REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1992	Final Jul 89 - Jul 91

4. TITLE AND SUBTITLE

Comparative Effects of Antihistamines on Aircrew Mission Effectiveness under Sustained Operations

5. FUNDING NUMBERS

C - F33615-87-D-0601
PE - 63764A
PR - 7930
TA - 19
WU - 17

6. AUTHOR(S)

Douglas R. Eddy
Mathieu A. Dalrymple
Samuel G. Schiflett

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Systems Research Laboratories, Inc.
A Division of Arvin/Calspan
2800 Indian Ripple Road
Dayton, OH 45440-3696

NTI, Inc.
4130 Linden Avenue
Suite 235
Dayton, OH 45432

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

US Army Medical R&D Command
HQ USAMRDC SGRD-PLE
Fort Detrick, MD 21701-5012

Armstrong Laboratory
Crew Systems Directorate
Brooks AFB, TX 78235-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AL-TR-1992-0018

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Dr. Samuel G. Schiflett (512) 536-3464

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Seldane (terfenadine) and Benadryl (diphenhydramine) were tested against a placebo in six Command, Control, and Communications (C^3) simulations involving a team of three AWACS weapons directors. Three of the simulations were high difficulty, three were low. An eclectic, hierarchical performance measurement system was developed to address several research questions. The results of these variables on mission effectiveness, the highest level of the performance hierarchy, are addressed in this report from the point of view of a mission commander.

High difficulty scenarios degraded performance compared to low difficulty scenarios, but antihistamine treatments had no statistical effects on the team performance measures. Trends in the data and subjective ratings of a subject matter expert, however, indicated that the Benadryl groups were poorer performers. Some learning was evident across the three days of testing. Future data analyses are discussed.

14. SUBJECT TERMS

Seldane (Terfenadine); Benadryl (diphenhydramine); Airborne Warning and Control System (AWACS); simulations; antihistamine drug effects.

15. NUMBER OF PAGES

56

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number		Page

LIST OF TABLES

PREFACE

A primary mission of the Sustained Operations Branch, Crew Technology Division of the Armstrong Laboratory, formerly the USAF School of Aerospace Medicine (USAFSAM), is to develop procedures and provide guidance to operational commands on maintaining and extending crew performance during sustained operations and continuous duty.

The USAFSAM developed the Aircrew Evaluation Sustained Operations Performance (AESOP) facility under the sponsorship of the Office of Military Performance Assessment Technology (OMPAT), formerly the Chemical Defense Joint Working Group on Drug Dependent Degradation of Military Performance (JWGD^3 MILPERF), to meet the triservice research and mission requirements for team performance metrics. Continuous technical guidance was received from OMPAT during the development of the AESOP facility. Dr. Frederick Hegge, OMPAT's director, was especially helpful. Partial funding was provided by Army Medical Research and Development Command.

Scientists at the AESOP facility conducted the study, Comparative Effects of Antihistamines on Aircrews under Sustained Operations, to evaluate the interactive effects of medications, as well as workload, fatigue, and stress on Airborne Warning and Control System (AWACS) aircrew performance. We acknowledge the assistance of the Tactical Air Command and the 28th Air Division in preparing for the study. Special thanks are due to personnel at Tinker Air Force Base including the 963d, 964th, and 965th AWAC Squadrons (assigned to the 552d AWAC Wing) for providing 36 AWACS Weapons Director volunteers to participate in the weeklong scenarios. We gratefully acknowledge the contribution of Merrell Dow Pharmaceuticals, Inc. in providing the medications for the study. Thanks are due as well to Joseph R. Fischer, Jr. and Carolyn Oakley (AL/CFTO) for their roles in data analysis and to Janet Trueblood (Systems Research Laboratories) for editing and final copy preparation.

COMPARATIVE EFFECTS OF ANTIHISTAMINES ON AIRCREW MISSION EFFECTIVENESS UNDER SUSTAINED OPERATIONS

RATIONALE

The Office of Military Performance Assessment Technology (OMPAT)[1] has attempted to determine the impact of certain classes of drugs and medications on the performance of aircrews solving a range of mission-related tasks in stressful environments. One area of interest involves the effects of antihistamines on complex Command, Control, and Communications (C^3) decision-making performance by Weapons Director (WD) teams during sustained operations. Because of the drowsiness side effects, United States Air Force (USAF) flight surgeons ground aircrew personnel who are taking centrally acting antihistamines, such as Benadryl, for seasonal allergies or nonallergic rhinitis symptoms. The common use of over-the-counter antihistamines results in frequent interruption of flying schedules, loss of training, and disruption of crew rest schedules for nonsymptomatic crew members, especially during sustained operations. However, several antihistamines purporting to have no drowsiness side effects have now become available to USAF flight surgeons. A triservice committee for the OMPAT chose a nonsedating antihistamine, Seldane, available only by prescription at the time of this study.

