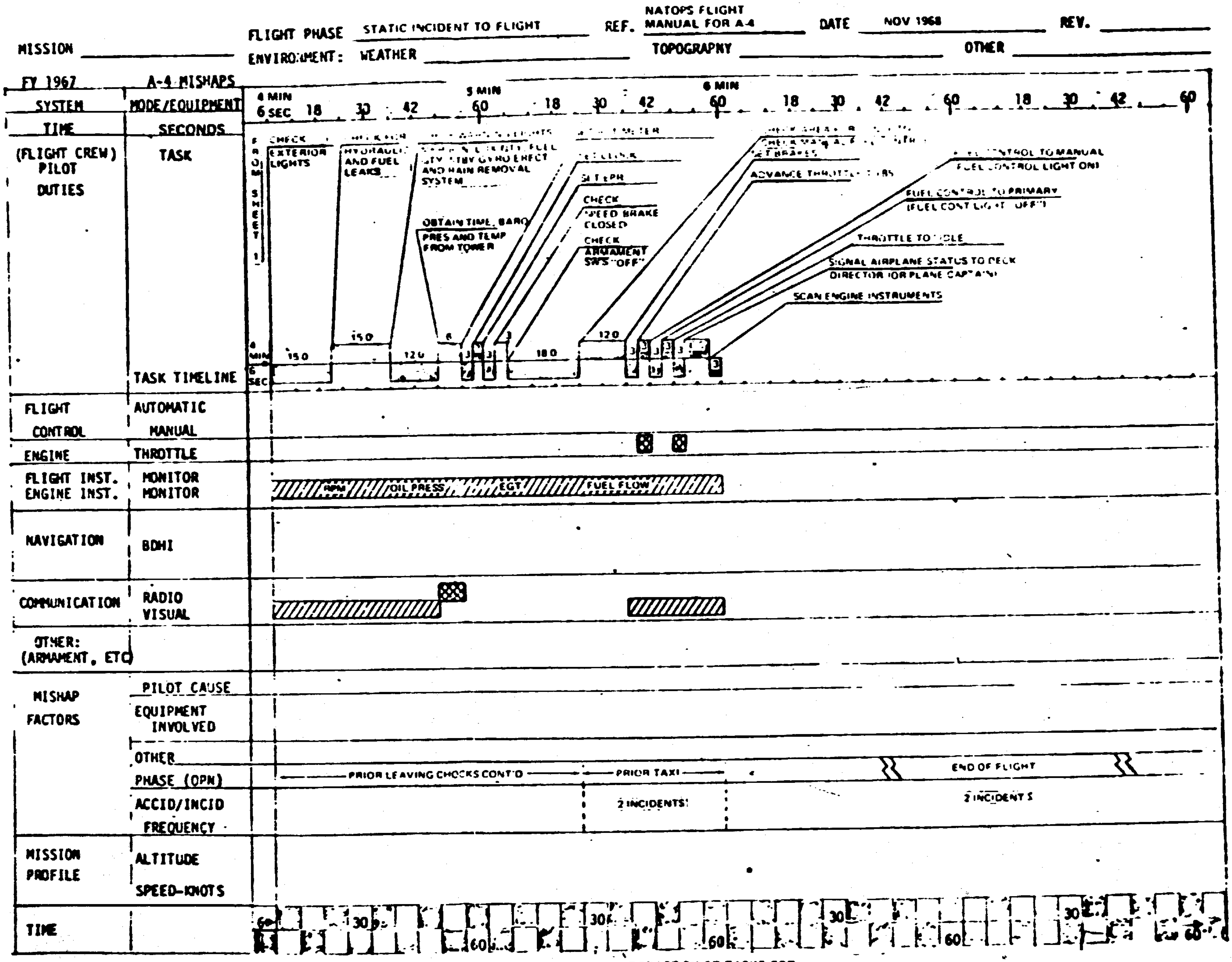


FIGURE 4. TASK ANALYSIS OF A-4 AIRCRAFT PILOT TASKS FOR STATIC INCIDENT TO FLIGHT PHASE AND RELATED MISHAPS (2 OF 2)

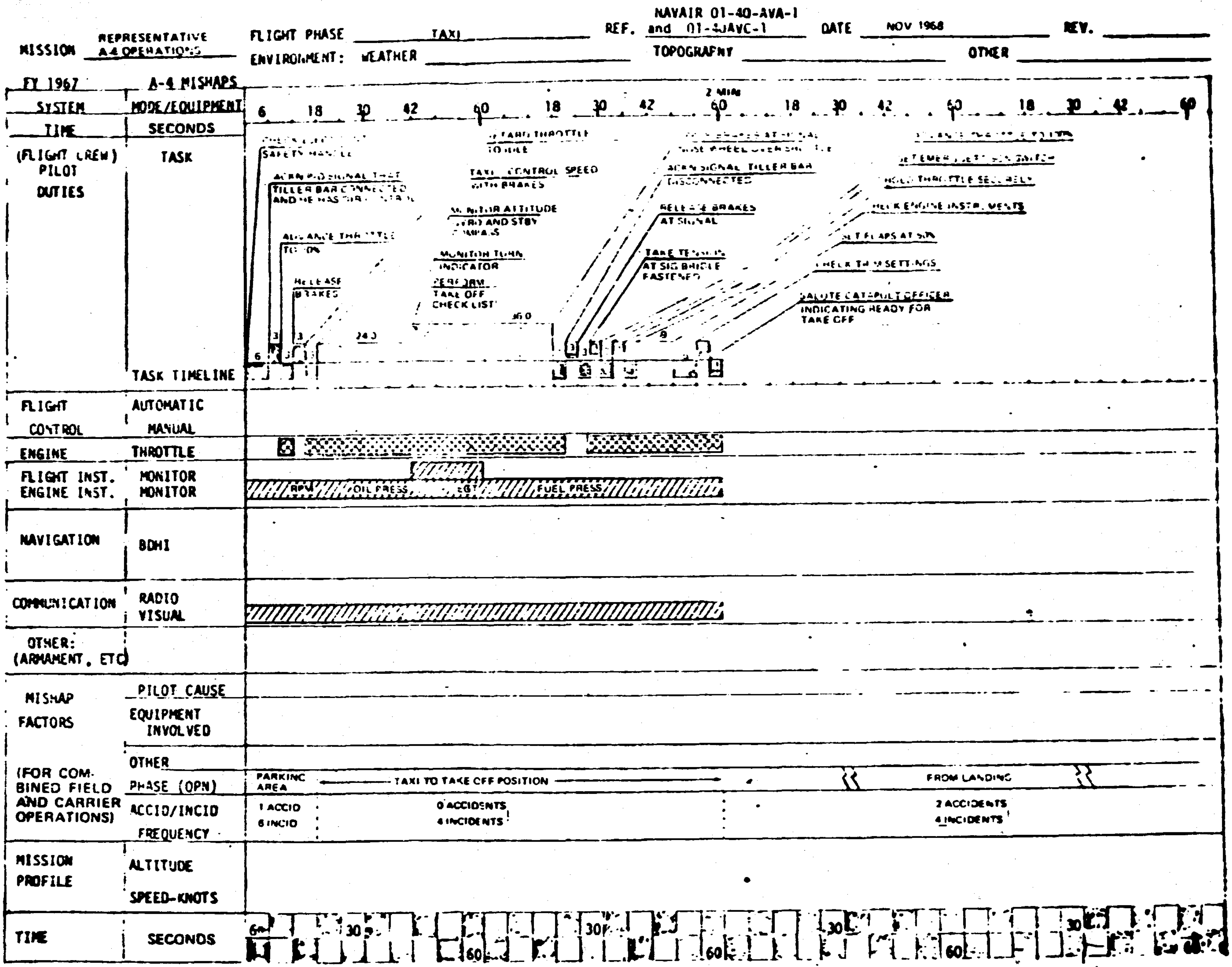


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT PILOT TASKS FOR THE TAXI PHASE AND RELATED MISHAPS

MISSION CARRIER OPERATIONS

FLIGHT PHASE TAKEOFF (CARRIER)

ENVIRONMENT: WEATHER

REF. NAVAIR 01-40-AVA-1 and 01-40AVC-1

TOPOGRAPHY

DATE NOV 1968

REV.

OTHER

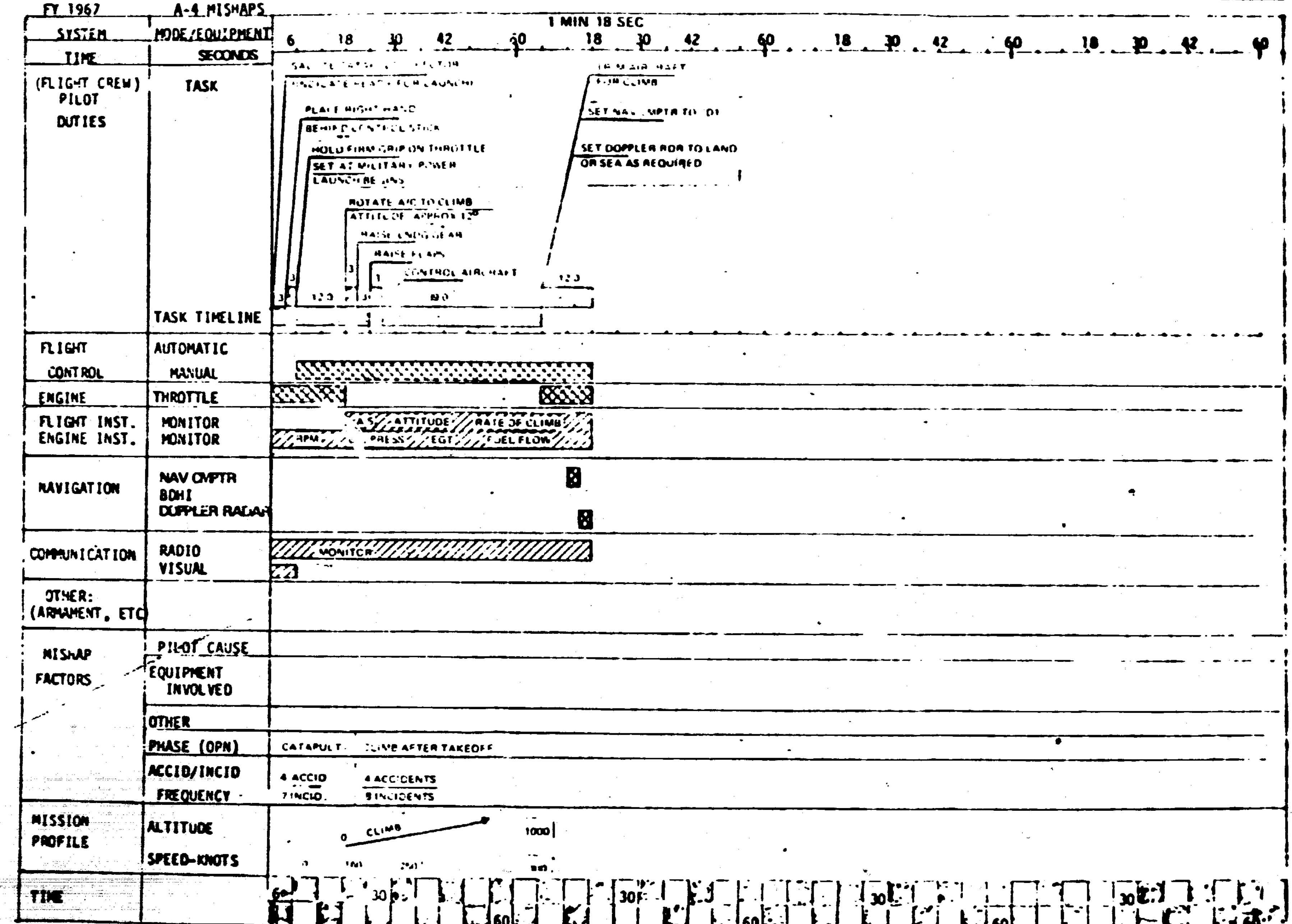


FIGURE 6. TASK ANALYSIS OF A A-4 AIRCRAFT CARRIER OPERATIONS TAKEOFF PHASE AND RELATED MISHAPS (1 OF 1)

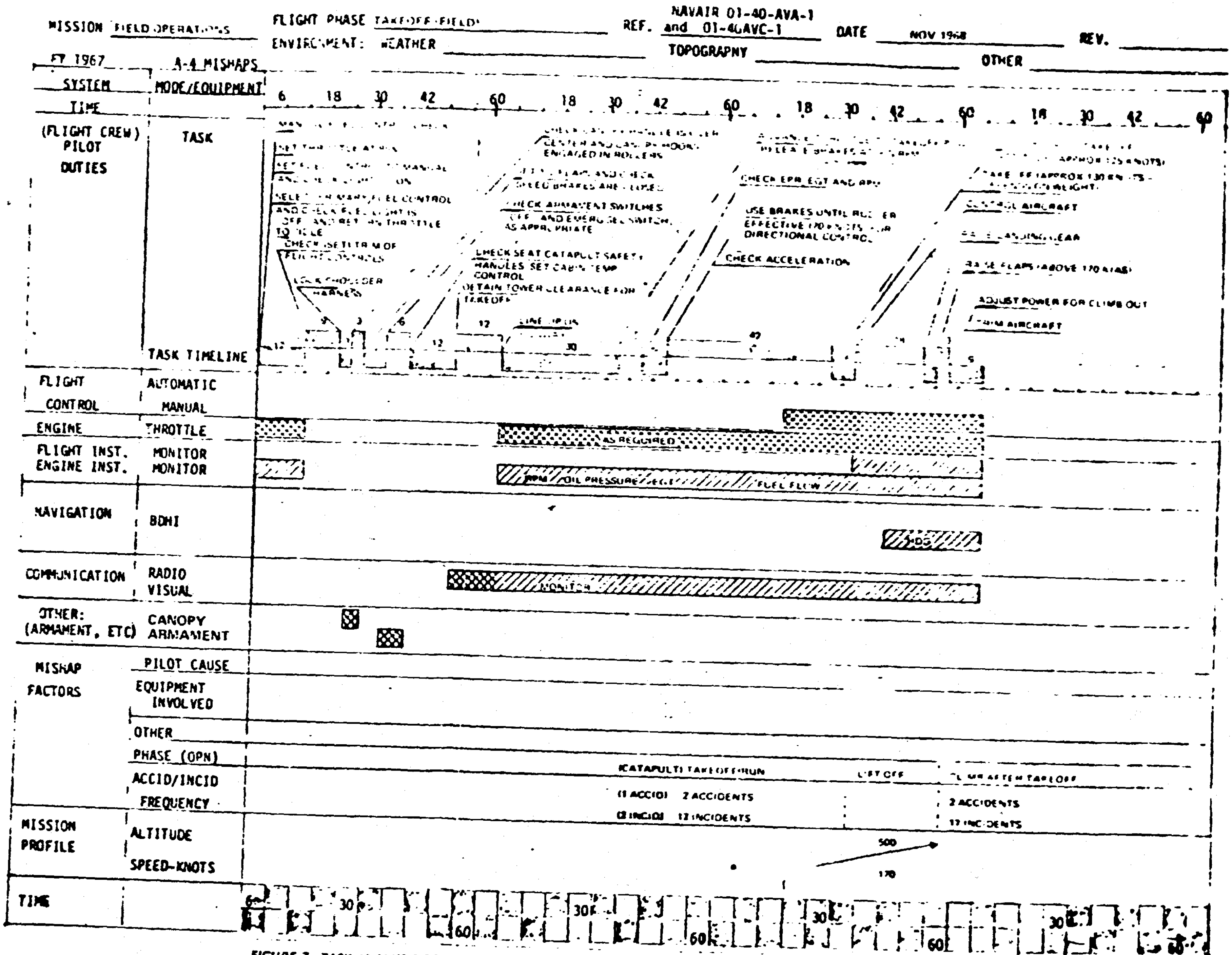


FIGURE 7. TASK ANALYSIS OF A 4 AIRCRAFT FIELD OPERATIONS TAKEOFF PHASE AND RELATED MISHAPS

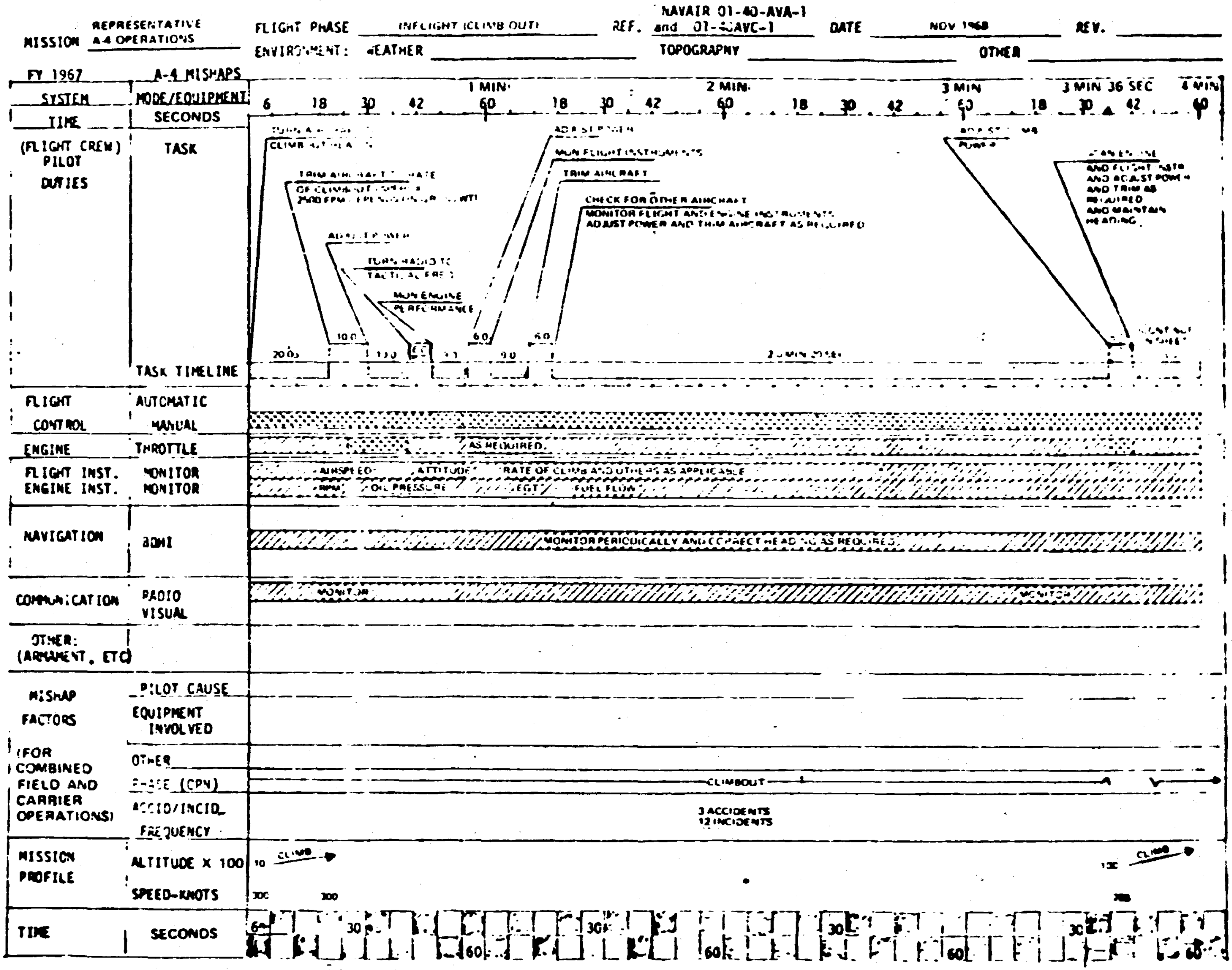


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (1 OF 7)

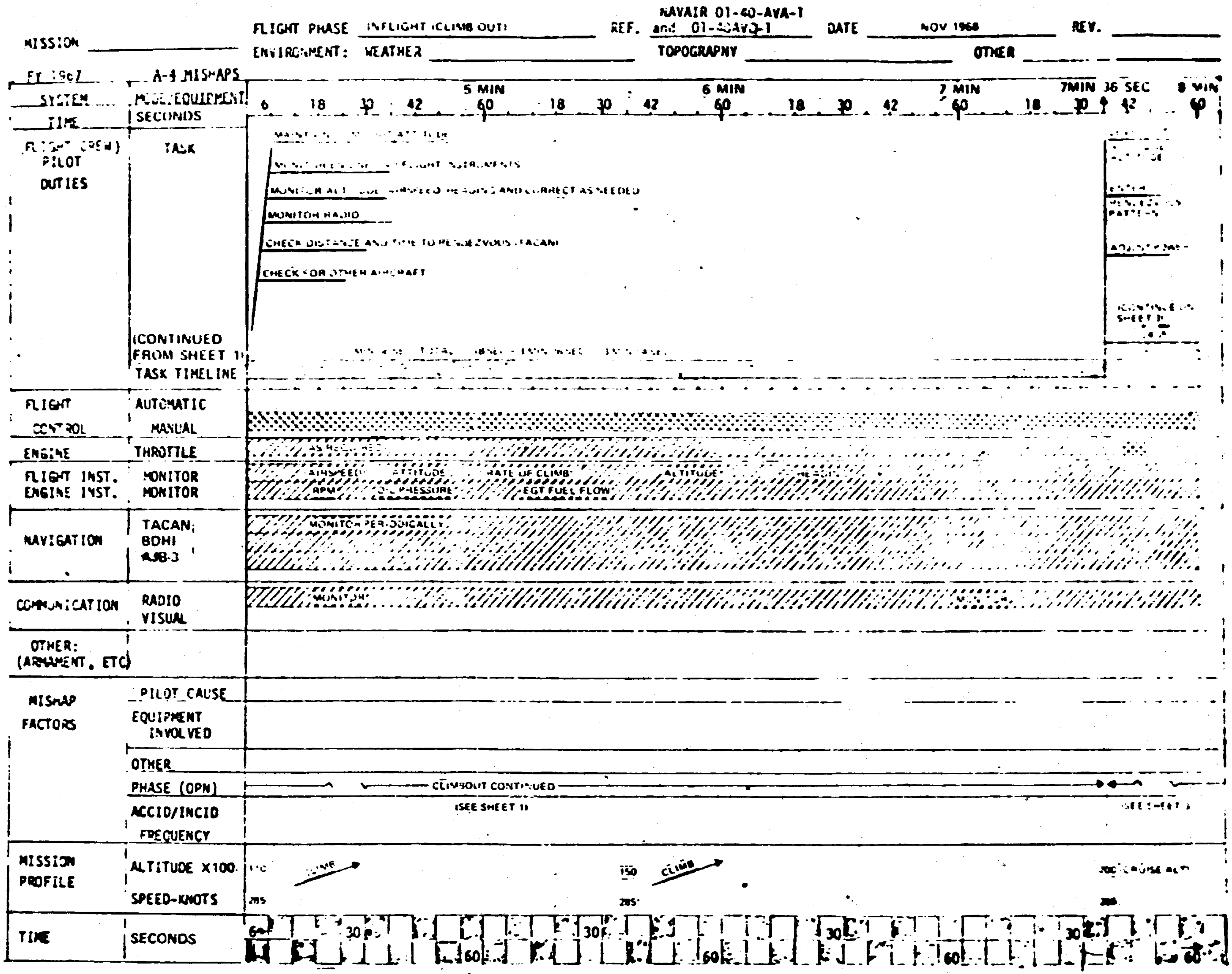


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (2 OF 7)

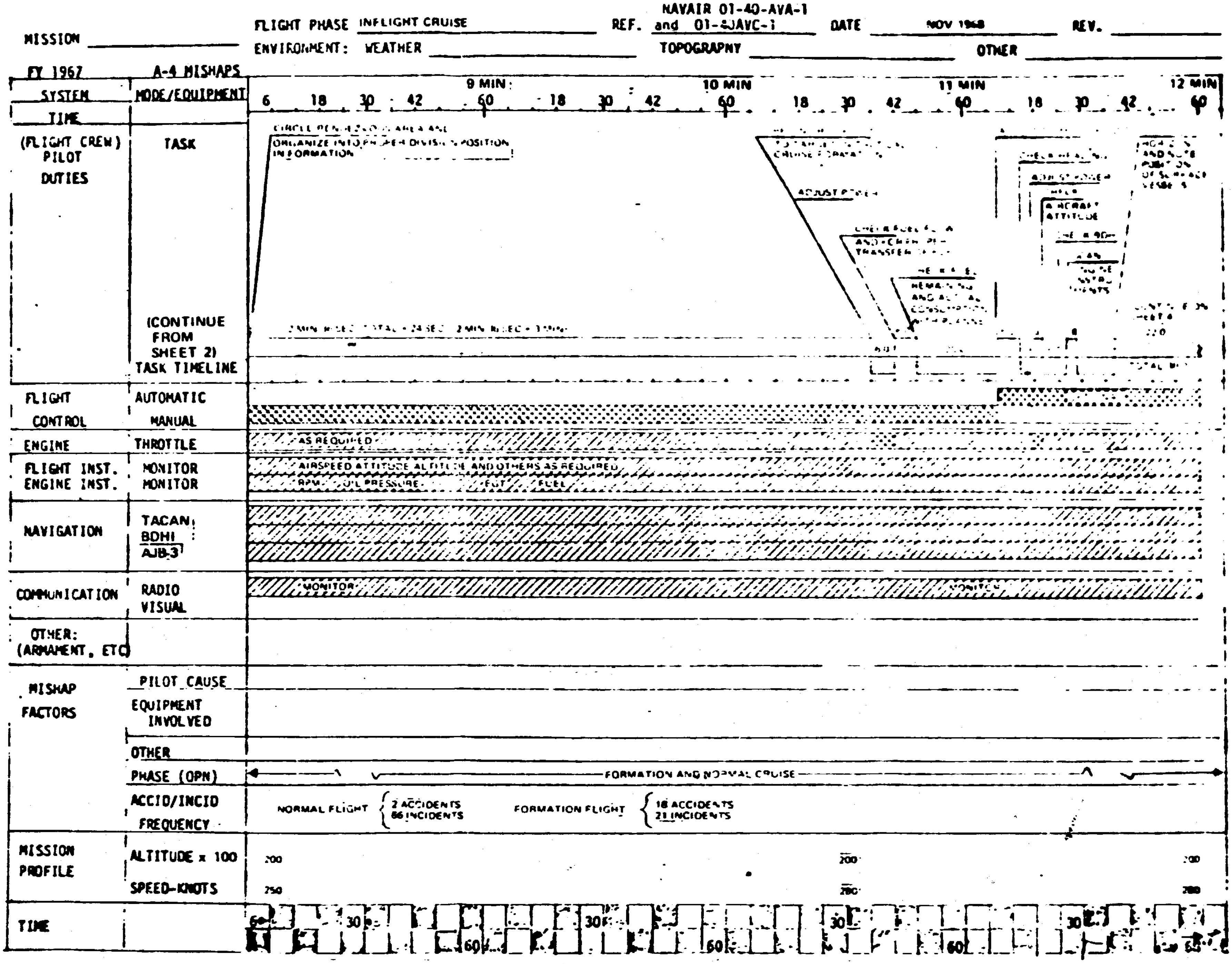


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (3 OF 7)

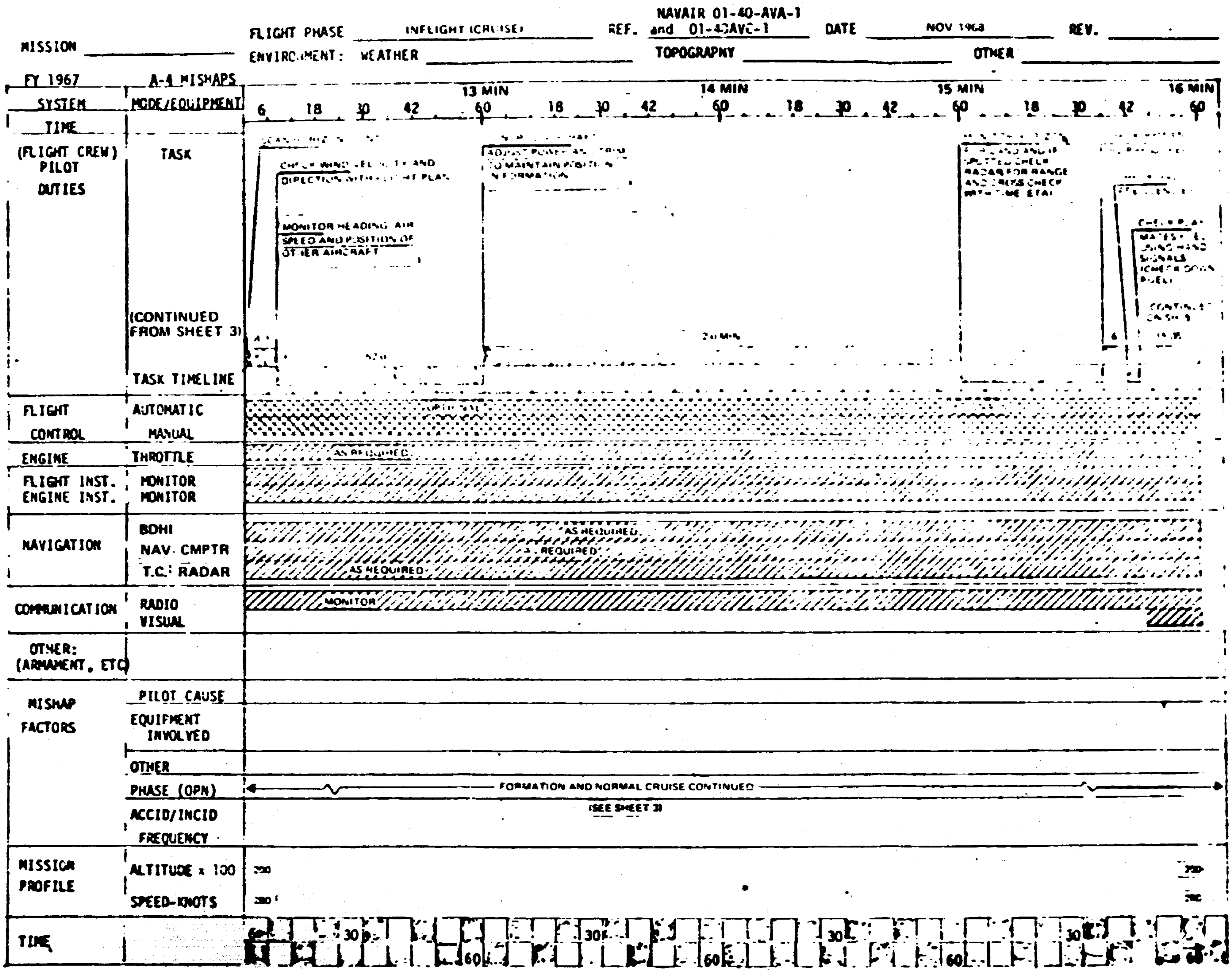


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (4 OF 7)

MISSION ______ FLIGHT PHASE INFLIGHT (CRUISE) REF. NAVAIR 01-40-AVA-1 and 01-40AVC-1 DATE NOV 1968 REV. ______

ENVIRONMENT: WEATHER ______ TOPOGRAPHY ______ OTHER ______

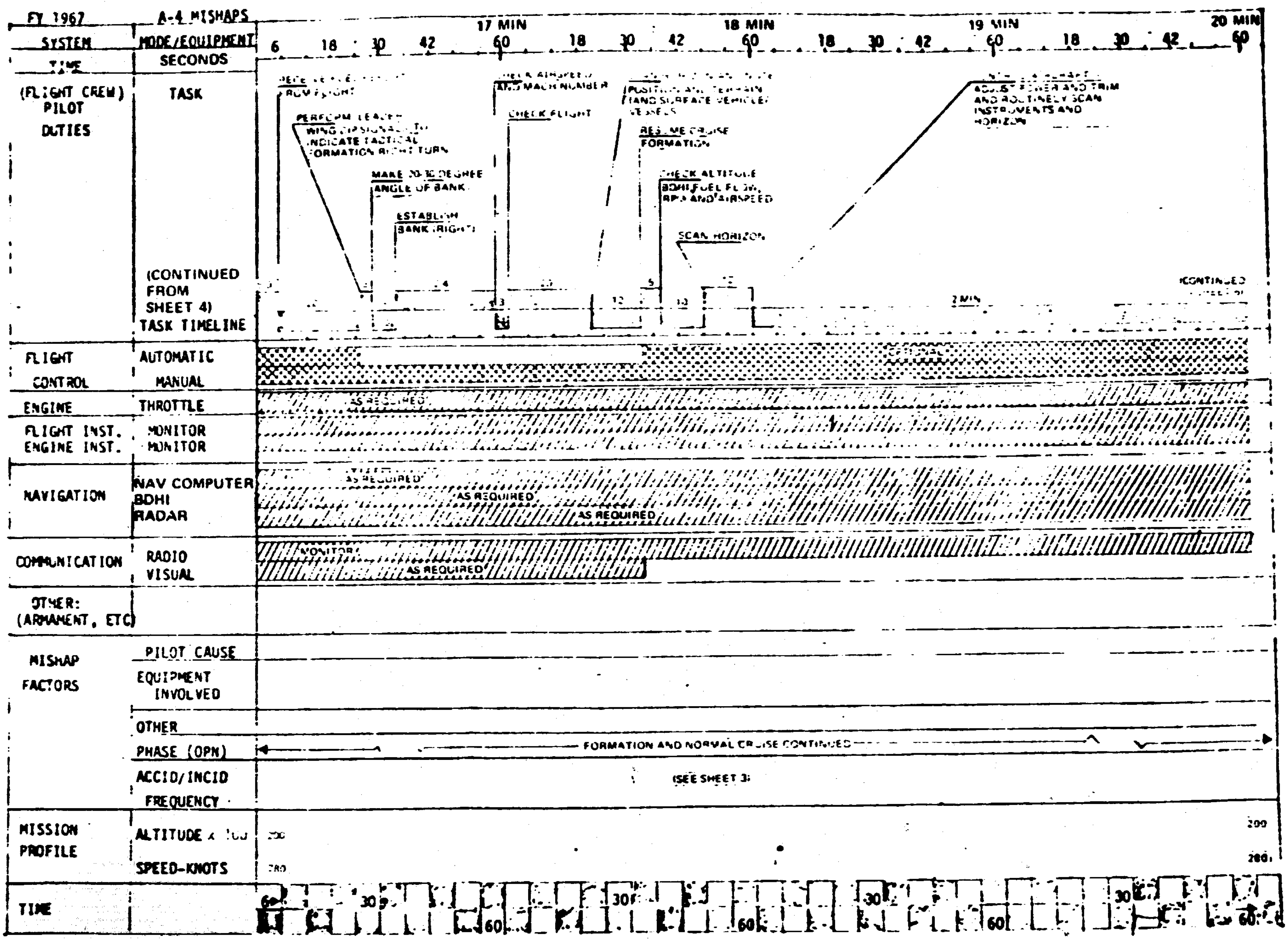


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (6 OF 7)

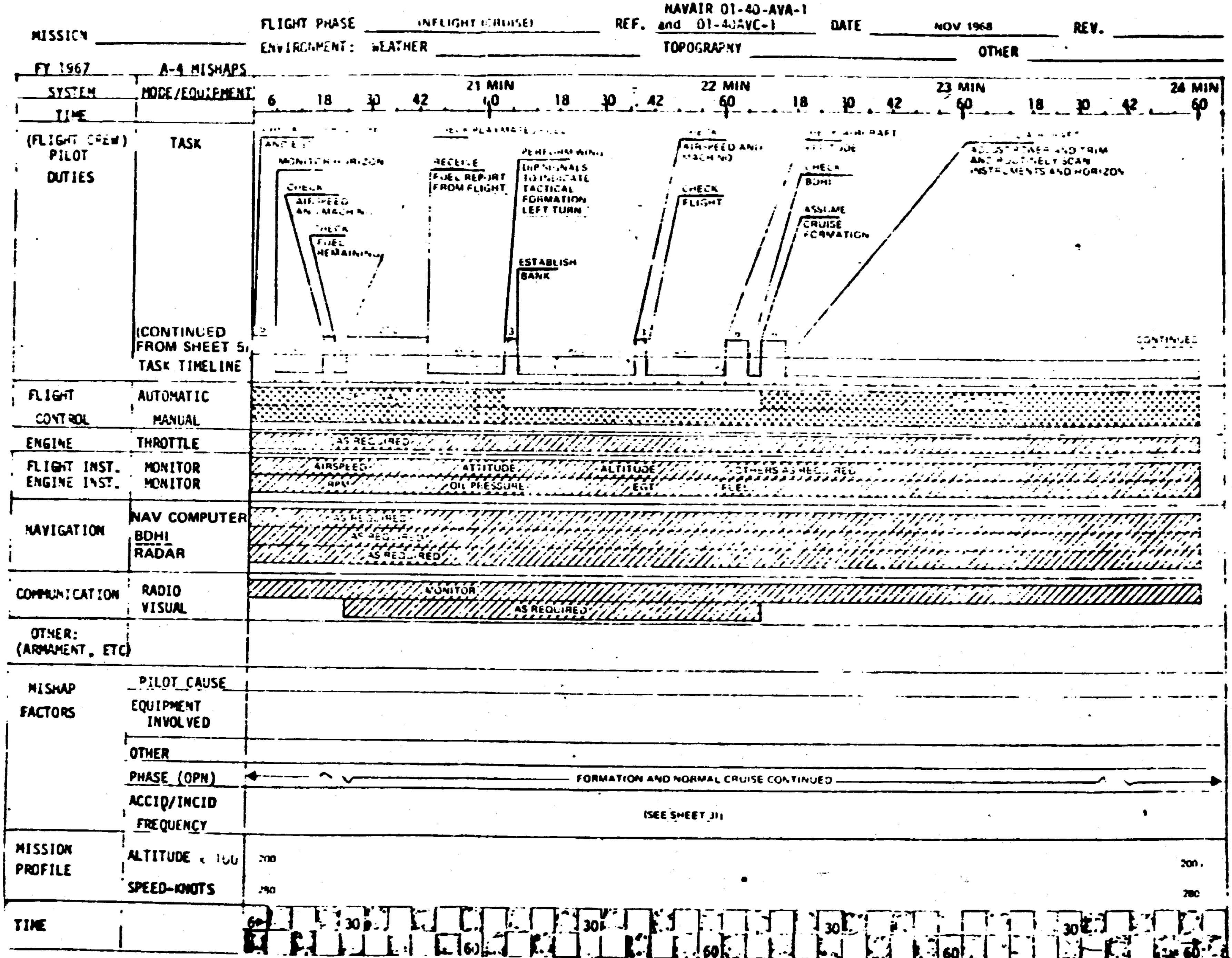


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (6 OF 7)

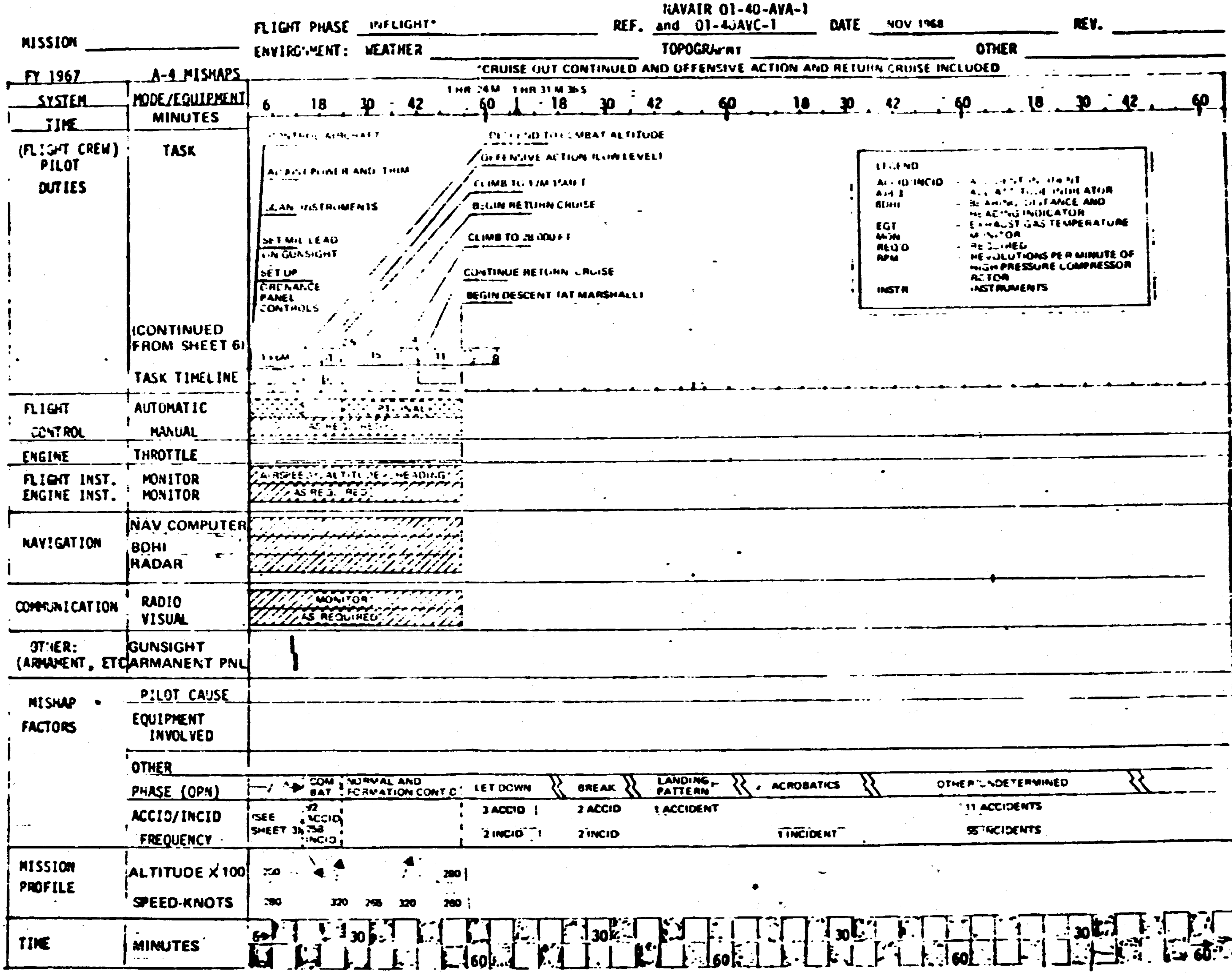


FIGURE 8. TASK ANALYSIS OF A-4 AIRCRAFT INFLIGHT PHASE AND RELATED MISHAPS (7 OF 7)

MDC - J0698/01

HUMAN ERROR RESEARCH AND ANALYSIS PROGRAM

Part III - Fatigue Study

A Laboratory Investigation of the Influence of 18 hours of Continuous Work on Human Performance

April, 1970

Prepared for the Naval Safety Center
Naval Air Station
Norfolk, Virginia

Under Contract No: N00189-70-C-0012

PREPARED BY:

L. R. Creamer, Ph.D.
Consultant, Science Research
Research and Technology

D. E. Wheeler, Ph.D.
Life Sciences
Science Research
Research and Technology

R. F. Gabriel, Ph.D.
Manager, Life Sciences
Science Research
Research and Technology

APPROVED BY:

A. A. Burrows, Ph.D.
Director
Science Research
Research and Technology

ABSTRACT

A state of fatigue, induced by continuous performance of tasks over prolonged working time, was studied in a laboratory environment. The tasks were selected to resemble the airborne activities necessary to a prolonged ASW mission.

Two loading conditions were imposed such that 12 subjects in a High Work Load (HWL) group worked without rest for approximately 18 hours and 12 different subjects in a Low Work Load (LWL) group worked for the same time span, but were permitted equal time on and off work, such that their workload was approximately one-half that of the HWL group. Psychological, physiological, biochemical, and performance measures were taken for each individual subject.