MEDICATIONS

Terfenadine (Seldane) is a noncentrally acting, H-1 type antihistamine with nonsedating properties (Boggs, 1987; Meltzer, 1990). Mann, Crowe, & Tietze (1989) and Woodward (1990) have described the chemistry, pharmacology, pharmacokinetics, clinical efficacy, adverse effects, and dosages of many of the nonsedating histamine H1-receptor antagonists and their differences with traditional antihistamines such as diphenhydramine (Benadryl) and chlorpheniramine. Terfenadine has shown little or no performance impairment when compared to the significant performance impairments shown with centrally acting antihistamines such as diphenhydramine (Betts, Markman, Debenham, Mortiboy & McKevitt, 1984; Clarke & Nicholson, 1978; Cohen, Hamilton, & Peck, 1987; Fink & Irwin, 1979; Gaillard, Gruisen, & de Jong, 1988; Goetz, Jocobsen, Murnane, Reid, Repperger, Goodyear, & Martin, 1989; Kulshrestha, Gupta, Turner, & Wadsworth, 1978; Moskowitz & Burns, 1988; Nicholson, Smith, & Spencer, 1982; Nicholson & Stone, 1986; and Schilling, Adamus, & Kuthan, 1990). Performance was assessed in asymptomatic adults with simple tasks such as reaction time, adaptive tracking, continuous memory, visual search, visuo-motor coordination, dynamic visual

---

[1] OMPAT was formerly the Military Performance Joint Working Group on Drug-Dependent Degradation (JWGD[3]), Walter Reed Army Institute of Research.

acuity, digit symbol substitution, divided attention, vigilance, finger tapping, body sway, eye movements, critical flicker fusion and with subjective scales such as mental status surveys, self-rating scales to assess mood state, and symptom questionnaires. However, Bhatti & Hindmarch (1989) did show impairment on laboratory tests analogous to driving an automobile with terfenadine doses of 240 mg, four times higher than normal.

Benadryl (diphenhydramine) is also an H-1 type antihistamine, but often produces a sedative effect due to direct central nervous system (CNS) activation (Spector, 1987; White & Rumbold, 1988). Benadryl was chosen for the present study as a positive control to establish the sensitivity level of the performance measures to detect antihistamine side effects.

PERFORMANCE

All of the studies cited above used simple performance tasks. The impact of the newer terfenadine medication on complex tasks is unknown. Demonstration of an absence of adverse effects on USAF, mission-relevant tasks under terfenadine could potentially reduce grounding time for aircrews by supporting a medical flying waiver. Complex laboratory performance tasks, such as the Complex Cognitive Assessment Battery (CCAB), are beginning to appear (Samet, Marshall-Mies & Albarian, 1987). Intano, Howse, & Lofaro (1991) have used tests from the CCAB to assign aviator candidates to one of four helicopters prior to day 100 of training. Their research group simultaneously pursued two avenues of research. In one, available test instruments were considered and evaluated for their potential to discriminate among aviators. In the other, groups of Subject Matter Experts (SMEs) developed lists of criticality-rated aviator candidate abilities and traits for specific operational helicopters. Four computerized tests were evaluated. The underlying abilities, traits, and skills purportedly measured by the tests matched the abilities, traits, and skills identified as necessary by the SMEs for each of the helicopters. High-time aviators were given the experimental battery to develop scoring profiles for specific aircraft and to generate the data for the statistical analyses.

In initial validation studies, Intano and Lofaro (1990) have shown that the battery of tests distinguished among helicopter training groups and assigned students to different helicopters. The battery also predicted actual flight performance in each group, performance in the common core flight training, and setbacks (retraining). Final validation in the training environment is in progress. The tests have not, however, been normed or validated against complex, real-world work environments.

At the present time, assessing the performance effects of antihistamines on complex task decision-making can best be accomplished in a simulation of real-world C^3 complex tasks under sustained operations.

BACKGROUND

To meet the objectives of this study, two challenging problems required solutions:

1. Objective measures of team and individual complex task performance were not available.

2. There was no military or industrial C^3 simulation facility capable of embedding such measures if the measures had been available.

Government researchers, under the direction of OMPAT, decided to develop a facility for simulating complex, team decision-making problems and quantifying the effects of various independent variables, e.g., drug effects, fatigue, etc. A hardware/software system was designed around networked VAX computers, terminals, voice synthesizers, and Silicon Graphics workstations to run air defense scenarios for WD teams, a Senior Director (SD), simulator pilots, a ground controller, and an experimenter. The components and capabilities of this system are described in Strome (1990). This system aided the development of unclassified scenarios with embedded performance measurement tasks described in a succeeding section. The Aircrew Evaluation Sustained Operations Performance (AESOP) facility is described in Schiflett, Strome, Eddy, and Dalrymple (1990).