Performance decrement was obtained for the HWL reflecting at least moderate fatigue. A specially designed Discrete Tracking Task was the most sensitive index of fatigue. The task was moderately difficult with almost excessive temporal demands. The HWL group also made larger time estimation errors late in the session than the LWL group. No other psychological, biochemical, or physiological variable was correlated with performance decrement (Fatigue) in this study.

PREFACE

This report is one of the three parts of the Final Report of Contract No: N00189-70-C-0012.

Human Error Research and Analysis Program (HERAP).

The three parts are:

I. ___________(JACK TARR).

II. ___________(JACK TARR)

III. ___________(CREAMER, L. R.; WHEELER, D. E.; GABRIEL, R. F.)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

(Continued)

LIST OF TABLES

INTRODUCTION

Fatigue is an elusive phenomenon. Although virtually everyone has experienced it, fatigue is extremely difficult to define to everyone's satisfaction, and perhaps even more difficult to measure. Some authorities consider the term of negative value and prefer not to use it. Nonetheless, "FATIGUE" is frequently thought to be a contributor to accidents and reduced effectiveness of personnel. If so, understanding fatigue is crucial to all areas of endeavor, and particularly to military commanders, who may find it desirable to schedule personnel for frequent and/or extended missions.

Many studies have been performed to improve our understanding of fatigue. Fatigue, for most of the scientific studies, has been operationally defined as the degradation in performance or affective state resulting from previous work. This definition, of course, could include studies such as those in the area of vigilance, in which boredom may play a prominent role. Indeed, boredom may be on one end of an activity continuum with fatigue at the opposite end with normal daily activities along the mid points. However, boredom need not be a concomitant of work or even extended work. Therefore, it is considered for the purposes of this present discussion as a confounding variable and as something to be avoided in fatigue studies insofar as possible at the present time. Sleep deprivation for extended periods is also viewed as a separate, although relevent, variable and is purposely ignored.

Once fatigue is well understood, it will be possible to include its interaction with the variables of boredom, environmental stress, anxiety, abnormal sleep deprivation, diurnal cycle effects, and similar stresses. Thus, the present authors would accept the operational definition above only if qualified to exclude boredom and other confounding stresses.

Most of the fatigue studies reported are limited by the relatively limited nature of the investigation. In fact, Harris and his coworkers (1956) state after an extensive review of the literature that the level of performance as a function of time has never been studied systematically under long-term or short-term conditions.

A number of laboratories have conducted programs which included several studies. Especially notable were the studies conducted by the R.A.F. Institute of Aviation Medicine (Fraser, 1954; Fraser, 1955; Fraser, 1957; Fraser and Samuel, 1966; and Jackson, 1956). Among the conclusions reached by these workers are: (1) a significant fatigue effect occurs after continuous ten hours of flight in a piston engine aircraft; and after three one-hour sorties in jet fighter aircraft or after two one-hour sorties at night; (2) no significant fatigue effect is observable after continuous sorties for 3 - 4 hours in a jet bomber during the day; (3) considerable fatigue may be produced by flying four 15-hour sorties at night with one day's rest between each sortie for long-range piston engined aircraft; and (4) marked changes in performance and subjective states after prolonged repeated flights can occur without very extensive and, as yet, detectable physiological changes (Fraser, 1958). The specific dependent measure which seems to be the most reliable index of fatigue was the variance of the subjects' estimates of the time at which a slowly varying signal passed through the center of a projected display (the Z function). This task was performed before and after flight. Although one report in the series (Jackson, 1956) reported a decrement in holding a constant heading after forty minutes, this may be a function of boredom rather than fatigue.

Efforts to measure a decrement in on-the-job performance have been generally unsuccessful. A typical study (Shaw, 1954) involved Air Traffic Control operators who were administered a test before and after regular seven-hour duty watches when traffic was heavy enough for near continuous work. Although subjects occasionally complained of fatigue, no performance impairment was obtained. Grandjean (1969) notes that such studies have one inherent drawback; work has a stimulating effect on (a person) which may overcome existing fatigue. This is particularly true if the worker knows he is being observed or tested.

A number of studies have attempted to correlate physiological or psychological modifications to performance decrement. Physiological concomitants have not been promising, to date. Grandjean (1969) states that there are no practicable physiological or psychological objective tests of fatigue which can be used in industry with success.

In spite of the lack of systematic knowledge after considerable effort, there is a need for the definition of a quantitative relationship between the duration of activity and the degradational effects which everyone is sure exists, and objective methods of predicting decrement. Carpenter, Creamer, Gabriel, and Burrows (1969) defined a set of goals which appear valid. These included:

1. Define performance parameters influenced by continual work load.

2. Determine the work load which results in the degradation in performance.

3. Define physiological and/or behavioral indices of the output of fatigue.

4. Develop an understanding of the underlying mechanisms involved in performance degradation.

5. Discover methods of mitigating or alleviating degradation without adverse side effects.

In the Carpenter, et al., study, a large number of potential behavioral and physiological indices were evaluated in eight-hour simulated airborne ASW missions. They also attempted to evaluate Bartlett's hypothesis of increased variability and perceptual break-up (Bartlett, 1953). One of the limitations in the study, however, was the slight degree of fatigue brought about by the eight-hour simulated missions. It was decided, therefore, to conduct an experiment which would attempt to increase the fatigue state and use some of the most promising indices of fatigue used in the

Carpenter study. Psychological dependent variables selected were near and far phoria, time estimation and size constancy. Physiological variables deemed showing promise for application in the operational setting were heart-rate, respiration rate, basal skin resistance, and core temperature.

Several additional promising measures were added. Kalsbeek (1968) found that heart-rate variability ("sinus arrythmia") was directly related to the work load. This variability decreased as a function of the difficulty of the task. It was hypothesized that heart-rate variability over longer time periods might be related to fatigue.

The phospholipids yielded statistical significance as a function of load in the Carpenter study. Analysis techniques are involved and it was decided that operational use would be limited. Urine analysis of sugar, protein and pH was substituted since these analyses are easily performed.

To allow for an extension to 16 hours or more, a laboratory study was designed. Tasks, however, were selected to represent those typically required in airborne ASW missions.

In addition to determining sensitive indices of performance decrement, the usefulness of the variables in predicting specific performance decrement was to be assessed. This latter factor greatly influenced the experimental design.

In summary, the objectives of the current study were defined as:

1. Validate specific results of the Carpenter, et al. (1969) study.

2. Extend the fatigue state until greater performance decrement was obtained.

3. Define the correlation of performance decrement and psychological and physiological measures; that is, determine the amount of change for a given degree of performance decrement.

4. Evaluate heart-rate variability as a measure of fatigue.

METHOD

SUBJECTS

Twenty-four male graduate and under-graduate students from California State College, Long Beach, were used as subjects (Ss). Volunteers were solicited for paid participation. From those volunteering, Ss were selected on the basis of good health (see Appendix for screening questionnaire), prior military service (not possible in all cases), and willingness to complete the experiment or go without pay. The age of those selected ranged from 19 - 29 and averaged 22 years.

EXPERIMENTAL DESIGN

A Lindquist "Type I" experimental design was used (Lindquist, 1963). Twenty-four Ss were randomly assigned to one of two groups, twelve to the "Low" Work Load group (LWL), and twelve to the "High" Work Load (HWL) group. Work period, the second independent variable, was a within subjects' comparison. Subjects participated for one day.

Two Ss were tested each experimental day. Ss for a given day were assigned to the same workload condition to reduce motivational confoundings. Other effects for days were controlled by counterbalancing.

EQUIPMENT AND TASK DESCRIPTION

Experimental Area

The experimental room is represented in Figure 1. The room was partitioned by portable screens into three general areas: A work area, which included three work tables on which the various tasks were performed; a sleeping area, which

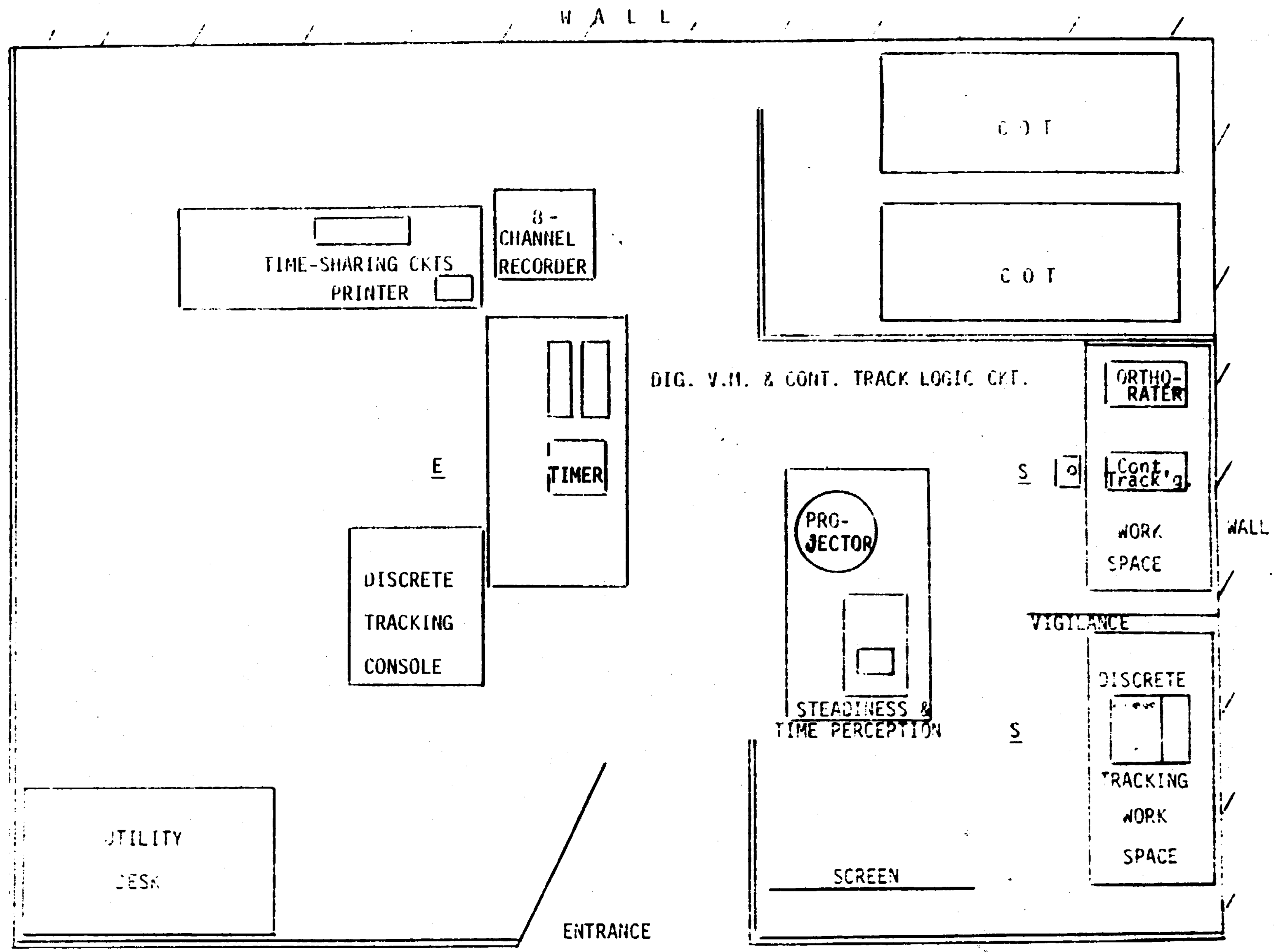


Figure 1 . LABORATORY OPERATING AREA FLOOR PLAN SACLE: 3/8" = 1'

III - 6

contained two Army cots and bedding; and an experimenter's area which contained task programming, data recording, and processing equipment. The arrangement was designed to minimize visual intrusions and distractions for the Ss. Lavatories were located approximately 50 feet away.

Temperature within the test chamber was maintained between 74° and 76°F. except during sleep periods, when it was raised to 78° to ensure comfort with a single blanket. Illumination was from diffused fluorscent lighting. Incident light on the work surfaces was 80 foot-candles. Noise levels in the laboratory ranged from 64 - 68 dBC. Much of the noise resulted from the air-conditioning and was greatest for the lower octave bands.

Physiological Recordings

Physiological data recorded included heart-rate, basal skin resistance, thoracic and abdominal respiration, oral temperature, and urinalysis of pH, protein and glucose. For the first three (continuous) measures an automatic timer controlled the circuitry so that for each successive six minutes, three minutes of data were collected for each S in alternate three-minute intervals.

To obtain heart-rate data, Beckman biopotential skin electrodes were applied to the shaven chest after cleansing the area with isopropyl alcohol. A detergent electrode paste was used. The heart-rate signals were amplified by a differential amplifier with a gain of 15,000 and a band pass of 5 - 45 Hz to enhance the signal-to-noise ratio.

To obtain values of variability in heart-rate, the basic data recorded were the times required for 10 successive heart-beats. This was accomplished by detecting the R-wave of each successive beat by a Schmitt trigger. An integrated "divide by 10" circuit was used to gate the initial pulse and reset after the 10th beat. An oscillator counted the time required for the ten beats and was recorded by a digital printer.

Data were acquired for both thoracic and abdominal respiration. Specially developed strain gauges, consisting of a thin brass seam, encapsulated in R.T.V. silicone rubber, were used. The strain gauges were connected to one-half-inch elastic bands sewn to a fabric belt. Velcro was used on the ends of the belt to allow adjustments. One belt was worn around the chest, while another one enclosed the waist. The outputs of the transducers were fed to self-balancing bridges and recorded on a Beckman strip chart recorder. Three minutes of data were acquired for each subject in a given six-minute interval.

Basal skin resistance was sensed with electrodes placed approximately 8 - 10 inches above the ankle on the right tibia. A 10 microampere 100 Hz activating current was passed between this electrode and the common ground electrode located approximately 2 inches above the right ankle. One value was obtained for each ten beats of the heart, processed through a digital voltmeter and printed out.

A belt pack, consisting of batteries, etc., was worn at the waist. Leads were bundled into a 45-inch cable for flexibility in movement from one task to another around the laboratory. For sleep, the battery pack was mounted on the wall adjacent to the head of the cots. All digitized heart-rate data were also recorded in analog form on the strip chart recorder for redundancy.

Oral temperature was obtained with a standard clinical oral thermometer.

Urinalysis to determine pH, glucose, and protein concentrations was performed by the Combistix method.

Block and/or wiring diagrams of this physiological data system(s) are provided in the Appendix.

Performance Tasks

Performance tasks were divided into two types - loading tasks and test tasks. The loading problems were designed to generate fatigue, while the test tasks were designed to measure fatigue unconfounded by differential practice. Tasks were selected to relate to typical ASW operator requirements.

The continuous tracking task was the primary electro-mechanical device required for the loading portion of the experiment. The block diagram for this system is provided in Figure 2. The S was required to center a "blip" on a five-inch cathode ray oscilloscope by manipulating a control stick. The "blip" was perturbated, in the horizontal axis only, by a random forcing function from a random noise generator. The control stick was spring-loaded to center, presented little resistance to displacement, and was free to move through 360°.

A feedback loop in the circuit increased the difficulty of the task as a function of the operator's (Ss) performance. The output of the averager was continuously fed back into the analog multiplier to vary the display sensitivity. The display sensitivity varied from 2cm/volt error in the easiest state to 20cm/volt error in the most difficult state. The voltage output of the averager was also fed through an amplifier and integrated over one-minute intervals of time. This value was displayed on an electronic counter as the performance measure. Thus, the dependent variable was the forcing function displayed as a unit of voltage in arbitrary units. Poor performance and/or low effort were represented by a low forcing function voltage. Difficulty level varied from 105 to 400 volts.

Additional loading tasks consisted of cognition tasks, such as solving navigational, plotting and fuel consumption problems. A total of 82 such problems were prepared. Twenty-seven of these problems were estimated-time-of-arrival problems, while thirty-four were plotting problems, and twenty-one were fuel consumption problems. Forms were prepared which provided the necessary data and Ss were to carry out the mathematical solution. The Appendix provides an example of each type of problem.

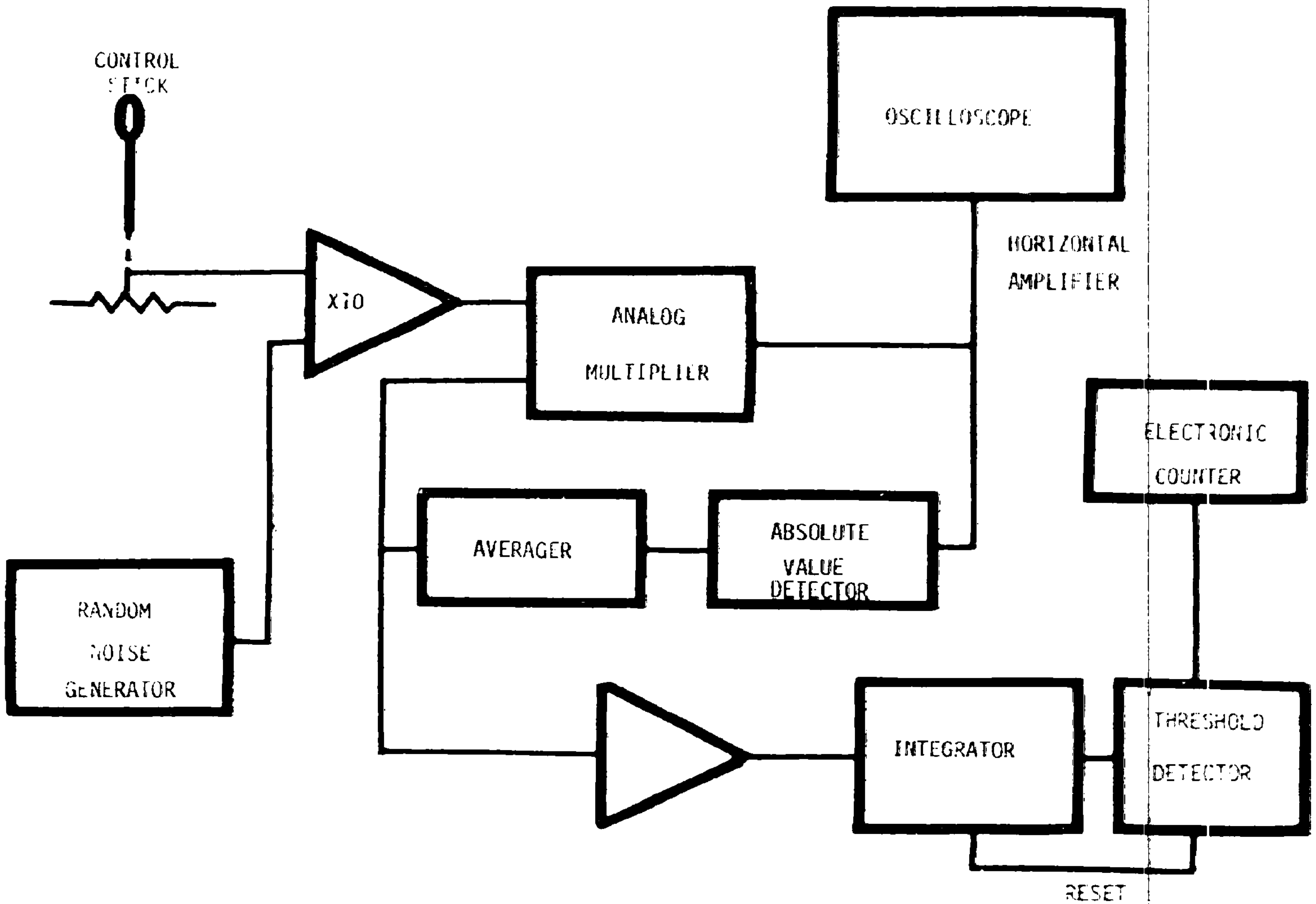


Figure 2 . CONTINUOUS TRACKING TASK.

III - 10

The test tasks included Discrete Tracking, Track-to-Intercept (TTI), and psychophysical tests of phorias, limb steadiness and time perception, and size constancy. The Discrete Tracking (DT) task consisted of six vertical and six horizontal lights in an " ⊢ " arrangement mounted on a vertical panel. Push-button switches on a keyboard were similarly arranged. If a light was on, the depression of the corresponding push-button would extinguish it. Two lights were illuminated at a time, one in the vertical row and one in the horizontal row, in a random order. The S's task was to turn out the lights as soon as he could, using the left hand for the vertical column of switches and the right hand for the horizontal row of switches. The presentation rate of the lights could be programmed for 26, 30, or 53 light pairs per minute and was independent of S's response. A switch on the experimenter's console was used to establish the desired rate. The circuit provided for the accumulation of the total time the operator was on target (i.e., lights turned out) for each one-minute interval.

A vigilance task was performed during the discrete tracking test. This task involved detecting the illumination of a small neon light located approximately at seated, eye-level and 30 inches to the left of the operator at a 45^{o}-angle. A push-button switch mounted on the table-top to the S's left was used to indicate signal detection and turn off the light. The light was presented at random, with the following constraints: a total of 5 presentations in each block of ten one-minute discrete tracking periods, no more than three presentations would occur in one five-minute period, and no lights presented before 15 seconds or after 45 seconds of a given one-minute interval had elapsed. A digital timer connected in series with the light and switches allowed the experimenter to record detection times. The light was turned off by E if not detected within one minute.

The S used a standard Bausch and Lomb Orthorater to measure near and far phoria.

Limb steadiness and time perception, although two separate tasks, were administered simultaneously. Limb steadiness involved inserting a one-eighth-inch metal stylus into a one-quarter-inch circular aperture with the preferred hand. Contact of the stylus with the side of the aperture completed a circuit which activated a digital timer. The recorded measure was error. The S initiated the trial when he was "set" and terminated when he judged his time estimation interval had elapsed. Time intervals to be estimated were between 34 and 73 seconds. One value was randomly selected for each trial. The S turned a switch starting the time estimation timer and actuating the limb steadiness circuit. When he judged that the given interval had elapsed, he turned the switch off and recorded his times. All Ss counted backward from 100 by 3's while performing these tasks.

The test for size constancy required S to estimate the actual length of a projected vertical line from a distance of 9 feet. Eighty-one photographic slides of narrow black rectangles, ranging in length from 5 to 50 cm in 1 cm units, were projected on a white paper screen. Ten slides were displayed during each test unit, by means of a Carousel slide projector. The order of the slides was randomly determined and repetition of lengths was also random. The S viewed the projected image and using a reference scale provided at his station, estimated the length of the image in centimeters. Figure 3 provides representative views of the apparatus. The Appendix provides the specific task instructions.