OBJECTIVES

The first objective of this investigation was to evaluate the <u>sensitivity</u> of selected C^3 Mission Effectiveness measures and synthetic performance measures to detect any differences in the effects of 2 antihistamine medications, Benadryl and Seldane. The study used 6 empirically derived, unclassified, air defense Airborne Warning and Control System (AWACS) scenarios to evaluate the 2 antihistamine medications against a placebo using a wide variety of performance measures. Three of the scenarios were high difficulty and 3 were low difficulty, as verified by SMEs and AWACS instructor evaluation teams. A second objective was to assess the magnitude of individual- and team-performance <u>impairment</u> of the mission produced by the antihistamines during high- and low-difficulty C^3 scenarios.

The AWACS WD team function was chosen as the complex task because it contained C^3 task elements common to all Department of Defense (DOD) services. More of the behaviors were accessible to performance measurement, compared to other positions on the AWACS team, e.g., surveillance.

SCENARIO DEVELOPMENT

Six 3.5-hr scenarios were designed by AWACS SMEs. The SMEs balanced realism with performance measurement repeatability in defensive counter air (DCA) mission scenarios. Briefly, in a DCA mission, the WD's goal is to defend friendly lines of communication, protect friendly bases, and support friendly land and naval forces while preventing the enemy from carrying out offensive operations. The primary operations are conducted to **detect, identify, intercept,** and **destroy** enemy aircraft attempting to attack friendly forces or penetrate friendly airspace. Five replications of a low-difficulty scenario were modified to appear unique to the WD teams. Aircraft tracks were rotated, land masses and names were changed, and prebrief situation documents were modified. Also, by increasing the variability of elements such as altitude and lane crossovers, three scenarios were modified to high difficulty. Embedded performance measurement tasks were created by timed voice inputs from an SD or other voices digitized offline and presented at critical points in the scenarios by a speech synthesizer. Further details of the scenario development are described in Schiflett et al. (1990).

PERFORMANCE MEASUREMENT HIERARCHY

Of primary importance to this study was the assessment of drug effects on team decision making in performing complex tasks. Although there are several models for evaluating teams, most require inputs from trained observers making subjective ratings. Reliable detection of subtle medications and fatigue effects requires objective, repeatable measures. After a review of team performance literature, Eddy (1989) and Dyer (1986) concluded that no one has systematically developed and empirically tested a comprehensive theory of team performance. As a result, Eddy and Shingledecker (1988) and Eddy (1990) developed a hierarchical performance assessment system to provide structure for understanding performance in WD tasks. This system provides an implicit underlying structure that weights the significance of each measure and relates it to the others. Each level of the hierarchy contains groups of measures that jointly determine the measures available at the next level higher in the framework. This system includes four interrelated levels of metrics (see Figure 1). From the top down the levels are:

- Mission Effectiveness,
- System/Team Performance,
- Individual Performance, and
- Performance Capabilities and Strategies.