PROCEDURE

Preliminary Study

Four Ss participated in a preliminary study consisting of two test days, one day for each work load condition. This "pilot" study served as a dress rehearsal for the experimenters, established S training requirements and assisted in refining the experimental schedule. These data were not included in the results.

Daily Schedule

On a test day, two Ss reported for duty at 0845 hours. Both Ss were assigned to the same work load group. After a preliminary briefing and familarization with the laboratory area, Ss filled out a questionnaire and had their oral temperature taken. Next, a brief familarization with the task equipment was provided. During the briefing, Ss were informed of the monetary bonus of $20.00, $12.50, and $5.00, for the best 3 performances in their group. Although Ss knew about the general nature of the study, they were not informed of the existence of two groups. The physiological recording preparation consisted of the attachment of sensors and the physiological harness to the Ss. The foregoing activity required about one-half hour.

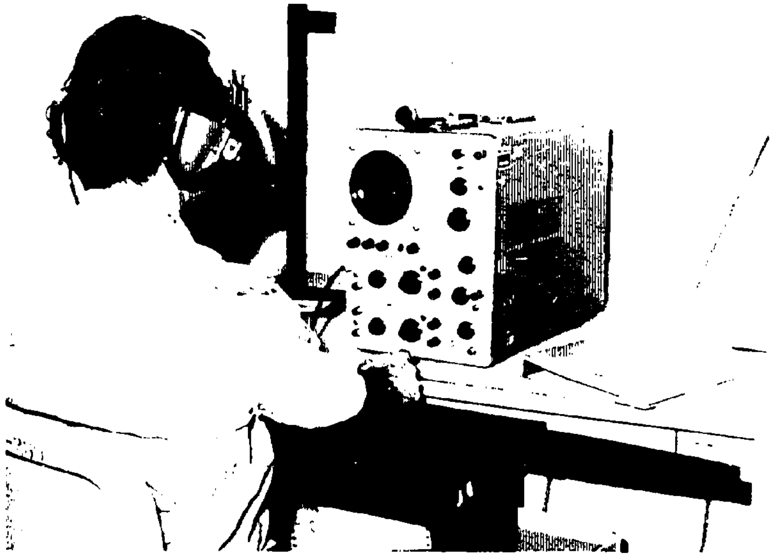


Figure 3a CONTINUOUS TRACKING TASK

Figure 3b DISCRETE TRACKING TASK

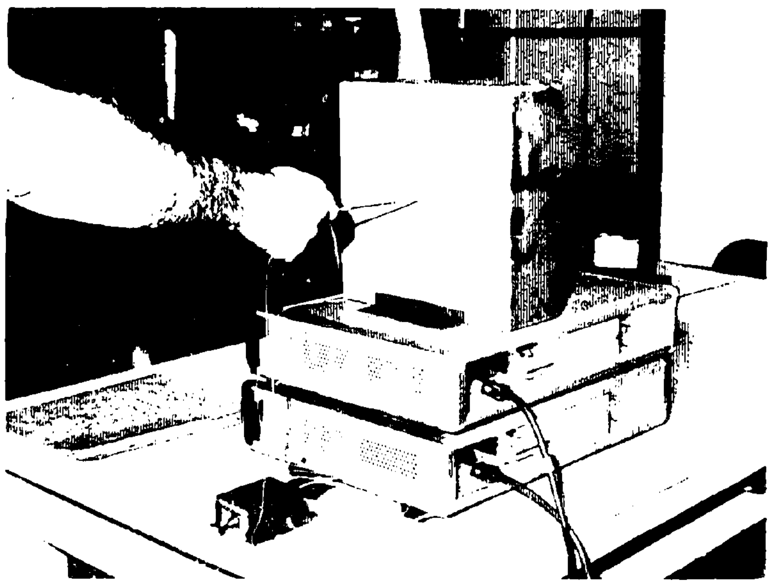


Figure 3c LIMB STEADINESS TEST

Figure 3d. VIEW OF THE WORK AREA FROM BEHIND EXPERIMENTER

Figure 3. Representative Views of the Laboratory

Specific, in-depth training on all tasks was given next. An average of about two hours was devoted to this training, although this time varied somewhat, due to individual differences. Since most of the tasks were largely self-administered, it was important to be certain that the Ss were thoroughly familiar with the requirements. Training consisted of demonstrating the task, instructing in the specific components, and providing practice with feedback.

Two complete test cycles were performed as covariant measures before the load-test cycles were started. Since only one set of equipment was available, Ss were sequenced through tasks in the same order but offset by one task. The only difference between covariant tests and later tests was the presentation rate on the discrete tracking task, which was increased from 26 to 53 light pairs per minute.

Subjects started the loading tasks immediately after completing the second covariant session.

The general plan of the experiment was to load the Ss for a series of consecutive time periods with test periods in between. The loading period lasted approximately 70 minutes, while the test period required 20 minutes (see Figure 4). One S worked on continuous tracking while the other S worked on cognitive problems. After 35 minutes, they changed place for 35 more minutes.

A similar procedure was used during the test tasks. One S would work for the assigned length of time (approximately 5 minutes) at one task, e.g., discrete tracking, while the other would work at the psychological tasks, etc. in the test phases. However, the Ss completed each test task just once, before going back to the loading tasks to start the whole cycle over.

The Ss in the HWL group received no breaks, except for 1 or 2 trips to the lavatory, until just before the sleep period. Eating (sandwiches, crackers and peanut butter, fruit, cookies, and candy bars) took place while the S was working on work load tasks. Food was provided at 1330 and 1900 hours.

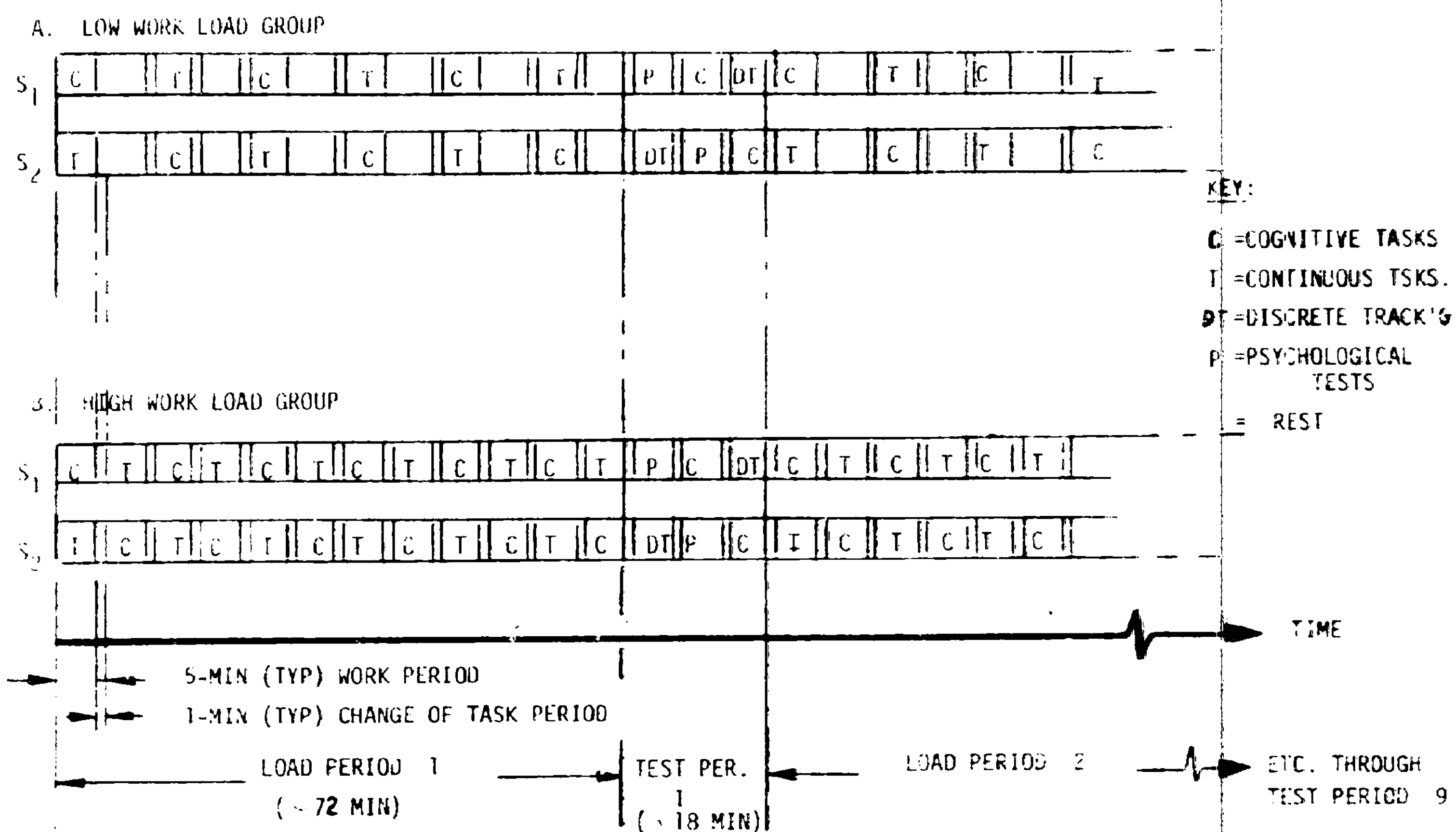


Figure 4. EXPERIMENTAL SCHEDULE

The LWL group followed the same schedule as the HWL group, with but one change. After each 5 min. in the work load phases, a 5-min. rest was provided. There fore, the LWL Ss were loaded only one-half as much as the HWL Ss, excluding test periods.

The completion of the ninth cycle terminated the work portion of the test day. The physiological harness was uncoupled and the Ss taken to the lavatory, where a urine sample was obtained and tested. When the Ss returned to the laboratory, they completed the self-inventory which was designed to determine their subjective state of fatigue. The physiological harness was then re-coupled and the Ss retired to the sleeping area. Laboratory lighting was extinguished during the sleep phases.

The Ss were awakened at 0730 hours, after which they completed one additional test period identical to those of the preceeding day. The Ss were then released and the laboratory secured.

Three experimenters (E) worked 8-hour shifts. Neither Ss nor Es were allowed out of the laboratory except for the required lavatory visits.

RESULTS

The results are presented under three separate headings. First, evidence is provided to determine whether fatigue was produced in the HWL. Potential correlates of fatigue are then considered and, finally, data testing Bartlett's Hypothesis are presented.

During the long, continuous experimental sessions, some data were lost. Equipment failure was the primary cause of lost data, but in two cases the Ss failed to understand the instructions. In most cases where data were lost, data from a comparable S in the other group were discarded, so that analyses were balanced for order effects. A very few physiological values were estimated by using adjacent values. In these cases, the Degrees-of-Freedom (df) in the error term was corrected.

The experimental design is shown schematically in Figure 5. Subjects are nested within workload, but are crossed with test periods (or loading periods or time). For Discrete Tracking, two additional variables were added (Minutes within test periods and horizontal versus vertical Time-on-Target (TOT)). The basic design remains unchanged, however, with these added variables being crossed with all other variables.

The .05 level was used for all significance tests. Any F ratio followed by an asterisk is significant at $p \leq .05$.

OPERATIONAL DEMONSTRATION OF FATIGUE

It was important for the purposes of this experiment that there was evidence of fatigue. Further, magnitude of fatigue at different times in the experimental session would allow for correlation of potential fatigue indices with performance decrement attributable to fatigue.

Discrete tracking was the primary task to demonstrate fatigue, but continuous tracking, the cognitive track to intercept problems, and the Self-Report Inventory, were also used. Since a good deal of learning was anticipated, the divergence of the LWL and HWL functions was considered to be the best operational demonstration of fatigue. Further evidence of fatigue could be inferred from a decrement for the HWL late in the experimental session, should a decrement occur.

Discrete Tracking

The performance measure for Discrete Tracking (DT) was a Time-on-Target (TOT) in seconds cumulated over one-minute-periods within the Horizontal and Vertical dimensions. The pace was rapid producing the relatively low mean TOT scores shown in Figure 6.

The most salient feature of these data is the significant WL X Test Period interaction (see ANOVA Summary in Table I). The groups begin to diverge on the fifth test period and the difference is obvious on the seventh test period.

Both groups show approximately equal performance decrement on "the morning after", Test 10. The relative difference between the groups is virtually unchanged by the short sleep period (4 - 6 hours).

Tract-to-Intercept Cognitive Problems

The cognitive test problems, track-to-intercept, were scored in three parts, each having a value of one. Therefore, a S could obtain 0, 1, 2, or 3 correct for a particular test period. The mean number correct responses are shown in Figure 7. The ANOVA F values are also shown.

In order to reduce the error variance and also to statistically equate the two WL groups, an analysis of covariance was run. The two preliminary covariant trials were combined as the covariant and test periods 6, 7, 8, 9, and 10 were summed for the variate. The WL, $F\ (1,21) = .849$ was not significant.

TABLE I

Analysis of Variance Summary Table for Discrete Tracking

Source	df	Mean Square	F.
I Work Load	1	657.46	0.48
K Test Period	9	540.03	2.02
L Minutes in Test Period	4	33.98	2.91 *
M Horizontal/Vertical	1	509.65	0.73
J(I) Subject (nested within)	14	1365.52	
IK	9	267.78	3.97 *
IL	4	7.77	0.67
KL	36	6.62	1.13
IM	1	13.79	0.02
K[illegible]	9	12.41	0.62 *
LM	4	6.25	4.43 *
JK(I)	126	67.46	
JL(I)	56	11.67	
JM(I)	14	694.84	
IKL	36	5.50	0.94
IKM	9	18.79	0.90
ILM	4	1.41	0.07
KLM	36	2.92	0.90
JKL(I)	504	5.88	
JKM(I)	126	20.83	
JLM(I)	56	4.34	
IKLM	36	3.27	1.00
JKLM(I)	504	3.27	

* Significant at .05 level

Continuous Tracking

The adaptive Continuous Tracking (CT) task yielded a response measure which reflected the level of difficulty. This forcing function was measured in arbitrary voltage units, such that large values represent hard work and good performance combined. With the control stationary, a score of 105 resulted in a one-minute-period. The best score obtained was just above 400.

The means are shown in Figure 8 . It is obvious that the groups differ for all loading periods. The WL X Loading Period interaction was not significant to support the trends for the High and Low WL functions to diverge across loading periods.

The main effect between groups dissipated when an Analysis of Covariance was used. Loading period one was used as the covariant and periods 6, 7, 8, and 9 were combined as the variate. The $\underline{F}$ (1,21) of 1.704 was not reliable. The LWL adjusted mean was 372 and the HWL adjusted mean was 348 with the difference still favoring the LWL group.

Using the DT function for an estimate of the onset of fatigue, one further analysis was performed to see if the WL functions diverged significantly after the fifth loading perid. The $\underline{F}$ (3,66) = 4.651 was significant for the WL by Loading Period interaction. The HWL decrement was significant for these four loading periods.

During the last four load cycles performance did deteriorate with the loading cycle. Means were computed for each of six five-minute work periods for the cycle. From the first to sixth, the means were: 354, 341, 336, 328, 326, and 327.

Self-Report Inventory

The Ss self-report responses are summarized in Table II. Means for High and Low WL's and Estimated Standard Error of Means are given.

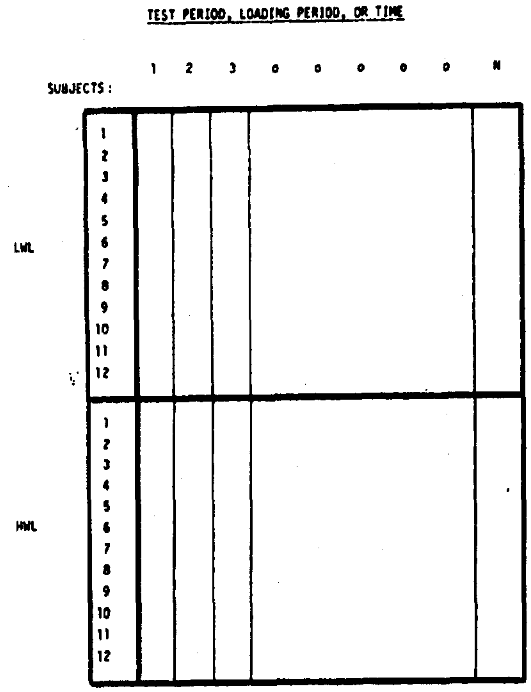


Figure 5. Schematic presentation of the experimental design. The columns indicate test periods, loading periods, or time, depending on the dependent variable being considered.

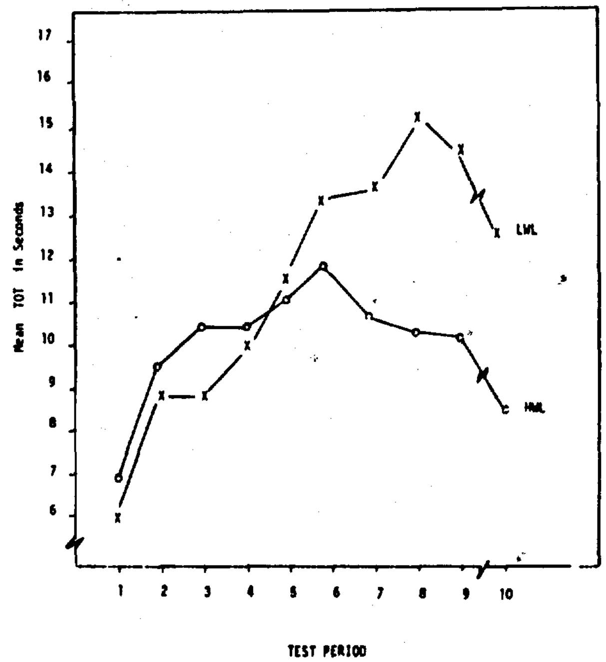


Figure 6. Mean Discrete Tracking TOT in seconds for one-minute test periods as a function of WL and Test Period.

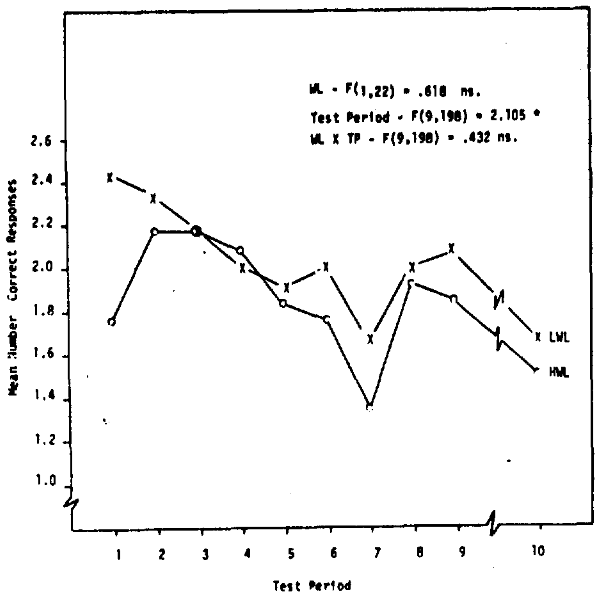


Figure 7. Mean number correct Track to Intercept cognitive problems as a function of WL and Test Period. The WL, Test Period, and WL X TP interaction F ratios are also presented.

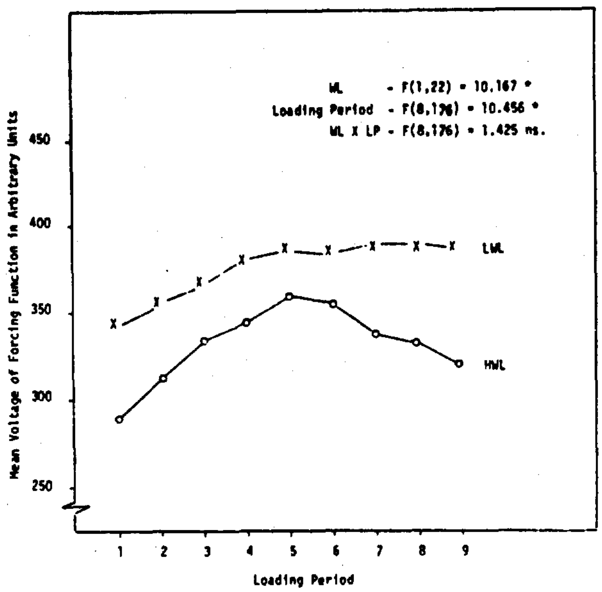


Figure 8. Mean Continuous Tracking Voltages of the Forcing Function as a Function of WL and Loading Period. The F Values are also presented.

TABLE II

SELF-REPORT INVENTORY WITH ITEM MEANS AND STANDARD ERRORS OF MEANS FOR HIGH AND LOW WORK LOAD CONDITIONS

1 = Strongly Disagree
2 = Disagree
3 = Agree
4 = Strongly Agree

ITEM	LOW WL MEAN	LOW WL S.E.	HIGH WL MEAN	HIGH WL S.E.
3. When it was time to eat lunch today, I seemed to have lost my appetite.	1.42	.15	1.33	.19
4. Time seemed to pass very slowly.	2.00	.17	2.17	.17
5. The break for lunch seemed to refresh me.	3.00	.21	3.25	.22
6. Any errors I made today were the fault of other people.	1.42	.15	1.42	.15
8. I did not maintain the same level of performance throughout the day.	2.58	.23	3.00	.28
9. It was difficult to wake up this morning.	2.42	.19	2.33	.22
11. The tasks seemed to be related to flight.	3.17	.11	3.17	.24
13. I felt as though I were pushing myself to complete the tasks.	2.50	.26	2.83	.24
15. I had difficulty falling asleep last night.	1.67	.19	1.92	.23
17. I felt that if I could have just stopped and closed my eyes for ten minutes, I could have continued working indefinitely.	1.58	.15	1.75	.18
18. Now that the day is over, I feel tired.	2.83	.17	2.92	.23
19. I could not have continued working for several hours more.	2.00	.17	2.42	.26
20. I did not get a good night's sleep last night.	2.08	.19	2.17	.24

1 = Never
2 = Sometimes
3 = Often
4 = Constantly

ITEM	LOW WL MEAN	LOW WL S.E.	HIGH WL MEAN	HIGH WL S.E.
21. . . .I became irritated and impatient with other people's errors.	1.33	.14	1.33	.49
22. . . .I could see my performance deviating, but I felt no particular need to attempt to correct the deviation.	1.67	.14	1.92	.51
23. . . .my eyelids felt heavy, and my eyes would begin to close.	1.75	.22	1.83	.72
24. . . .my leg fell asleep.	1.08	.08	1.08	.29
25. . . .my joints became stiff.	1.50	.19	2.00	1.04
27. . . .my responses seemed sluggish.	1.92	.08	2.25	.75
28. . . .the physiological instrumentation interferred with my normal duties.	1.58	.15	1.67	.65
29. . . .it was hard to concentrate on the task I was doing.	1.42	.15	1.92	.79
30. . . .I became sleepy.	2.00	.17	1.92	.51
31. . . .numbers on dials and charts appeared fuzzy.	1.42	.19	1.58	.67
32. . . .I became bored.	1.83	.11	2.42	.90
33. . . .I felt that I wanted to quit.	1.50	.19	1.92	.79
34. . . .my speech became slurred, and I mispronounced words or revised the order of words in a sentence.	1.75	.13	1.42	.67
35. . . .I found myself day-dreaming.	2.25	.13	2.58	.79
36. . . .my responses seemed sluggish.	2.00	.17	2.50	.80
37. . . .I developed a headache.	1.17	.11	1.08	.51
38. . . .I began to lose interest in the tasks and my performance.	1.58	.15	1.67	.65
39. . . .a ringing developed in my ears.	1.00	0.0	1.33	.65

1 = Never
2 = Sometimes
3 = Often
4 = Constantly

ITEM	LOW WL MEAN	LOW WL S.E.	HIGH WL MEAN	HIGH WL S.E.
40. . . .I began to sweat.	1.17	.11	.92	.29
41. . . .I became sleepy.	2.00	.12	1.75	.75
43. . . .the noise became annoying.	1.75	.13	1.33	.65
44. . . .my leg fell asleep.	1.08	.08	.92	.29
45. . . .I rubbed my eyes.	2.00	.21	1.83	.83
46. . . .the instrument harness felt uncomfortable.	2.25	.18	1.67	1.07
47. . . .it became difficult to make decisions quickly.	1.83	.11	2.08	1.00
48. . . .a pain developed in my neck and shoulders.	1.58	.26	2.08	1.00
49. . . .I yawned or stretched.	2.42	.15	2.67	1.07
50. . . .the ear phones became uncomfortable.	1.58	.23	2.92	1.31

Several items offer evidence of the Ss perception of fatigue symptoms. The HWL groups felt they did not maintain the same level of performance over trials, felt more tired at the end of the day, felt less motivated to maintain their performance at a high level, felt that their responses were more sluggish, and felt they had more difficulty making quick decisions than the LWL group.

These and many other items indicate that the HWL group verbalized more fatigue symptoms than the LWL group.

CORRELATES OF FATIGUE

Given performance decrement as a function of continuous work, certain physiological and psychological behaviors may change systematically with the decrement. Several variables were measured to determine their relationship with the performance decrement.

Phoria

Near vertical, near horizontal, far vertical, and far horizontal phoria were measured each test period. Zero phoria on the vertical measures was indicated by a response of five, the center value of a nine-point scale. Zero phoria for horizontal judgments was a response of 8.

One measure was obtained in each of the four phoria tests in each test period. These responses were then transformed to deviations from 5 (or 8 for the horizontal). The algebraic means (use sign) of these deviation scores are shown in Figures 9 through 12. The data offer no suggestion of progressive changes in phoria across Test Periods.

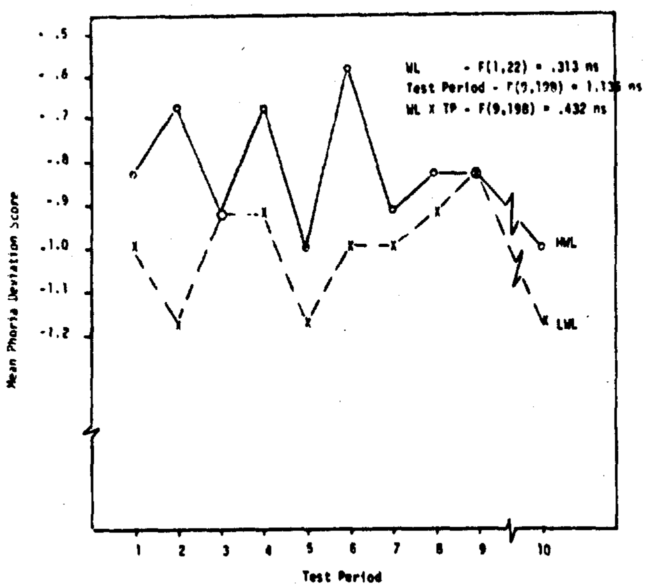


Figure 9 . Near Vertical Phoria means using algebraic deviation from the Zero Phoria score of five for WL and Test Period.

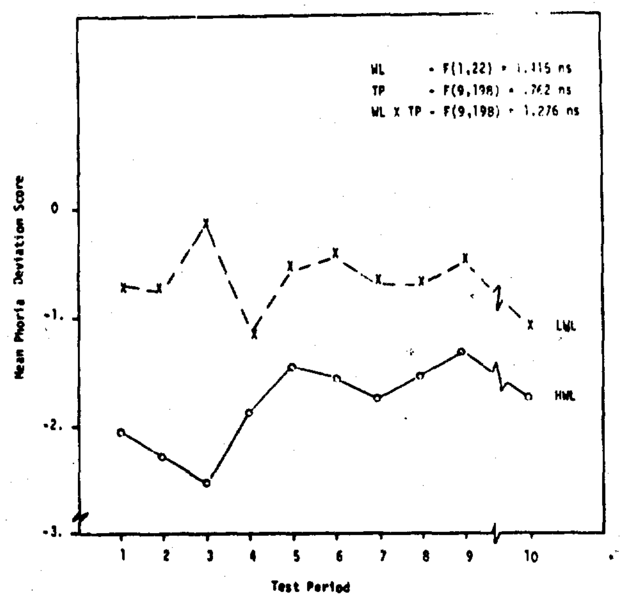


Figure 10 . Mean Near Horizontal Phoria Score using Algebraic Deviations from eight for WL X Test Period.

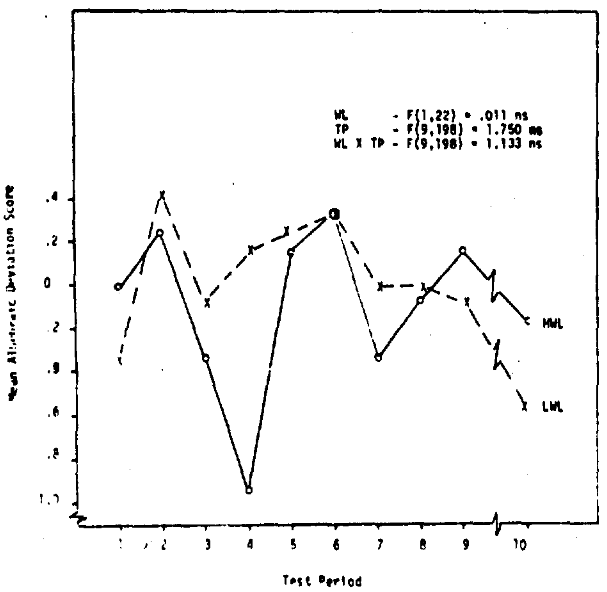


Figure 11 . Mean Far Horizontal Phoria Score for Algebraic Deviations from eight for WL X Test Period.

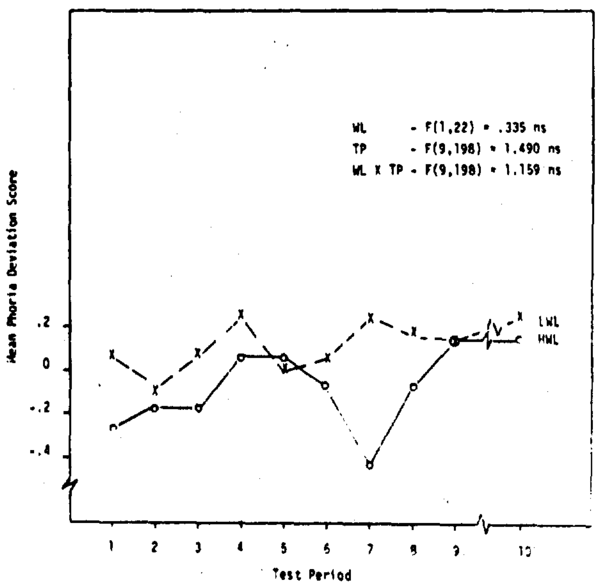


Figure 12 . Mean Far Vertical Phoria Score for Algebraic Deviations from five for WL X Test Period.

Using the covariant measures, the phoria scores were also analyzed with an analysis of covariance. Test scores 6, 7, 8, 9, and 10 were summed and used as the variate. The WL F ratios were .834, .127, .976, and .939 for Near Vertical, Near Horizontal, Far Vertical, and Far Horizontal, respectively, with 1 and 21 df. No significant effects were obtained.

Since tendencies toward positive and negative phoria for later test periods might balance one another, the absolute means (disregard sign) were also analyzed. These means are shown in Figures 13, 14, 15, and 16. There were no significant effects and no meaningful trends.

Again, an analysis of covariance was used to reduce the error variance and statistically equate the WL groups. The WL F ratios were 1.921, .066, .076, and 2.163 for Near Vertical, Near Horizontal, Far Vertical, and Far Horizontal, respectively, all with 1 and 21 df. None of these values are significant.

Time Estimation

During each test period, the Ss made one time estimate. Their responses were recorded to the nearest .1 second and these values were transformed to absolute deviations from times to be estimated. The mean absolute error scores as a function of WL and test period are shown in Figure 17. The HWL groups made slightly greater errors and this tendency increased in the later work periods but these differences were not reliable. The discrete tracking data are also presented for a direct comparison of the two variables. The point of divergence for both functions occurs at approximately the fifth loading period.

An analysis of covariance using two preliminary trials as the covariant and combining Test Periods 6, 7, 8, 9, and 10 as the variate, was significant, F (1,21) = 6.158. This analysis statistically equated the two groups on the basis of preliminary trials so that the significant F represents decrement for the HWL group in Test Periods 6, 7, 8, 9, and 10. The LWL and HWL adjusted means were 5.78 and 11.53 seconds, respectively.

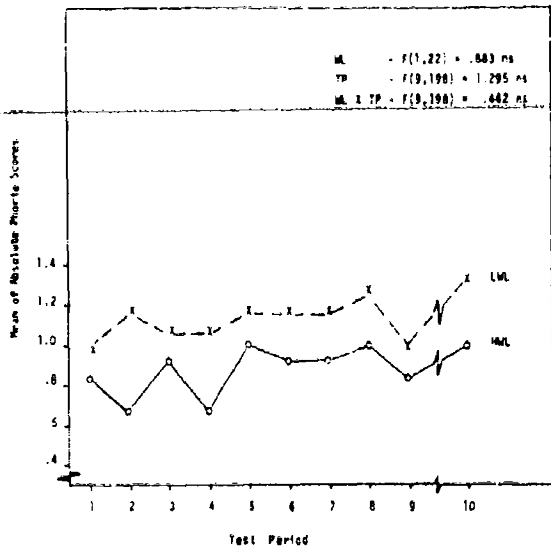


Figure 13. Mean of Absolute Near Vertical Phoria scores as a function of WL and TP.

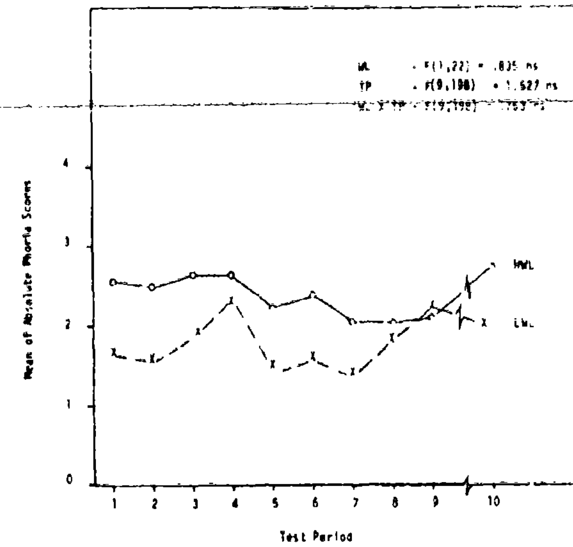


Figure 14. Mean of Absolute Near Horizontal Phoria Scores as a Function of WL and TP.

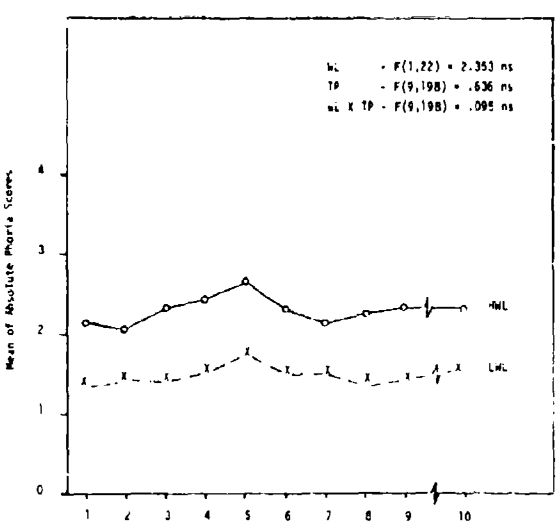


Figure 15. Mean of Absolute Far Horizontal Phoria Scores as a Function of WL and TP.

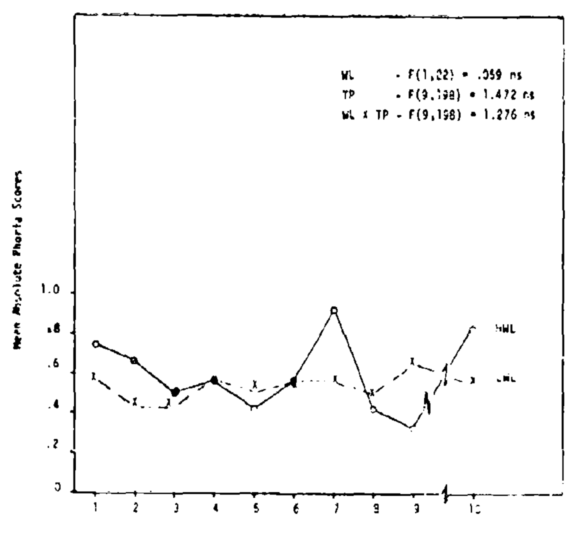


Figure 16. Mean of Absolute Far Vertical Phoria Scores as a Function of WL and TP.

Steadiness

Hand steadiness was measured by the percent of time the stylus touched the aperature during a trial. A low score, therefore, reflects good performance.

The means are shown in Figure 18 as a function WL and Test Period. The mean differences are not reliable. An analysis of Covariance on Trials 6, 7, 8, 9, and 10 combined failed to reach significance, $F(1,21) = .099$. Figure 18 provides a graphical representation of this data.

Size Estimation

The unit of measure for Size Estimation was absolute centimeters error. Ten judgments were made each test period. The means and standard deviations of the absolute error were computed and used as raw scores for the analyses.

Figure 19 shows the average error when the means of test periods were used. Significantly larger errors were made in the later work periods but they were not associated with WL.

The means of the Size Estimation standard deviation score are shown in Figure 20. There were no meaningful trends in these means.

Vigilance

During the Discrete Tracking task a light was presented on some trials. When S detected the light in his peripheral vision he was to extinguish it with a push-button switch. The response measure was RT in seconds and median RTs were computed for blocks of test trials. None of the mean differences shown in Figure 21 are significant.

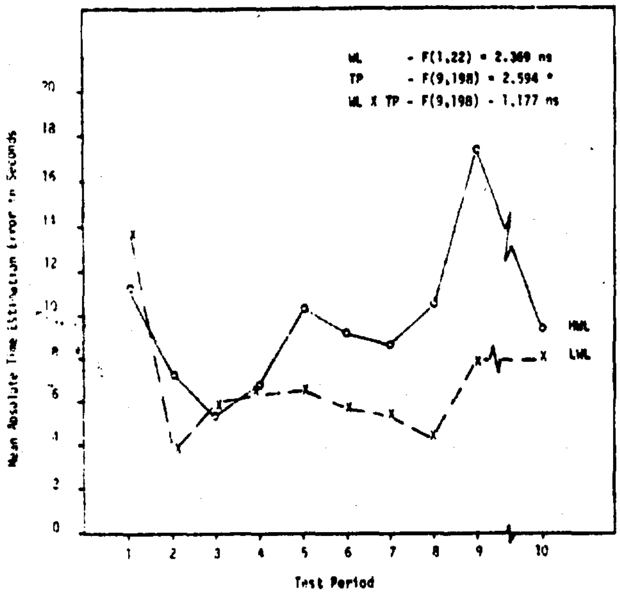


Figure 17. Mean Absolute Time Estimation Error in Seconds as a Function of WL and Test Period.

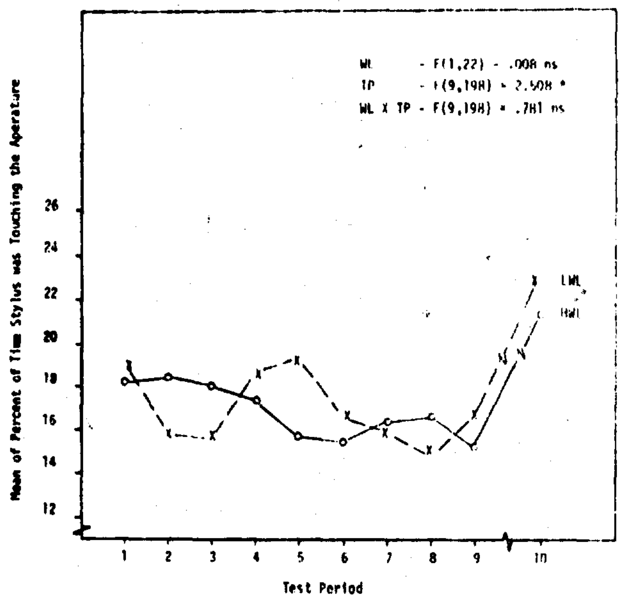


Figure 18. Mean Steadiness Scores. Means are taken on percent of time that the system touched the side of the aperature. Means shown for WL X Test Period.

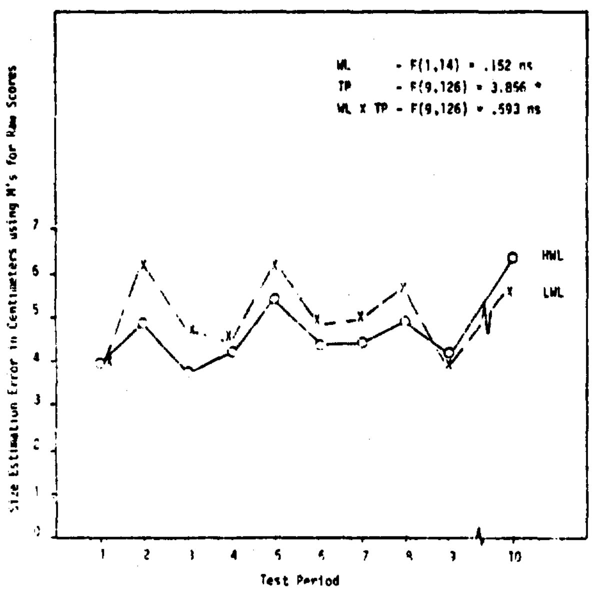


Figure 19. Mean Size Estimation Error in Centimeters for WL X Test Period. The Means of ten measures in a test period were used as raw scores for the analysis.

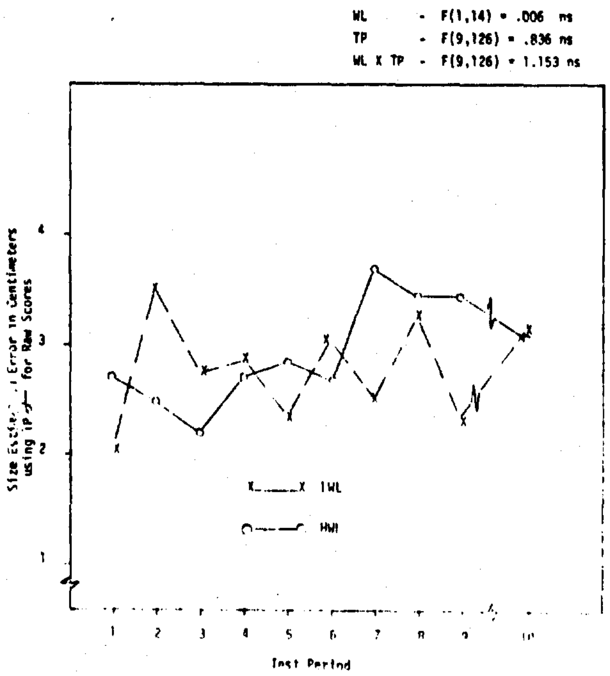


Figure 20. Mean Size Estimation Errors in centimeters for WL X Test Period. The Standard Deviation of ten measures in a test period were used as raw scores for the analysis.

Temperature

Oral temperature was taken four times: 9:00 A.M., 4:00 P.M., 3:00 A.M., prior to sleep, and 7:30 A.M., just after waking. The diurnal cycle was confirmed (See Figure 22), but no other means were reliably different. An Analysis of Covariance, using the 9:00 A.M. values as the covariant and the 7:30 A.M. values as the variate failed to show reliable differences, $F(1,31) = 0$. An analysis of covariance for the 3:00 A.M. values indicated that the HWL had significantly lower temperatures, $F(1,21) = 4.089$. The adjusted means were 97.7 and 97.1 for the LWL and HWL groups, respectively.