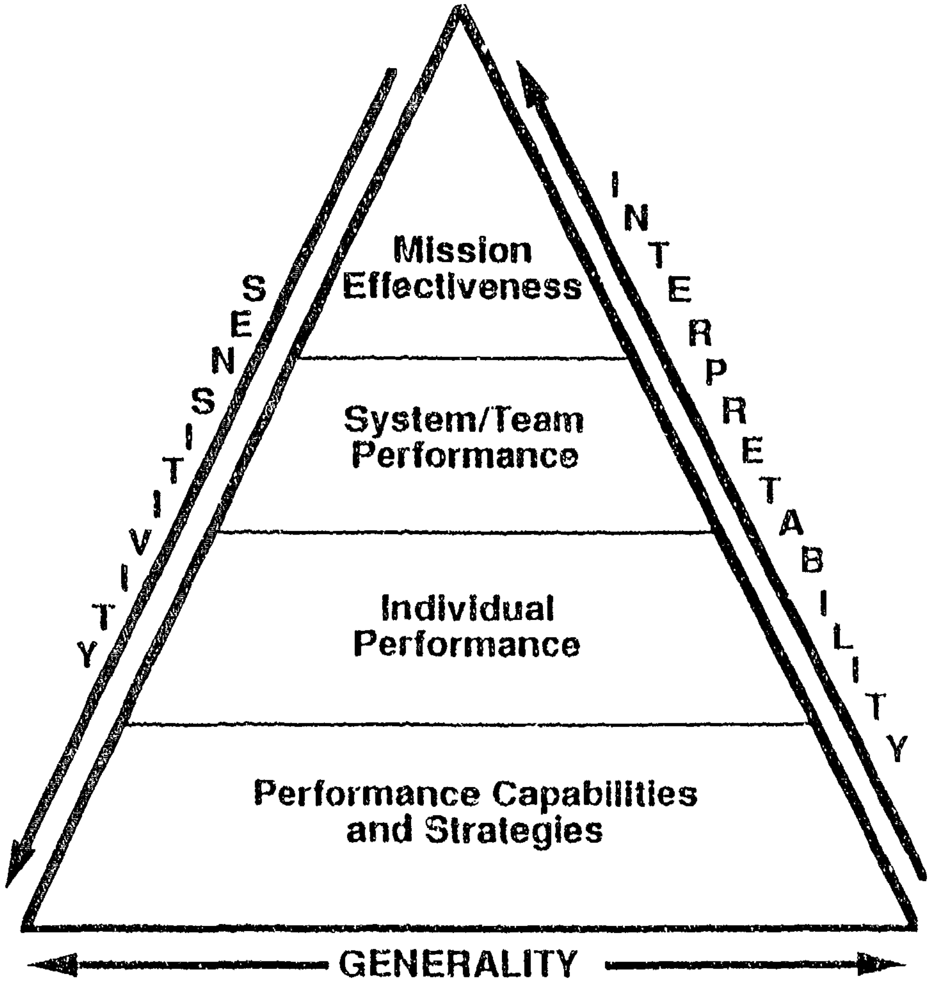

Figure 1. Performance measurement hierarchy.

Each level of the Performance Measurement Hierarchy was developed in conjunction with operationally experienced SMEs in AWACS C^3 tasks. The *mission effectiveness* level is assessed exclusively by *outcome* measures, i.e., measures of the team's results. The *system/team performance* level is assessed by several types of *multi-dimensional* measures combined to quantify changes in situational awareness, cooperation, cohesiveness, adaptation, and distribution of work. The *individual performance* measures consist mainly of *process* measures. Process measures are measures of activities used to accomplish the mission and produce the final results. They include task completion times and response variability, and information processing rates as they relate to unique task assignment. *Performance capabilities and strategies* are measured by skill assessment batteries administered separately from the scenarios.

Mission Effectiveness Measures

Mission Effectiveness measures are derived directly from the specific objectives of the mission assigned to the system. For the C^3 AWACS system the objectives include:

1) protection of a specific sector of air and ground space from infiltration by enemy aircraft (protection of assets),

2) minimization of resource expenditure (fuel, weapons) in protection of assets, and

3) maximization of resource survivability (interceptor aircraft as well as self).

Measures that flow from these high-level objectives and that assess performance in terms of Mission Effectiveness include, among others, the number of enemy infiltrations, fuel and weapons expended, and the ratio of systems returning to systems deployed.

System/Team Performance Measures

The second level of the hierarchy, *System/Team Performance,* contains groups of measures reflecting factors that immediately affect Mission Effectiveness. These include the threat environment (composition and performance of enemy forces), the physical environment (weather, etc.), and the performance of the C^3 system itself. Since the emphasis of the simulations was to measure the factors under at least partial control of the human operator, it was the latter group of determinants that was of interest.

Such measures of System/Team Performance reflect the degree to which the combined human-machine system has accomplished those tasks required to meet mission objectives. These metrics do not reflect the individual contributions of different human behaviors or various hardware and software component performances. Instead, they are more global indices of the degree to which the total system successfully accomplished the tasks essential to mission success.

In order to derive such measures, it was necessary to obtain a detailed description of the specific methods by which the system accomplishes its mission. For example, the weapons director/workstation system is required to meet its mission objectives by accomplishing a weapons control function aimed at directing interceptor aircraft to defeat threat aircraft. This weapons controller task was broken down into a number of essential subtasks such as pairing of interceptors with targets, providing target data to interceptors, and maintaining target correlation, among others. Performance measures of these system tasks include the proportion of time that targets are uncorrelated and the accuracy and speed of data transfer to interceptors, among others.

Individual Performance Measures

The third level of the hierarchy, *Individual Performance,* contains process measures that assess the individual contributions of hardware/software and human components to overall system performance. Measures of the Individual Performance level of the hierarchy are designed to reflect the quality of the individual behaviors required of the WD expressed primarily in terms of latencies, errors, and rate of correct responses. These metrics are derived by examining the system functions required to meet mission objectives to identify the specific contributions of the operator. For example, the system performance requirement to pair targets with interceptors tasks the WD to identify a target's location on the workstation display and to communicate this information to an interceptor aircraft via radio. The quality of the operator's performance in achieving this objective can be measured by evaluating the time needed to complete the full sequence of required behaviors and by assessing the accuracy of each manual and verbal response.