Urinary Tests

Specimens of urine were taken after the final test cycle and after a short sleep period. Three separate tests were run, using the combistix technique so that specimens were placed in categories.

Table III gives a frequency count for the protein concentration, glucose, and pH tests as a function of WL and time of measurement. The results are virtually identical for High and Low WL with almost no change from the first to the second measure.

Heart-Rate

Heart-rate was recorded alternately for the two Ss each session in three-minute blocks. Time elapsed during 10 heart-beats was digitally recorded and these times were then converted to Beats-Per-Minute (BPM). Six to ten of these samples were taken to represent each hour and means and estimated standard deviations were computed for each hour. Missing data prohibited the use of 10 samples every hour, but no sample less than six was used.

TABLE III

RESULTS OF URINARY TESTS AFTER THE
FINAL TEST CYCLE AND AFTER A SHORT SLEEP PERIOD
FOR THE HIGH AND LOW WORK LOAD GROUPS

(Combistix Method. All call entries represent the number of subjects in that category)

	After Final Work Cycle					After Sleep				
Protein	Neg.	Tr.	1+	2+	3+	Neg.	Tr.	1+	2+	3+
LWL	12	0	0	0	0	11	0	0	0	0
HWL	11	0	0	0	0	10	0	0	0	0
Glucose	Neg.	Light	Med.	Dark		Neg.	Light	Med.	Dark	
LWL	12	0	0	0		10	2	0	0	
HWL	11	0	0	0		10	0	0	0	
pH	5	6	7	8		5	6	7	8	
LWL	6	5	1	0		8	4	0	0	
HWL	7	4	0	0		6	3	1	0	

Figure 23 shows the means for WL and Time of Day (TD) when the mean BPM for each hour were used as raw scores. The WL groups differ initially and the difference remains essentially constant across TD. Correcting for initial differences the curves are quite similar with peaks corresponding to feeding times.

Figure 24 shows heart-rate variance where SDs were used as raw scores. Again the WL groups differ throughout with essentially no WL X TD interaction.

For the sleep period, heart-rate scores were recorded and treated similar to waking values except that they were not analyzed by hours. Rather, they were combined into a single mean and SD for each S.

When means were used as raw scores, the LWL mean was 58.0 BPM and the HWL mean was 57.0. The F (1,16) = .084 was not significant. An analysis of covariance statistically equating the groups also failed to be significant, F(1,15) = .058.

Using SD scores, the mean SD for the LWL group was 8.45 BPM and 7.44 for the HWL. This difference was not significant, F (1,15) = .970.

Basal Skin Resistance

Basal Skin Resistance (BSR) was taken in ohms resistance and converted to log ohms resistance. One sample was taken each hour and the WL by TD means are plotted in Figure 25. Resistance decreases throughout the day, but there is no differential decrease for the two WLs.

Sleep BSR is shown in Figure 26. Again, resistance decreases across hours, but not differentially.

Respiration Rate

Respiration was recorded in wave form on strip charts. Ten one-minute samples were taken for each hour and a mean was computed. These means were then used as raw scores and they are summarized in Figure 27. RR decreased over time, but there was no differential decrease as a function of WL.

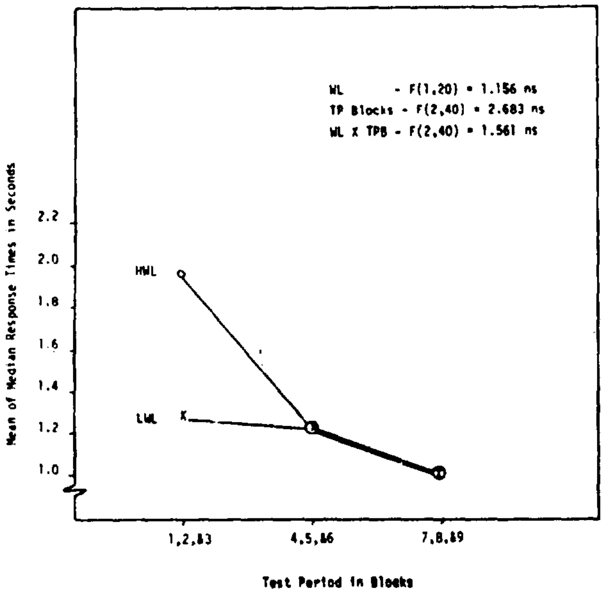


Figure 21. Mean of nine Median Response Times in seconds, as a function of WL and block of Test Periods.

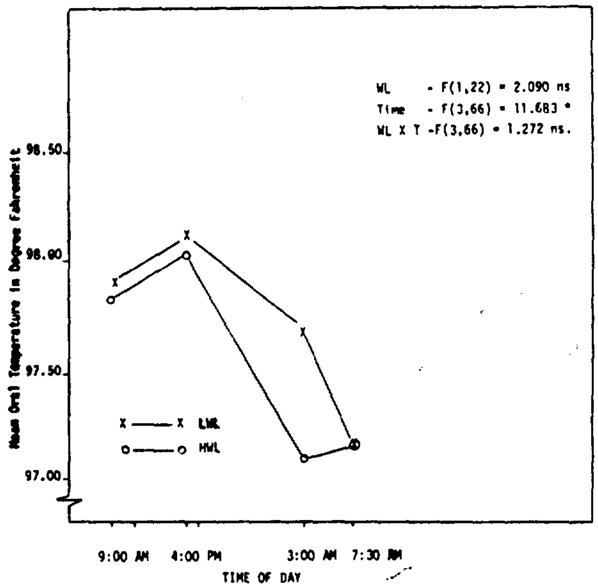


Figure 22. Mean Oral Temperature in degrees Fahrenheit as a function of Time of Day and WL.

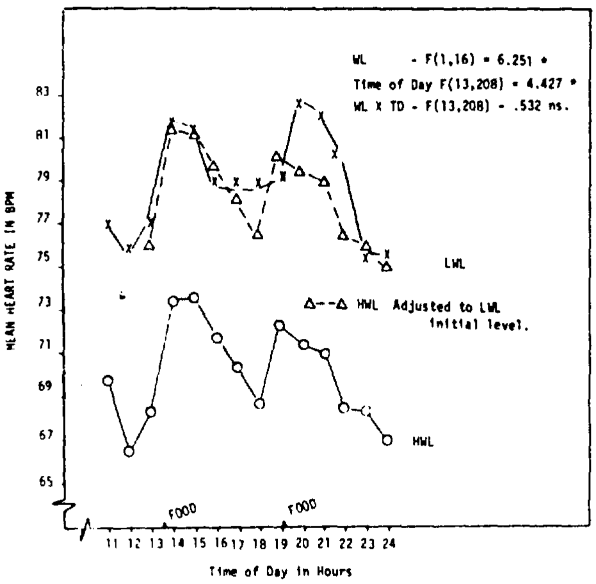


Figure 23. Mean Heart-Rate in BPM as a function of WL and Time of Day. HWL was also adjusted for initial Heart-Rate differences and these adjusted values are plotted.

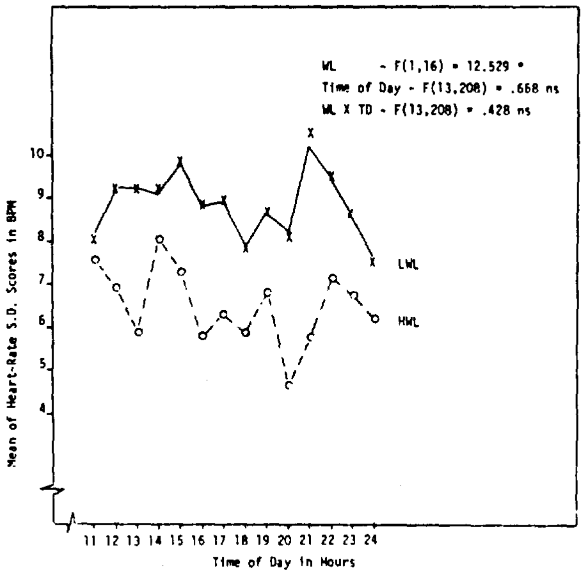


Figure 24. Mean of Heart-Rate Standard Deviations Scores as a Function of WL and Time of Day. Standard Deviations were computed from approximately ten samples for each hour.

Comparison of Discrete Tracking, Continuous Tracking, and Time Estimation

In order to see the relationship among DT, CT, and Time Estimation Error, they are plotted together in Figure 28. The CT HWL function was adjusted for easier comparison. The adjustment is the mean difference between the LWL and HWL on the first two Test-Trial cycles. Comparisons are also given in Figures 29, 30, 31, and 32, for S1, S2, S11, and S12, as examples of individual performance curves.

TEST OF BARTLETT'S HYPOTHESIS

Two variables allowed for a test of Bartlett's Hypothesis, Discrete Tracking, and Size Estimation. The hypothesis predicts increased variance, perceptual break-up, operator distress, and performance decrement - in that order.

Discrete Tracking

To determine whether performance became more variable in the HWL condition in successive test periods, the Standard Deviation for each Test Period for the ten TOT scores each S produced, were used as raw scores.

The means of these SD scores as a function of WL and Test Period are shown in Figure 33. There was a perfect reversal from the prediction. The LWL performance became significantly more variable than the HWL. The F ratios are shown on the figure.

Perceptual break-up would be reflected in the interaction between Horizontal-Vertical and Test Period for the HWL. These means are shown in Figure 34, and it is apparent that the functions are virtually parallel indicating little or no interaction. The ANOVA Summary in Table I confirms this with a WL X H-V X TP interaction F ratio (9,504) = .902.

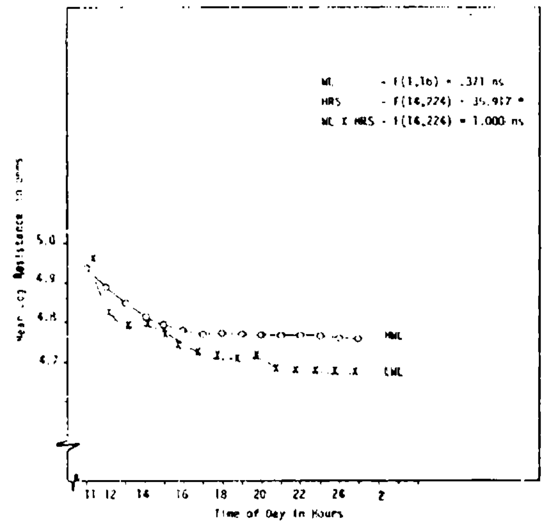


Figure 25. Mean of BSR Log Resistance Scores as a Function of WL and Time of Day.

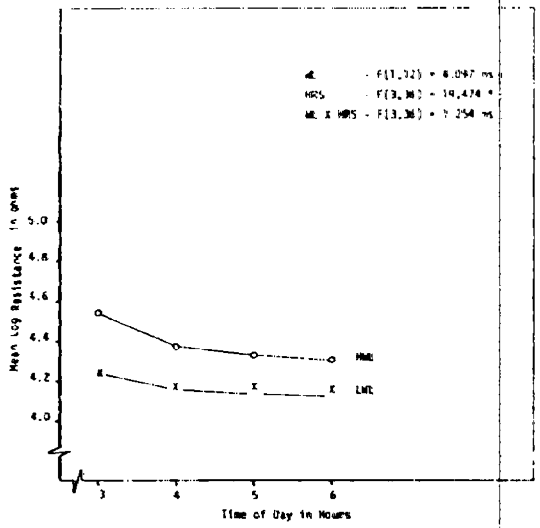


Figure 26. Mean of BSR Log Resistance Scores During Sleep as a Function of WL and Time of Day.

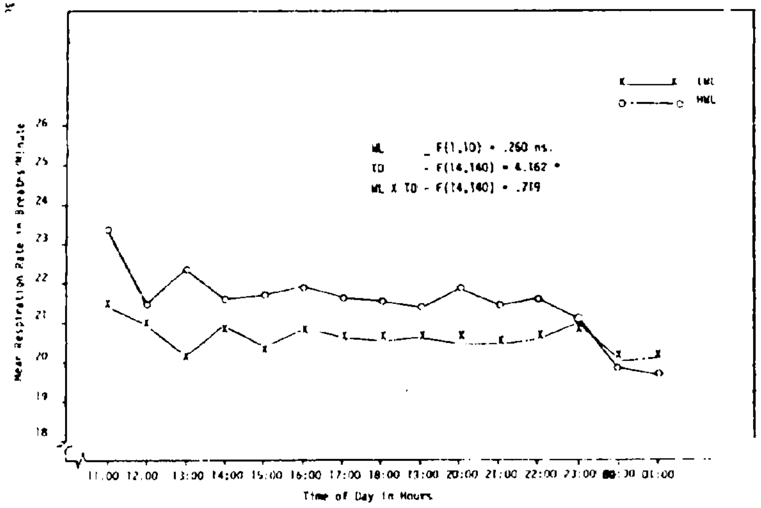


Figure 27. Mean Respiration Rate in BPM as a Function of WL and Time of Day.

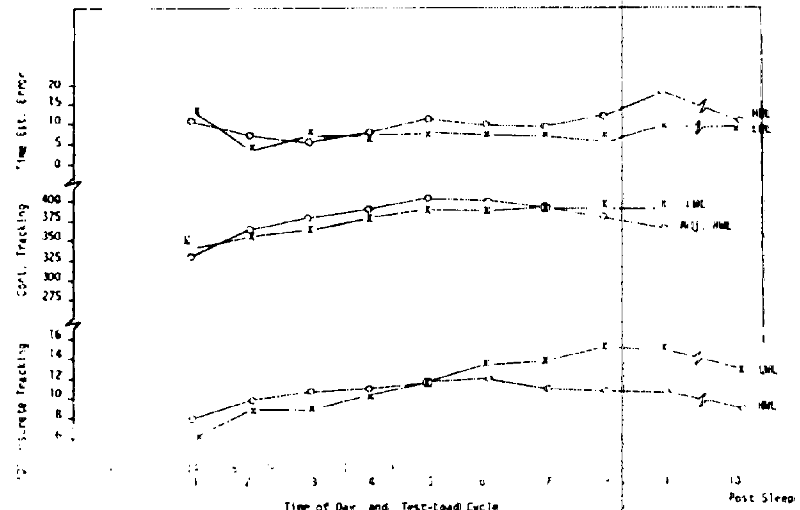


Figure 28. Comparison of Time Estimation Error, Continuous Tracking, and Discrete Tracking. The HWL Function for Continuous Tracking was adjusted by using the average difference between LWL and HWL on the first two cycles.

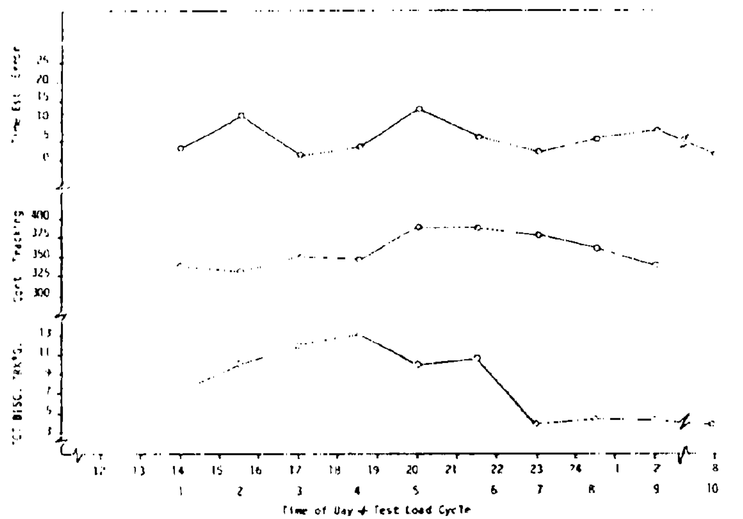


Figure 29. Comparison of Discrete Tracking, Continuous Tracking and Time Estimation for S_1 in the HWL

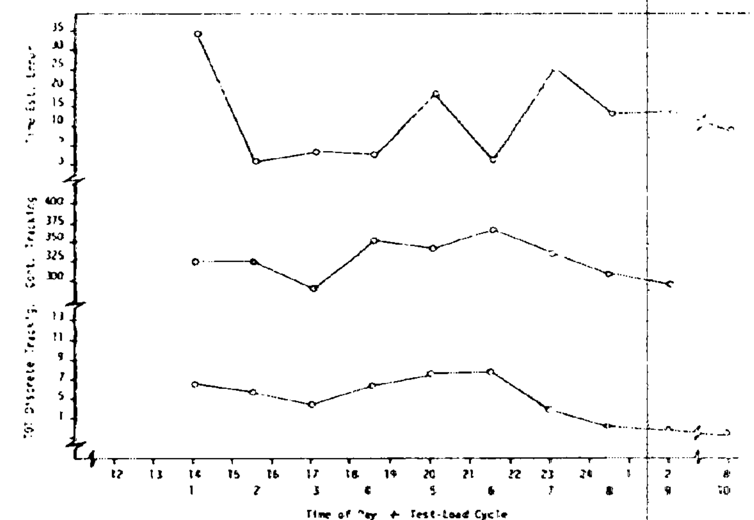


Figure 30. Comparison of Discrete Tracking, Continuous Tracking and Time Estimation Error for S_2 in the HWL.

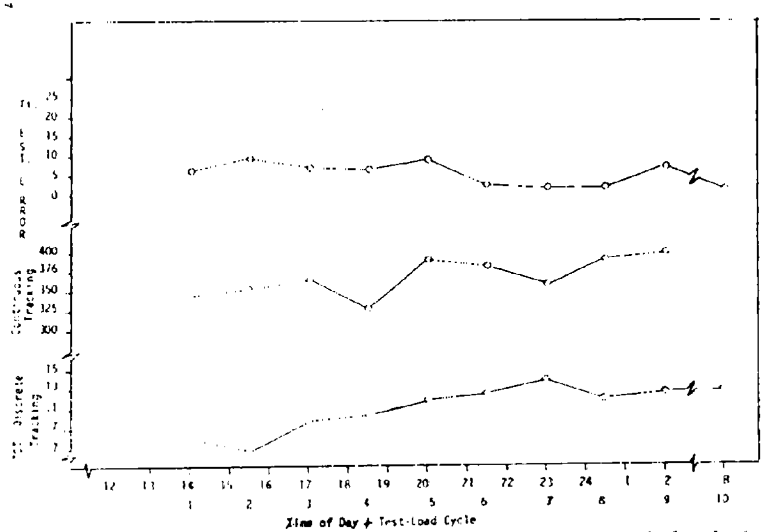


Figure 31. Comparison of Discrete Tracking, Continuous Tracking and Time Estimation Error for S_{11} in the HWL.

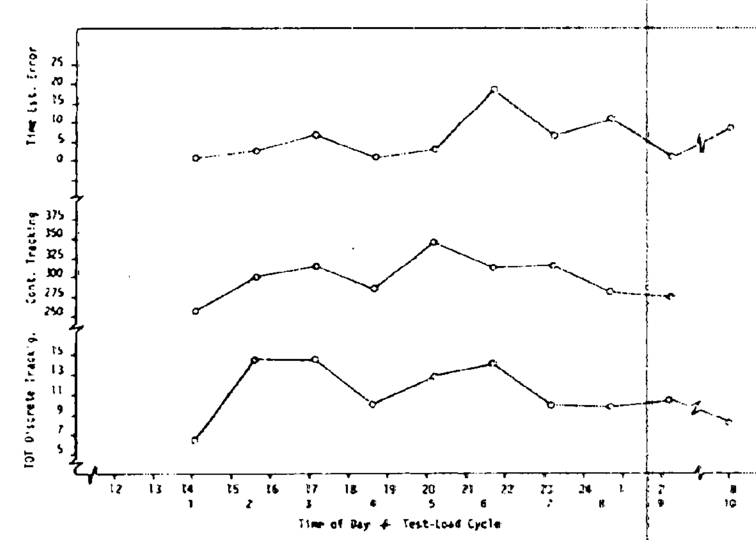


Figure 32. Comparison of Discrete Tracking, Continuous Tracking, and Time Estimation Error for S_{12} in the HWL.

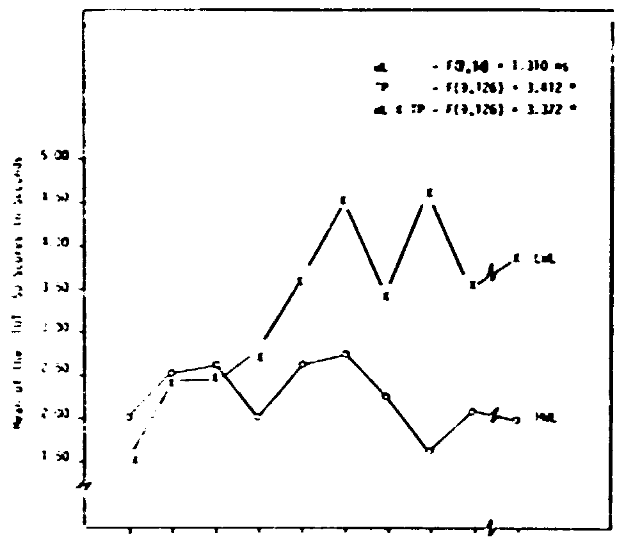


Figure 33. Mean of the Discrete Tracking TOT Standard Deviation scores for Test Periods in seconds as a function of WL and Test Period.

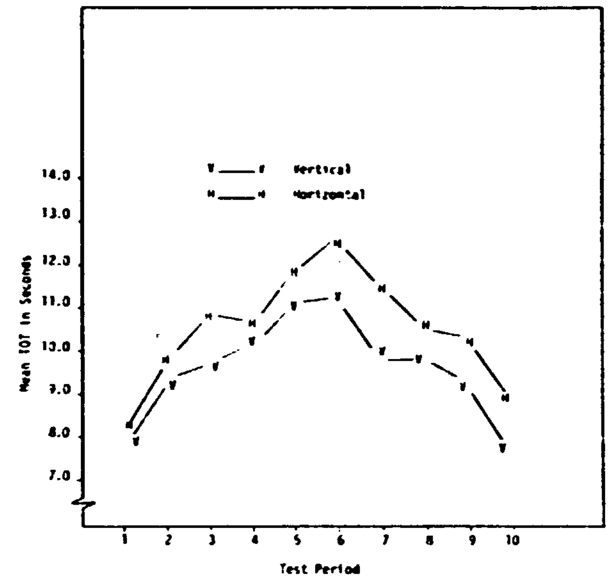


Figure 34. Mean Discrete Tracking TOT Scores in Seconds as a Function of Horizontal-Vertical and Test Period for the HWL Only.