In deriving the Individual Performance Measures, it is crucial to ensure that the aspect of performance assessed is a true contributor to system performance. For example, assessing response time on a task component not time-critical could easily lead to erroneous conclusions about the operator's performance.

Performance Capabilities and Strategies Measures

The final level of the hierarchy, *Performance Capabilities and Strategies,* contains measures that assess factors directly affecting the individual performance capacities of primary system components. For hardware, these measures might include data transfer rates, component reliabilities, etc. For the human operator, measures of Performance Capabilities and Strategies are composed of a large group of potential human state and ability metrics that combine to determine overt performance. These metrics include indices of workload or reserve processing capacity, fatigue, mood, arousal level, experience level, and individual perceptual, cognitive and motor abilities that make up the total productivity of the operator.

Hierarchical Relationships

The multilevel classification of performance measures has the advantage of placing metrics into logical subordinate and superordinate groups that indicate the predictive relationships among them. Measures at each of the levels differ in their sensitivity, generalizability and practical interpretability. Examining the hierarchy, it is obvious that the data provided by the highest level of measurement is easily interpreted while that from lower levels offers information increasingly remote from the ultimate criterion of mission success or failure. However, this disadvantage is countered by the fact that measures at lower levels of the framework are both more sensitive and more general than those at higher strata. For example, while kill ratios are direct indices of Mission Effectiveness, these measures are influenced by a host of individual factors that make them insensitive to small but significant variations in such things as operator decision time. Furthermore, Mission Effectiveness measures are highly specific to the individual characteristics of the test scenario. Hence, an effectiveness metric obtained under one set of conditions may give little indication of the system's performance in a different situation. Conversely, a measure of operator reserve capacity, such as a response time on an embedded secondary task, is difficult to relate directly to a criterion such as survivability. At the same time, however, such a measure is generalizable across a wide range of simulation scenarios and will be extremely sensitive to variations in operator capability.

These features of the different levels of performance measurement make it extremely important to identify the specific assessment goals of a system simulation in order to ensure appropriate data are collected. Since a primary goal of the simulations was to explore the impact of operator variables on system and mission performance, it was necessary to collect detailed measures of Mission Effectiveness and System/Team Performance in order to identify operationally significant effects of the medications and stressor variables. However, because of the predicted limited sensitivity and generality of these measures, it was also necessary to obtain measures from the lowest levels of the performance hierarchy. Such Individual Performance and Performance Capabilities and Strategies metrics extend the utility of necessarily constrained research studies and permit generalization to a wide range of systems and mission scenarios.

CORRELATING MEASURES

In attempting to measure complex decision-making performance, correlations with other simpler performance measures should be explored. These simpler measures may be predictive of the complex decision-making performance. If the simpler measures are found to be predictive, they may be useful in selecting future WDs.

The study used several classes of measures and subjective instruments: cognitive and psychomotor performance measures, standardized complex task measures, personality measures, sleep survey, mood scale, fatigue scale, subjective workload scale, biographical sketch, and a WD experience form.

PROBLEMS IN MEASURING PERFORMANCE IN A COMPLEX, 2-SIDED ENVIRONMENT

The AWACS DCA mission scenarios can be considered a 2-sided environment in that the actions of the defenders affect the reactions of the aggressors. Kubula (1978) described many of the problems in attempting to measure performance in a 2-sided test. Although the realism of an aggressor force adds to the reality of the scenario, it also makes each test unique. Two of the problems include: (1) the nonrepeatability of events from one team to the next, allowing members of a team to overextend themselves on a problem to such a degree that they are not ready for the succeeding events programmed into the script, and (2) group responses may be unique to only one team and hence cannot be compared to the responses of other teams.

We solved some of the problems in our simulations by having a single SD who was a part of the experimenter's team of players. The SD kept the team in bounds with regard to having enough resources to fight the war and to breaking off intercepts and other distractions that would remove the WD from significant upcoming events requiring specific responses. These "assists" by the SD were weighted and counted against the team as necessary interventions.

METHODS

Subjects

For twelve weeks between July 10, and October 20, 1989 (testing was not conducted during the weeks of July 17, July 24, and September 4), the 552d Air Wing at Tinker AFB assigned teams of 3 WDs, who had previously volunteered, to spend their work week in support of this study. All subjects had successfully completed the required USAF training courses for qualification as WDs. Each team was randomly assigned to a drug treatment condition. All subjects signed the Human Use Committee's approved consent form prior to any data collection. Female subjects had a negative pregnancy test within the previous 30 days and signed a pregnancy disclaimer.

Tasks

WDs in an air defense scenario use their consoles to accomplish a number of tasks. The wartime tasks include the following:

- locating and identifying aircraft,
- maintaining track information on aircraft and targets,
- updating display information received from pilots,
- accepting aircraft hand-offs,
- performing a tactical controller function with appropriate level of control using voice communication or data link,
- communicating target information to interceptors,
- performing a tanker controller function through communications with tankers and interceptors,
- using communications to provide recovery assistance,
- safe passage monitoring,
- briefing the SD of any tracking or sensor data problems, and
- responding to alerts, alarms, and messages on the console.

The success of the C^3 mission results directly from the WDs' successful accomplishment of their duties as individuals and as a team.

The WD's goal in a DCA mission is to defend friendly lines of communication, protect friendly bases, and support friendly land and naval forces while preventing the enemy from carrying out offensive operations. The primary operations are conducted to detect, identify, intercept, and destroy enemy aircraft attempting to attack friendly forces or penetrate friendly airspace. All other operations are secondary: provide warning, command, and control to friendly forces, handle air refueling, conduct search and rescue (SAR) operations, etc.

The SD, supervisor for the WD portion of the AWACS team, assigns tasks to the WDs, maintains situational awareness, maintains a log of interceptor assignments by WD, and helps WDs requiring assistance. In our scenarios, the SD played the role of a passive-reactive leader, frequently found in the operational community. The SD in our simulations allowed the WDs to scramble their own interceptor flights, SAR aircraft, and return fighters to base. They also conducted their own refueling operations unless, as a team, they decided to assign that duty to one WD. The WDs also used their radios to query the battle manager on the ground for permission to violate Rules of Engagement, for information, or for other instructions. The SD interacted during the scenarios to ask questions of the WDs (embedded tasks), kept the action within the scope of the systems measurement capability (scramble interceptors if a WD was about to run out), and assisted WDs when they became so task-saturated that they could not continue providing at least tactical control.

Simulator pilots, fighter employment agencies, and ground controllers responded to the radio communications from the WDs. The simulator pilots, retired USAF command pilots with air combat experience, responded to WD directives and queries in a real-time fashion. By placing combat experience in the simulator cockpit, situations were prevented in which a WD could ask a question or give information on a topic that a naive simulator pilot might not be able to answer. Such a situation could easily detract from the realism of the simulation.

Due to space and equipment limitations, each pilot simulated more than one pilot/aircraft at a time. To prevent them from becoming task-saturated and losing control of the experiment, the friendly fighters were given some automated parameters. For example, at 21 nm on a cutoff intercept, the computer system took control of the interceptor flight and flew the final attack phase for the simulator pilot. The computer informed the simulator pilot of JUDY (pilot control of the intercept) and launched a FOX 1 missile. This methodology provided consistency across all pilots, scenarios, days, and teams. Differences among WD teams, different scenarios, etc., could not be accounted for by simulator pilots using different tactics within 21 nm of their target. As a result, the quality of the intercept could be attributed to the skill and tactics of the WD who placed the interceptor in the most favorable position to destroy the target. However, friendly fighters and the E-3 AWACS aircraft itself could be destroyed by enemy aircraft.

The simulator pilots were instructed to perform their functions as if they were actually flying the aircraft. However, since their actions were the events that triggered actions by the WDs, they had to consistently interact with each WD controller by following a scripted communication language. Generally, simulator pilots were instructed not to correct problems created by a WD, such as flying a head-on intercept without radar ordnance. They were instructed not to disagree with the WD on the intercept strategy, post-attack vector, refueling, or other WD decisions. This strategy placed all the responsibility for the outcome on the WD, the subject of the study.

Tests and Scales

Subject Biographical Profile

The Subject Biographical Profile is a standard interview instrument requesting personal information about the subject. It requests the study title, location, date, subject's education, sex, age, etc. It also requests information on vision, hearing, medication usage, and sleep patterns. For this study, typing speed was also requested for prediction of keystroke errors in using the Generic Workstation. This instrument was administered during the subject's introduction to the study.

Weapons Director Experience

The WDs recorded their experience in directing aircraft on the Weapons Director Experience form. It included E-3 hours, simulator time, participation in exercises, and their experience with the other subjects on their team. This form was administered during the subject's introduction to the study.

Sleep Survey

The Sleep Survey, USAFSAM Form 154 (September 1976), is used to record a subject's sleep pattern. Subjects recorded their overnight sleep hours on the form and also included information on the quality of their sleep, trouble going to sleep, and whether or not they felt like they needed more sleep. The Sleep Survey was completed each morning.