Measures of operator distress were not taken systematically. Notes were made ~~in the log, but this source of data was not sufficient to test this portion~~ of the hypothesis.

Performance decrement for Discrete Tracking was described above.

Size Estimation

The size estimation scores allowed a test of increased variance across work periods. Ten judgments of size estimations were made in each Test Period and the Standard Deviation of the Absolute Error scores was computed. The means of these SD scores were presented in Figure 20. There was a slight trend for variance to increase across Test Periods, slightly more pronounced for the HWL, but the ANOVA failed to confirm the trend.

DISCUSSION

OPERATIONAL DEMONSTRATION OF FATIGUE

There was ample evidence that some fatigue occurred. The effect was best illustrated by the Discrete Tracking task where the Low Work Load (LWL) and High Work Load (HWL) curves began to diverge sharply with the HWL showing performance decrement. One might assume that the onset of fatigue was at the fifth Test Period but this is probably a gross over-simplification. Further testimony of fatigue was produced by Continuous Tracking and the Self Report Inventory.

The Continuous Tracking task was used primarily as a loading task to fatigue the Ss. It was analyzed for symptoms of fatigue under the assumption that a divergence of the curves especially with a decrement for the HWL could best be explained by fatigue. The HWL received twice as much practice on CT as the LWL so that a simple direct comparison was impossible. Also the task was self-paced placing only moderate demands on the S. The findings lend general support to the DT data.

Responses on the Self Report Inventory indicated that the HWL Ss perceived themselves as being more fatigued than the LWL Ss.

By some combined criteria it seems a certainty that the HWL group was more fatigued than the LWL group. But if this is so, how should the cognitive Track-to-Intercept data be interpreted?

Perhaps the most decisive comparison in this experiment was between Discrete Tracking and the Cognitive TTI problems. In pilot work a simplified DT task and the TTI were evaluated. This preliminary investigation indicated that neither task in the form evaluated was likely to be a sensitive index of fatigue.

The difficulty of the DT task was increased so that maximum effort was required just to stay with the light sequence. Even the slightest lapse would cause a complete miss such that the lights were demanding a new response before S could initiate his response to a previous signal. The pace was rapid and the pressure unrelenting. The task was an excellent index of fatigue. The Cognitive TTI problems were of moderate difficulty with virtually no time pressures. Alternative cognitive tasks were considered. Moderate difficulty with temporal pacing seemed to be the need, but it was impossible in the constraints of this investigation (realism and calendar schedule) to effect an alternate task. Indeed, this may have been fortunate since it served to verify and emphasize the preliminary conclusions.

CORRELATION OF FATIGUE

The onset of fatigue is most certainly coincident with the first response. There is probably a brief period after warm-up when the organism is best able to respond. With continuation of the task the organism becomes less capable but this deterioration is masked by the combination of increased effort and tasks that demand less, sometimes far less, than S's maximum effort and skill. Only a task which makes almost excessive demands on skill and effort will immediately reflect fatigue. Other measures may be partially sensitive but they will indicate fatigue only after it has become excessive.

Given the conclusion that the onset of fatigue begins with the first response, specific psychological and physiological variables should reflect fatigue changes. Initially, these variables would show extremely subtle deviations due to the gradual onset of fatigue. With progressive fatigue the correlates would show larger and larger deviations from some nominal value. This entire sequence of changes would be continuously retarded by recuperative mechanisms (Selye, 1950). As fatigue cumulates recuperative processes reduce fatigue and then with rest eliminate it. Correlates of fatigue should reflect this entire cycle. Ideally, the correlates would be simple physiological or psychological behavior for the applied setting.

Discrete Tracking and Time Estimation were the best correlates of fatigue. As the experimental session progressed the HWL performance progressively deteriorated on these tasks. No other potential indices of fatigue showed any promise of predicting fatigue.

The DT task manifested all the attributes of a potentially good fatigue correlate so it was not particularly surprising when DT correlated with WL and produced a significant WL by test period interaction. The behavioral processes that were fatigued in the CT task were presumably necessary for DT degredation. After successive loading periods the CT fatigue was reflected in DT performance. Therefore, you might predict from DT decrement that performance in the less demanding CT task was beginning and would be clearly evident if the work continued. This was indeed the case. The HWL and LWL DT curves begin to diverge at the third or fourth loading-test cycle whereas the CT divergence gradually begins at the sixth or seventh load-test cycle (also, recall that the test measures of a load-test cycle follow the load measures).

Time estimation has been identified as a correlate in other investigations. Carpenter, et al. (1969) found significantly smaller time estimation errors in Low Load conditions relative to High Load conditions, contrary to the present data.

Bartlett (1943), however, found larger time estimation errors in conditions designed to produce high fatigue.

Time estimation can be measured with a simple psychological test. It would be ideal for many applications. Several (5 - 6) successive estimates would be necessary to produce a reliable measure. In most situations these estimates could be concurrent with on-going activities producing minimal interruption.

It could be argued that fatigue was not severe so that there was little opportunity for covariance. While it is true that Ss were not worked to complete debilitation, verbal reports and behavioral measures indicated at least moderate fatigue. This is precisely the level of fatigue that would be important to predict. Lower levels of fatigue would not be significantly detrimental for most tasks and higher levels would be self-evident.

Some variables deserve further consideration even though they failed to be productive in this experiment. Horizontal phoria may be an index of visual fatigue given tasks with greater visual load or poor lighting conditions. Oral temperature differed significantly between LWL and HWL at 3:00 A.M. Temperature is a simple direct measure which has shown promise before (Carpenter, et al.; 1969) and should receive further attention.

The recovery processes were not adequately sampled with the short 3-to-5-hour sleep period. Therefore, variables likely to be associated with recovery processes should receive further exploration. They would be ideal indices of crew readiness.

It now seems clear that special tests must be developed to ensure measurement of fatigue in operational settings. The tests must require traits very similar to those in the operational setting but demand maximum skill with almost excessive time pressures. Approximately 5 - 10 minutes would be required to produce reliable measurement. Such tests should predict performance decrement in typical operational tasks so that missions can be terminated before the impairment actually occurs. Other simple psychological and/or physiological correlates may be effective from time-to-time but should not be considered as primary fatigue indices.

BARTLETT'S HYPOTHESIS

The present investigation did not provide for a good test of Bartlett's Hypothesis (Bartlett, 1953). The perceptual display was not sufficiently complex at any one time to allow for perceptual break-up and only two measures were obtained which would reflect increased variance. Perhaps it would be best simply to conclude that Bartlett's Hypothesis received no support.

The significant reversal in DT variance demands some consideration. Why should performance become more variable for the LWL group and remain relatively stable for the HWL group? Since performance was superior in the LWL group, perhaps this is just one of those cases where means and variance are correlated, contrary to our typical statistical assumptions. Or this may be that "one in twenty" random error effect that we expect when we use the .05 level of significance. There seems to be no plausible theoretical explanation for this outcome.

COMPARISON WITH CARPENTER, ET AL. (1969)

Carpenter, et al. (1969) simulated ASW missions of 2-hour and 8-hour duration. Navy crews flew an S2 simulator on five successive days for one duration period and then after a weekend break flew five more missions of the other duration. The work environment was crowded, noisy, poorly ventilated, and poorly lighted. The tasks closely simulated ASW tasks. The high WL conditions made only moderate demands on the Ss.

The present study used tasks related to ASW but the study should not be considered a simulation of ASW missions. Lighting and ventilation were good. There was ample space with frequent opportunities for the Ss to move about and stretch. Background noise was of moderate intensity. There was a high work load condition extending over 14 hours producing extensive demands on the Ss.

Heart-rate and respiration rate seemed to be indices of fatigue for Carpenter et al. but not in the present study. The change in environmental conditions, the task differences, and/or successive versus one-day test periods could account for the discrepancy between the two studies. It may be that the physiological parameters will be sensitive only when there is extensive work in poor environmental conditions.

Phoria measures may be sensitive with a combination of high visual load and poor lighting. Either adequate lighting or low visual load may invalidate phoria as a fatigue index.

The Ss' self-reports of fatigue symptoms were similar across the two studies with perhaps just the slightest increase in intensity of fatigue symptoms in the present investigation.

The major difference between the experiments was the high correlation of tracking performance and WL in present study with no apparent relationship in the previous experiment. Again, this seems to certify the necessity of a moderately difficult task with extreme temporal pressure. Carpenter, et al. used pilot performance in a standard rate turn, a maneuver with low temporal stress requiring a moderate level of skill. The performance decrement trends for this turn maneuver may reflect fatigue while at the same time indicate low sensitivity for the measure. In contrast, DT in the present study reflected fatigue and proved to be an extremely sensitive index. The DT task demanded an extremely high rate of information processing per unit of time.

CONCLUSIONS

1. The HWL Ss were more fatigued than the LWL Ss.

2. Discrete Tracking was the best performance index of fatigue but other measures reflected fatigue.

3. Time estimation errors reflected the magnitude of fatigue.

4. Bartlett's Hypothesis was not supported; however, it may not have been adequately tested.

5. Indices of fatigue must demand moderate or high level skills with almost excessive time pressures.

6. Simple physiological or psychological measures may indicate fatigue for some applications while being insensitive for others.

REFERENCES

Bartlett, F. C. "Fatigue Following Highly Skilled Work". Proc. roy. Soc Sev. B, 1943, 131, 247-257.

Bartlett, F. C., "Physiological Criteria of Fatigue". Fatigue, (edited by W. F. Floyd and A. T. Welford), (London: Lewis and Company) 1953.

Carpenter, D. L., Creamer, L. R., Gabriel, R. F., and Burrows, A. A., Individual and Crew Fatigue in a Simulated Complex Airborne Weapon System. (Long Beach: Douglas Aircraft). Contract No: N00189-68-C-0565, Technical Report DAC - 67798, 1969.

Fraser, D. C., The Study of Fatigue. R.A.F. Insd. Aviat. Med. August, 1954.

Fraser, D. C., A Study of Fatigue in Aircrew: II. Comparison of the Effects of Day and Night Flying. Flying Personnel Res. Comm. (Gt. Brit) FPRC Rep No. 925, June, 1955.

Fraser, D. C., A Study in Fatigue of Aircrew: IV. Overview of the Problem. Flying Pers. Res. Comm. (Gt. Brit.), FPRC Rep. No. 984, Feb., 1957.

Fraser, D. C. and Samuel, C. D. "Effects on Vigilance", Physiological and Psychological Studies, Part IV. Aircrew Farigue in Long-Range Maritime Reconnaissance. Flying Personnel Research Comm. (Gt. Brit) FPRC Rep No. 907.10, August, 1956.

Grandjean, E. Fitting the Task to the Man - An Ergonomic Approach. (London: Taylor and Francis) 1969.

Harris, W., Mackie, R. R., and Wilson, C. L. Performance Under Stress: A Review and Critique of Recent Studies. ONR, Contract No: NONR 1241(00), Human Factors Research, Inc., July, 1956.

Jackson, E. F. Aircrew Fatigue in Long-Range Maritime Reconnaissance: Pilot Performance. Flying Personnel Research Comm. (Gt. Brit.). Rep. No. 907.2, August, 1956.

Kalsbeek, J. W. H. Measurement of Mental Workload and of Acceptable Load: Possible Applications in Industry. Int'l J. Prod Res., 1968, 7, 33-45.

Selye, H. The Physiology and Pathology of Exposure to Stress. (Montreal: Acta) 1950.

Shaw, W. J. The Effect of Continued Performance in a Task of Air Traffic Control. Med Res. Comm. (Gt. Brit.) APU 205/54. 1954.

APPENDIX A

SUBJECT SCREENING AND BRIEFING

Subject # _____

SUBJECT BRIEFING

1. Purpose: To study behavior in simulated long duration flight operations.

2. The task involved is not manual labor, but more like desk work. It goes on all day and it takes place in a laboratory.

3. You must be present for a full 24 hours. You cannot leave the laboratory during that time.

4. Part of the 24 hours is reserved for sleep, which will come at the end of the same day, after the task has been completed.

5. Sleeping quarters and bedding are within the laboratory. If you require more than 4 hours of sleep, you cannot count on it for this 24 hour period.

6. Toilet and shower facilities are available in the laboratory. You will be asked to provide your own towels. In general, bring with you whatever you normally require for an over-night stay.

7. Food (subjects only) will be supplied, but not at regular meal time. The food will be in the form of snacks and sandwiches.

8. You will be required to wear some electrical connections during both the day and the sleep period. There will be no danger of electrical shock. You will also wear elastic belts around your chest and stomach.

9. You cannot smoke except by permission at planned intervals.

I understand and accept the above conditions.
I understand that if I refuse to complete the full 24 hours of work I cannot be paid.

Signature ______________________

Address ______________________

Phone No. ______________________

My Age in Years _____

Date ______________

GENERAL MEDICAL

	Yes	No
Is your corrected vision normal		
Do you wear glasses		
Do you now have any disease or infection that you can report		
Is your hearing good		
Are you now taking any medicine		
Do you have any skin trouble		
Have you ever had a skin allergy		
Do you have diabetes		
Has any of your family had diabetes		
Have you ever had heart trouble		
Did you ever have rheumatic fever		
Has any one ever said you have epilepsy		
Would you object to giving specimens of your urine during the experiment		
Is there anything important about your health not already asked. If YES, state below.		

GENERAL QUESTIONS

	Yes	No
Can you avoid use of the toilet for as long as 4 hours during the experiment		
Are there foods you are not allowed to eat. IF YES, name them. ________ ________ ________		
Are you willing to go without regular meals during the experiment		
Do you have any unusual food requirement		
Can you arrange your affairs so that there will be no outside interruption during the experiment		
Do you have trouble sleeping when you are away from home		