Antihistamine Symptom Questionnaire

The Antihistamine Symptom Questionnaire, completed with each drug administration for assessing the potential symptoms resulting from antihistamine consumption, was used to monitor deleterious effects of the medication as well as other potentially disruptive symptoms, such as headaches.

Mood II

The Mood II scale, developed by Thorne et al. (1985) and modeled after the Profile of Mood States (POMS), records a subject's instantaneous feelings. It has only 36 items instead of the 65 of the POMS. The subject's response is the level, 1 to 3, of agreement with the item. The items are divided into 6 scales. The raw data are the sum of the values given by the subject on each scale. Since the total number of items differ in each subcategory, the scores require conversion to percent of maximum possible. The Mood II is administered on an IBM compatible computer and is taken at the beginning and at the end of a duty day. This test measures specific mood effects that can be correlated with general performance effects on both the simple cognitive tasks and the simulations.

Subject Workload Assessment Technique (SWAT) (Reid and Nygren, 1988)

At the end of a simulation, each subject evaluated the difficulty of the scenario along the SWAT's 3 dimensions: time load, mental effort, and psychological stress. These measures were weighted against each subject's individual assessment of workload to give an overall value. Each individual's assessment of workload was obtained by sorting all possible combinations of each level of the 3 dimensions. These data represent an independent and standardized assessment of the difficulty of each simulation. The average objective workload measures were compared against the SWAT scores.

AWACS-PAB

The AWACS-Performance Assessment Battery is composed of tests from 2 different performance batteries. The Unified Triservice Cognitive Performance Assessment Battery (UTC-PAB) was developed by representatives from the Air Force, Army, and Navy under the direction of OMPAT (Perez et al., 1987; Reeves et al., 1989). It consists of 25 tests selected for their potential sensitivity to the effects of protective chemical defense drugs on human perceptual, motor, and cognitive performance. An investigator may select those tests from the UTC-PAB most appropriate for the independent variables to be tested. The following tests from the UTC-PAB were selected because of their sensitivity to drowsiness and fatigue or because the WD tasks were built on the specific abilities assessed by the test: Matching to Sample, Code Substitution, Pattern Comparison, Logical Reasoning, Dual Task (Memory Search/Tracking), and Dichotic Listening.

The Complex Cognitive Assessment Battery (CCAB), developed by the Army Research Institute and also sponsored by OMPAT (Hartel, 1988), consists of 8 tests selected because of their similarity to many complex tasks routinely performed by DOD personnel. In using the CCAB, the investigator selects tests most appropriate for the independent variables to be measured. The following 2 tests were selected because of their similarity to WD tasks built on the specific abilities assessed by the tests: (1) Numbers and Words, and (2) Mark Numbers.

Standardized Personality Tests

The Standardized Personality Tests were included to investigate their potential as WD selection instruments. The tests included the Rotter Scale, which assesses the "locus of control" generally perceived by a person in causing changes to take place in one's life; the Personal Characteristics Inventory (PCI), which assesses attitudes and leadership qualities; the Life Style Questionnaire, which predicts a subject's performance under stress; the Least Preferred Coworker (LPC) Scale, which may identify a WD's leadership style; the Jenkins Activity Scale, which assesses a WD's personality characteristics of decision making; and the FIRO-B, which measures a subject's attitudes with regard to sociability and social interaction. A further explanation of these tests is discussed in Nesthus, Schiflett, Eddy, and Whitmore (1991).

Other Tests and Scales

The following tests and scales were also used: the USAFSAM Fatigue Scale, frequently used by the USAF, in which subjects describe their perceived levels of fatigue at that time; an Operational Impact Survey, which allows individual subjects to rate how well the team completed its mission and how well individual subjects completed their parts of the mission; and a Scenario Evaluation form, allowing each WD to rank the simulations with respect to difficulty.

Research Design

The study used a double-blind design with a different drug administered to each of 3 groups. The 3 drugs included Seldane, Benadryl, and placebo control. Twelve teams of 3 subjects each were tested together under placebo and 1 drug in both high- and low-difficulty conditions over 3 days (see Figure 2). Each team received 1 of 2 orders of difficulty to balance the order of these treatments during the morning and early evening sessions. Teams were randomly assigned without replacement to an order of difficulty. Table 1 shows the daily schedule of testing activities.

Day 1	Day 2 (placebo only)	Drug	Day 3	Day 4
Training Only	Easy* Hard	Benadryl	Easy Hard	Easy Hard
No Drug	Easy Hard	Seldane	Easy Hard	Easy Hard
	Easy Hard	Placebo	Easy Hard	Easy Hard

* Order of scenario difficulty level was counterbalanced within each drug group, N=12, for morning and evening.

Note: On Day 2 all groups received placebo.

Figure 2. C^3 research design.