APPENDIX B

APPARATUS DIAGRAMS

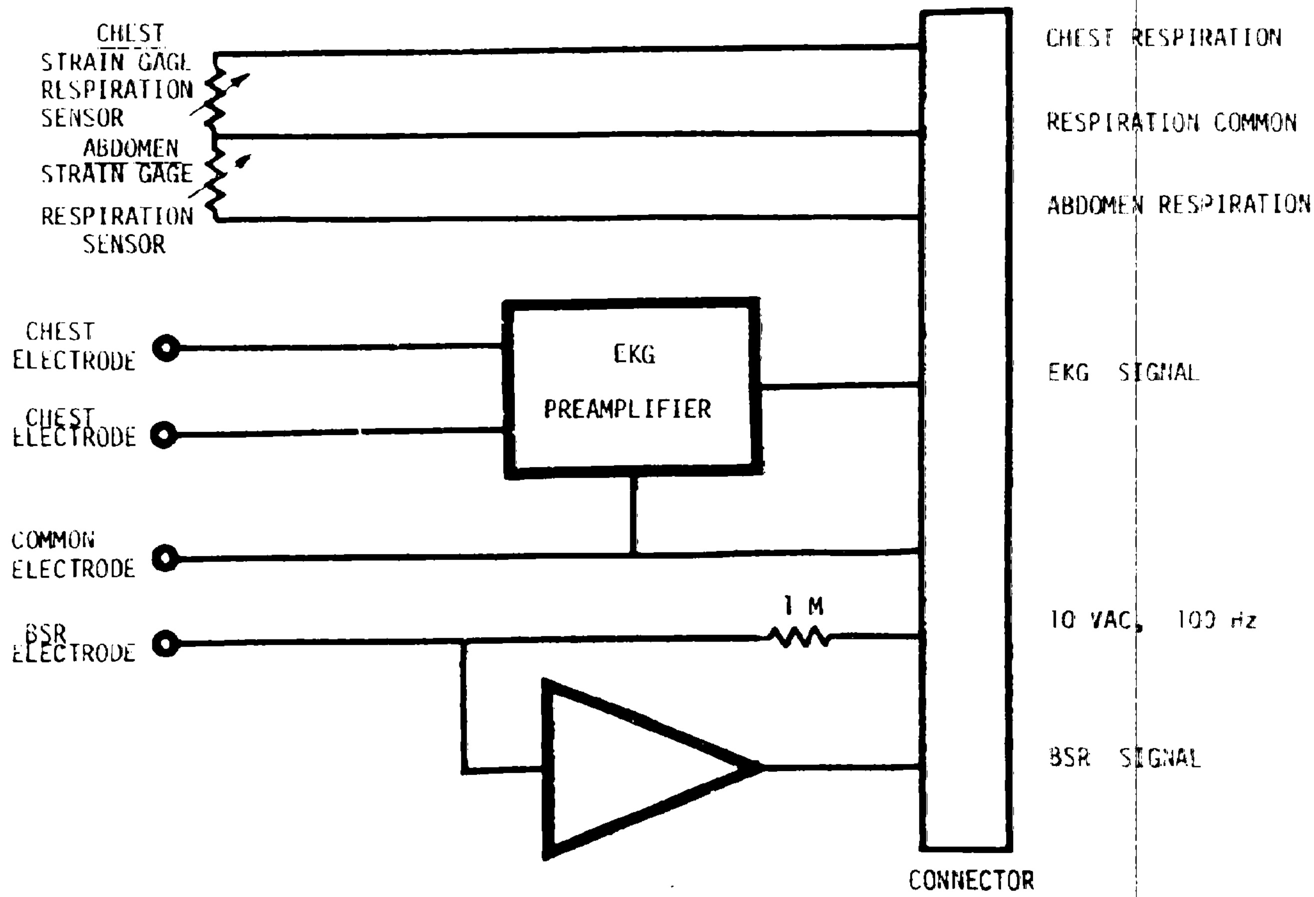


Figure A1 . BIO-ELECTRONIC PACKAGE

III - 53

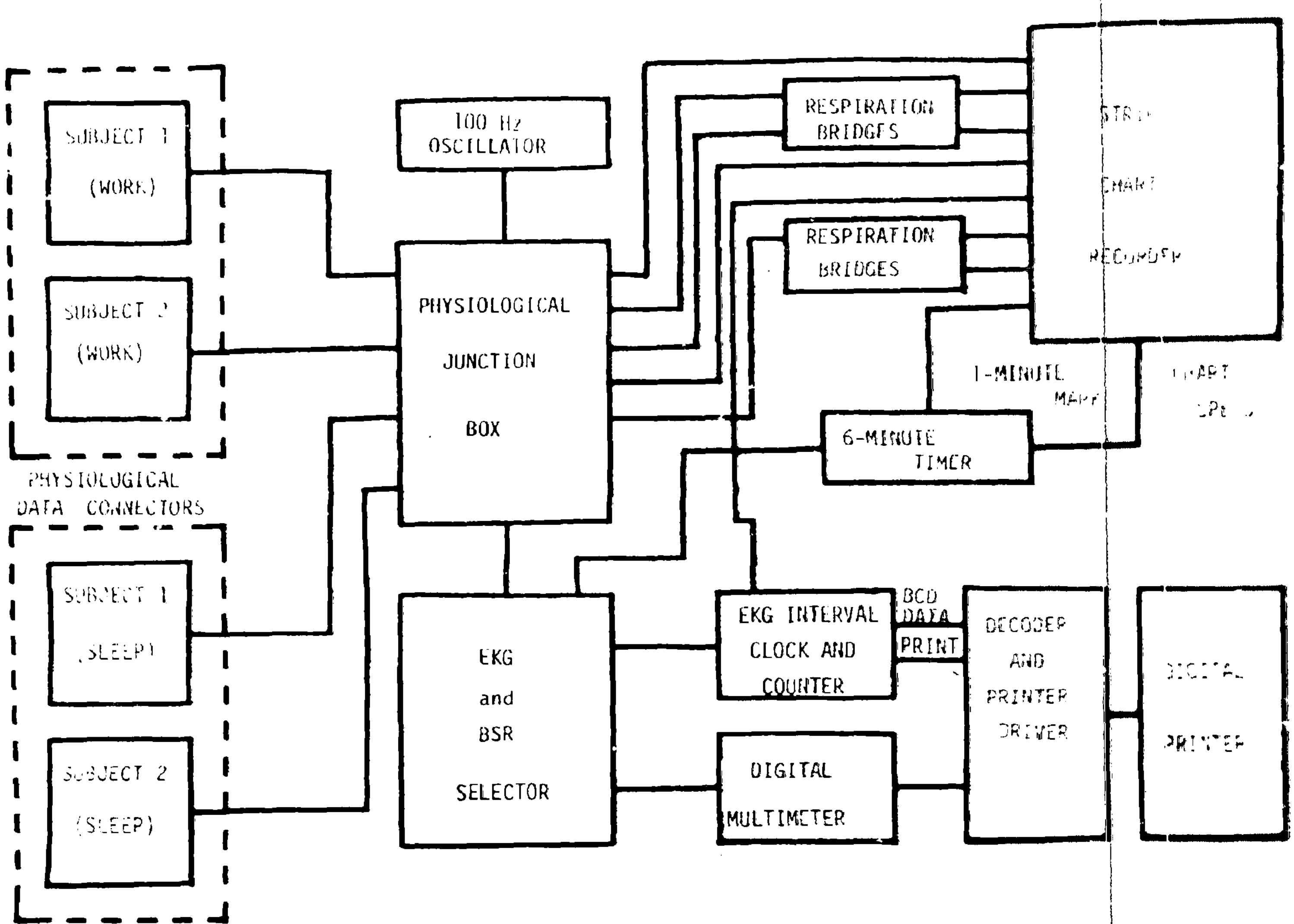


Figure A2. PHYSIOLOGICAL DATA RECORDING SYSTEM

III - 54

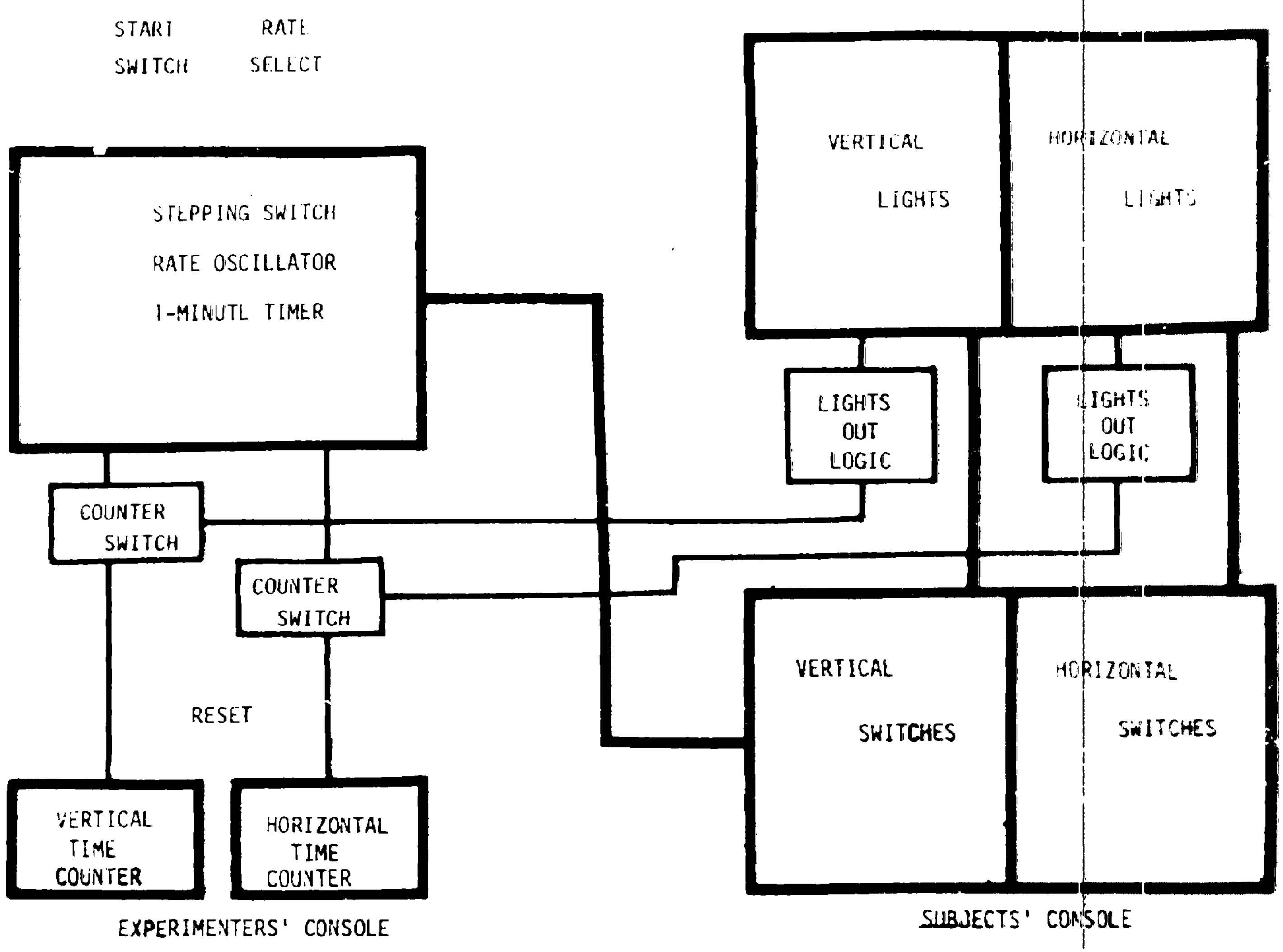


Figure A3. DISCRETE TASKS

III - 55

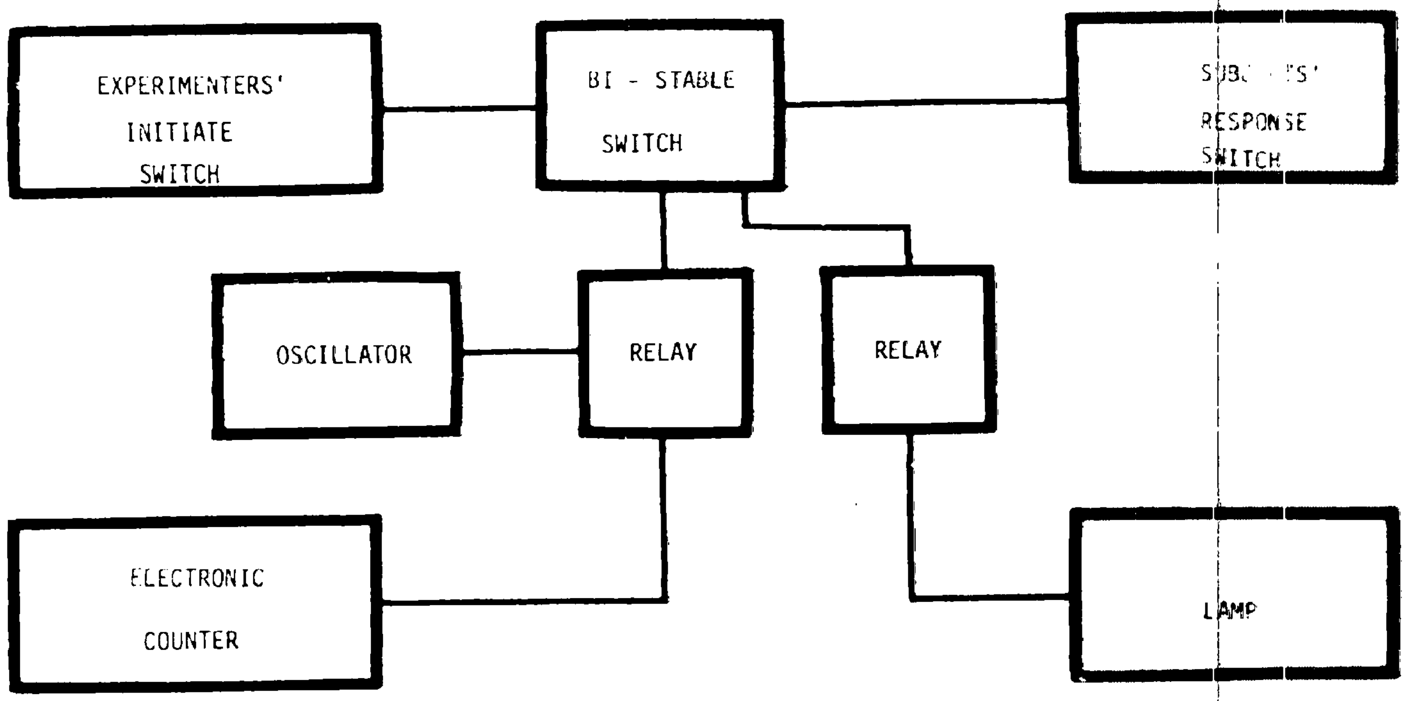


Figure A4 . PERIMETER LIGHT TASK (VIGILANCE)

III - 56

APPENDIX C

SAMPLE COGNITIVE PROBLEMS

PLOTTING

PROBLEMS

Instructions

After plotting the coordinates you are to compute the distance traveled, speed, and compass heading for the following problems.

I. Distance is derived simply by measuring the distance between two points (i.e., between fix 1 and 2). A ruler is provided for this purpose. You will be given further instruction.

II. Speed is derived by dividing distance by time. You get "time" by subtracting the appropriate times given in your problem sheets. You are to convert "time" into seconds. For example, in the following problem you would do as follows:

$$\begin{array}{lcl} t_2 & = & 07 : 09 : 09 \\ t_1 & & \underline{06 : 03 : 04} \\ & & 1 : 6 : 5 \end{array}$$

$$\begin{array}{lcr} 1 \text{ hr.} & = & 3600 \text{ sec.} \\ 6 \text{ min.} & = & 360 \text{ sec.} \\ 5 \text{ sec.} & = & \underline{5 \text{ sec.}} \\ & & 3965 \text{ sec.} \end{array}$$

Therefore, "time" equals $\frac{3965}{3600}$ seconds.

To get speed you simply divide the distance by time. So, if distance equals 46 nautical miles, then speed equals:

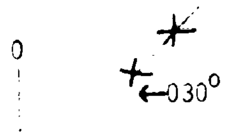
$$= 46 \div \frac{3965}{3600}$$

$$= 46 \times \frac{3600}{3965}$$

$$= \underline{\underline{40}}$$

Speed is equal to 40 knots.

III. Compass heading is obtained by plotting the coordinates provided, drawing a line between the points, and measuring the heading with a protractor.



Sample Problem and Its Solution

Given:

1) lat. 32^{o} 11' N, long. 153^{o} 13' E (t = 01:39:31)

2) lat. 32^{o} 15' N, long. 153^{o} 31' E (t = 02:30:11)

Questions:

A. What is the distance from fix 1 to fix 2?

B. What is the speed of the sub if it is traveling from fix 1 to fix 2?

C. What is the compass heading?

Solution

A. Distance equals 19 nautical miles. This was derived simply by measuring the distance between the two fixes.

B. Speed

(1)
```
  01 : 89 : 71
 -01 : 39 : 31
       50   40
```

(2)
```
   50        3000
   60          40
 3000        3040
```

(3) $19 \div \frac{3040}{3600} = 19 \times \frac{3600}{3040} = \frac{68400}{3040}$

(4) $\frac{68400}{3040} = 22.50$

Speed = 22.50 knots

C. You will be shown how compass headings are read.

Note:

1) t = time

2) A "fix" is a point that you have plotted.

ESTIMATED TIME OF ARRIVAL

PROBLEMS

All problems that involve air speed:

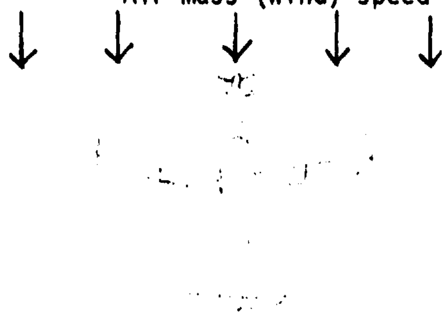
1. The effect of altitude

   Light plane airspeed indicators are calibrated for standard sea level operation (59° F -29.92" Hg). As you go to higher altitudes, the air will be less dense and therefore there will be less drag. The plane will move faster, but because of the lesser density, the airspeed will register less than your actual speed. This error can be corrected roughly by the rule of thumb, "add 2 per cent per thousand feet". This means that if your "calibrated" airspeed is 100 mph at 5,000 feet, your "true" airspeed is 5 x 2 = +10 per cent = (.10) (100) = 10mph + 100mph = 110mph.

2. The effect of wind

   Ground speed is like walking on an escalator. If you're walking <u>against</u> the escalator at its exact speed, you make no progress. By walking with the escalator your speed is added to its speed.

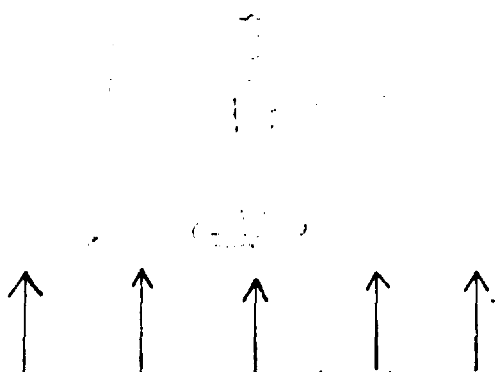
Air mass (wind) speed -20 mph



calibrated speed = 100 mph

Ground speed = 100 - 20 = 80 mph

calibrated = 100 mph



Air mass (wind) speed - 20 mph

Ground speed = 100 + 20 = 120 mph

EXAMPLES:

1. Altitude

At what altitude must a plane be flying in order for it to take 3 hours to fly 450 nautical miles if it had a 25 knot tail wind and was flying at 95 knots?

Solution:

a. air speed = 95 + 25 120 knots (speed plus wind factor)

b. land speed = 450 ÷ 3 = 150 knots (distance to be flown divided by time)

c. correction factor per thousand feet = .02 x 120 = 2.40 (correction factor times air speed)

d. 150 - 120 = 30 (land speed subtracted by air speed)

e. altitude = 30 ÷ 2.40 = 12,500 feet (difference between land speed and air speed divided by correction factor)

2. Time

Flying at 6,000 feet, with a 20 knot tail wind, at 100 k how long will it take to go 500 nautical miles?

Solution:

a. 2 x 6 = .12 x 100 = 112K (correction factor for altitude)

b. 112 + 20 = 132K = (correction factor for altitude)

c. time = 500 ÷ 132 = 3.79 (distance divided by speed)

3. Distance

A plane is traveling at 100 knots at 5,000 feet, with a head wind of 20 knots. How many nautical miles can it travel in 2 hours?

Solution:

a. 5 x 2 = .10 x 100 = 100 + 10 = 110 (correction factor for altitude)

b. 110 + 20 = 130 (correction factor for wind)

c. 130 x 2 = 260 nautical miles (distance that can be flown)

FUEL CONSUMPTION

INSTRUCTIONS

To determine the amount of fuel consumed, follow these steps:

1. Convert speed into true air speed as you did in the estimated time of arrival problems. Remember: you take 2% of every one thousand feet the plane is into the atmosphere. Then you multiply this amount by the speed, and add the amount to the speed. Finally, you add or subtract the wind factor. This value is your true air speed. (See instructions for ETA problems.)

2. Next you divide the distance flown by the true air speed to get the duration of the flight.

3. Finally, you multiply the value for flight duration by the value for the number of gallons consumed per hour. This latter value can be found by looking at Table 1 and finding the power rating given in the problem. You will use the accompanying number for the multiplication.

4. The product is your answer.

TABLE 1

% of Power Rating	Gallons of Fuel Consumed Per/Hr.
66%	12.8
90%	19.1
60%	12.0
50%	9.9
96%	22.6
45%	8.9

T.T.I.

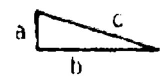
PROBLEMS

Formulas for T.T.I. Problems

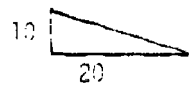
1. Time = $\frac{\text{distance}}{\text{speed}}$ ex: $t = \frac{20 \text{ n.m.}}{10 \text{ k.}} = 2$ hrs.

2. ~~Speed = $\frac{\text{distance}}{\text{time}}$ ex: $s = \frac{15 \text{ n.m.}}{3 \text{ hrs.}} = 5$ k.~~

3. The value of the hypotenuse is equal to the square root of the sum of the squarred legs of the triangle. That is, in the triangle c (the hypotenuse) equals $\sqrt{a^2 + b^2}$.



Example:



$c = \sqrt{10^2 + 20^2}$

$c = \sqrt{100 + 400}$

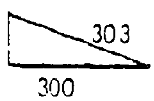
$c = \sqrt{500}$

$c = 22$

4. If the value of one leg and the hypotenuse are known, the value of the other leg may be found by: 1. squaring the value of the known leg and squaring the hypotenuse; 2. subtracting the squared value of the leg from the value of the squared hypotenuse; 3. Then taking the square root of the obtained value. That is

$a = \sqrt{c^2 - b^2}$ or $b = \sqrt{c^2 - a^2}$.

Example:



$a = \sqrt{303^2 - 300^2}$

$a = \sqrt{91809 - 90{,}000}$

$a = \sqrt{1809}$

$a = 43$

Formulas for T.T.I. Problems (continued)

For the first two parts (A and B) of the question you will use the information listed under "given". For the third part (C) you will use the information provided after the question.

d = distance

t = time

APPENDIX D

SPECIFIC TASK INSTRUCTIONS

SPECIFIC TASK INSTRUCTIONS

Instructions: Continuous Tracking

This is a continuous tracking task. It is your job to keep the green pip in the exact center of the screen, which is designated by the vertical line. The pip will move randomly but by moving the "joy-stick" in the direction you want the pip to go, you will be able to control its fluctuations. You are to use only one hand while manipulating the stick. Use the same hand for all trials.

The closer the pip is to the center and the longer you keep it that way, the more credit you will be given.

Instructions: Ortho-rater (Phoria)

You are to look at four targets, two at close range and two at a more distant range. The two displays seen at close range will be the same as the two seen at the more distant range.

You adjust for the two closer displays first by pulling toward you where it says "ortho-rater". Then you move the adjustor on the right side of the ortho-rater so that it is spaced between the 1 and 2. This is the setting for the vertical display. After you adjust to this setting, by looking into the lens you will see that there is a step-like ladder with a number designating each stap (range, 1 to 9). Also there is a discrete horizontal line that is formed by a number of red dots. Your task is to determine the step level that the horizontal line runs across. You record the number that corresponds between the numbers 2 and 3. This is the setting for the horizontal display. After you adjust to this setting, by looking into the lens you will see that there are some horizontally-spaced white dots with a number under each of them (range, 1 to 15). Laterally, there is an arrow with black dots within it. You are to determine where the arrow intercepts a dot. You are to record on your data sheet the number that is below the dot that you think the arrow intercepts.

After finishing that, move the adjustor back to its previous setting between 1 and 2. Then push forward where it says "ortho-rater" and now you are ready to work with the more distant display.

First, move the adjustor farthest from you (on the right side of the ortho-rater) to the number 5. This is the setting for the vertical display. You then follow the same procedure as in the vertical display task above.

Next, move the farthest adjustor to the number 6. This is the setting for the horizontal display. Then follow the same procedure as in the horizontal display task above.

Finally, move the setting back to the number 5, then go on to the next station (i.e., the Size Constancy station).

Instructions: Combined Hand Steadiness and Time Perception

In this task, you will do two things. You are required to make an estimate of a prescribed time interval; also, during that estimation period you are to keep the stylus from touching the sides of the circle, as well as the back-up plate.

Follow carefully these procedures:

1. Select the appropriate time interval to be estimated.
2. To reset the timer to zero, push the reset buttons in the bottom right corner of the bottom apparatus.
3. To reset the numbers on the top apparatus to zero, push the reset button in the bottom right corner of the top apparatus.

4. With your dominant arm, take the stylus into your hand with your palms faced down. You cannot hold the stylus beyond the plastic covering. Do not place your arm on the table while doing the task.

5. Take the plastic box into your free hand and place a finger on the switch. This operates both timers.
6. Make sure you are standing so that you cannot see the numbers of the two apparatus.
7. When you are ready, extend your arm and keep the stylus from touching the sides. Then flip on the switch on the plastic box. Be sure to keep the stylus from touching the sides and back plate as much as possible. The more you touch these, the poorer one of your scores will be.
8. As soon as you turn on the switch, begin counting aloud backwards - from 100 by threes. When you have judged that the prescribed interval has elapsed, switch off the plastic box. You can then release the stylus.
9. Look around and record in the appropriate spaces on your data sheet the actual elapse time (from bottom apparatus) and the time touching the plate (from top apparatus).

Reset both series of numbers and move on to the next station.

Instructions: Size Estimation

In this task, you are to estimate line length.

First, move the black switch from the "OFF" position up two notches to where it says "LOW". This turns on the projector and projects a black line onto the screen before you. You will notice a scale to the right of the projector with numbers on it, ranging from 1 to 52. This scale is in centimeters. Your task is to determine the length of the projected line in centimeters, using this scale. You are to estimate the line to the nearest centimeter. When you have done this, record on your data sheet the number which corresponds to the estimated line length in centimeters.

On your data sheet you will notice that there are 10 spaces in each row. During this task you will estimate 10 lines, using the procedure just described. There will be only one line on each slide. Therefore, you go through a series of 10 slides. To do this, you must push the grey button marked "FOR" on the black box beside the projector. By pushing the button once you will advance forward one slide. By holding the button down you can advance forward more than one slide. If you want to go back and look at a slide, push the button marked "REV". This button reverses the slide advancement. The longer you hold this button down, the farther back in the series of slides you will go.

On your data sheet there is a column, "slide number to start on". You will always start on that number. On the slide carousel there are numbers ranging from 0 to 80. Using the start number on your data sheet, you will line the corresponding number on carousel up with the notch on the side of the projector. To move the carousel, push and hold the grey button marked "SELECT" on the side of the projector down. While this button is down, you can move the carousel to any number that you want.

After you have completed this task (estimated the length of 10 lines), turn off the projector. To do this, move the black switch on the projector from the "LOW" position to the "OFF" position, then go on to the next station (Limb Steadiness and Time Perception).

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Douglas Aircraft Company McDonnell Douglas Corporation	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE

~~Human Error Research and Analysis Program, Summary and Final Report~~

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Summarizes work of four contracts from August 1965 to April 1970

5. AUTHOR(S) (First name, middle initial, last name)
Lyle R. Creamer, Douglas E. Wheeler, Richard F. Gabriel, John S. Tarr, Jack A. Baldwin, Hojat J. Rostami

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1970	150	36

8a. CONTRACT OR GRANT NO. N00189-70-C-1321	9a. ORIGINATOR'S REPORT NUMBER(S) MDC J0698/01
b. PROJECT NO. N00188-8290-2131	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT
All distribution of this report is controlled. Qualified DDC users shall request through U.S. Naval Safety Center, U.S. Naval Air Station, Norfolk, Virginia.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY U.S. Naval Safety Center U.S. Naval Air Station Norfolk, Virginia

13. ABSTRACT

Because of the high ratio of human error induced to total Naval aviation accidents the U.S. Naval Safety Center, Norfolk, Virginia, in 1965 funded the first of four contracts to study "human error" under the heading Human Error Research and Analysis Program (HERAP).

Data analysts have developed a data bank for records of 17,000 Navy pilots (personal data, training records and individual flight time) and accident data. Statistical techniques were applied to the data in an attempt to find predictive measures for human error accidents. Some success was achieved.

Two studies were conducted to provide additional insight into the recognition, measurement, prediction and alleviation of fatigue when defined as "The degradation in performance or affective state, resulting from previous work." In the first study two-man crews were monitored continuously for two weeks during high and low work periods in an ASW S-2 simulator. Some indicators were found but longer work periods were needed to insure fatigue.

In the second study two-man crews were run through 24 hour cycles of high and low work loads of up to 18 hours duration. Performance decrement was obtained and one measure was sensitive to fatigue.

DD FORM 1 NOV 65 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Naval Aviation						
Safety						
Human Error						
Fatigue						
Data Bank						
Accident Analysis						
HERAP						

Unclassified

Security Classification