TABLE 1. TUESDAY, WEDNESDAY, & THURSDAY TESTING SCHEDULE (DAYS 2, 3, 4)

Time	*Activity*
0600	Breakfast at AESOP
0630	**Drug/Placebo** & Questionnaires at AESOP Briefing Room
	1. Sleep Survey (pencil/paper--Briefing Room)
	2. Antihistamine Quest. (pencil/paper--Briefing Room)
	3. Mood II (computer--Room 24X)
0700	Pre-Brief
0730	USAFSAM Fatigue Scale (Simulation Room--before sim)
0730	**Simulation Morning**
1100	1. Operational Impact I
	2. USAFSAM Fatigue Scale (Simulation Room--after sim)
1100	Post-Brief
1130	Lunch, **Drug/Placebo** Questionnaire at AESOP
	Antihistamine Questionnaire
1230	USAFSAM Fatigue Scale (Room 24X--before PAB)
1230	PAB Testing I, III, V
1330	PAB Testing II, IV, VI
1430	1. USAFSAM Fatigue Scale (Room 24X--after PAB)
	2. Rotter Scale (pencil/paper--Tuesday)
	3. Life Style (pencil/paper--Tuesday)
	4. LPC Scale (pencil/paper--Tuesday)
1430	Mission Planning and snack
1500	**Drug/Placebo** & Questionnaire at AESOP
	1. Antihistamine Questionnaire (pencil/paper--Briefing Room)
	2. Jenkins Activity Survey (pencil/paper--Tuesday)
	3. PCI (pencil/paper--Tuesday)
1530	Pre-Brief
1600	USAFSAM Fatigue Scale (Simulation Room--before sim)
1600	**Simulation Early Evening**
1930	Post-Brief & Questionnaires
	1. Operational Impact II
	2. USAFSAM/SWAT Scale (Simulation Room--after sim)
	3. Mood II (computer--Room 24X)
2030	Supper, free time (see notes)
2230*	Phone calls to all subjects
	Drug/Placebo & symptom questionnaire
	Antihistamine Questionnaire (pencil/paper--take home)

*2230 events do not occur on Thursday

Because of the possibility that one group could receive, randomly, all above-average teams, a placebo condition during the first testing day was included to ensure the performance equivalence of the 3 groups. Should the groups be different on the placebo day, their scores on drug days could be weighted by subtracting the placebo day scores. This weighting would allow a statistical analysis neutralizing the before-drug differences. Accordingly, this testing day was single-blind in that the experimenters were aware of the drug condition on this day only. The subjects remained unaware of the drug condition beginning the evening of the training day and continuing throughout the study.

Procedure

Upon arriving at Brooks AFB on Saturday or Sunday evening, the subjects received a packet of materials explaining who was in authority, where and when to report, what to expect (brief schedule of the week's events), and where service facilities were located on base. The team of 3 subjects reported to the laboratory at 0700 after breakfast on Monday morning. Monday was used primarily to train the subjects on the cognitive performance tests, to acquaint them with the simulated WD workstations, and to obtain data on paper and pencil tests. Table 2 is the schedule followed for Monday's Training.

TABLE 2. MONDAY TRAINING SCHEDULE (DAY 1)

Time	*Activity*
0630	Breakfast (WDs on their own)
0700	Introduction & Questionnaires at AESOP Briefing Room
	1. Subject Bio-Profile (pencil/paper)
	2. Weapons Director Experience
	3. Sleep Survey (pencil/paper--for last night)
	4. Mood II (computer--Room 24X)
0730	SWAT card sort
0815	PAB Instructions & Training I
0930	PAB Training II
1030	Lunch
1130	Pre-Brief
1200	**Scenario Training Simulation**
1530	Post-Brief, snack; Rotter Scale (pencil/paper)
1630	PAB Training III
1730	PAB Training IV
1830	Questionnaire: Mood II (computer--Room 24X)
1845	Supper, free time
2230	Phone calls to all subjects
	Drug/Placebo & take-home symptom questionnaire
	Antihistamine Questionnaire (pencil/paper--take home)

As noted in the schedule, teams were briefed, signed the Human Use Committee's approved Subject Consent form, completed a biographical survey, WD experience questionnaire, and sleep survey, and then performed the SWAT card sort. Then they were taken to the performance assessment laboratory where they responded to the automated Mood II Questionnaire and were trained on 6 simple computerized tests and 2 complex tests: the AWACS-PAB. Four 60-minute training sessions were given on the computerized tests: two in the morning and two in the afternoon. After lunch and a pre-briefing, the subjects ran a 3.5-hr C^3 training scenario to familiarize them with the simulated AWACS crewstations and scenarios; no drugs were administered. The Rotter Scale was given after the simulation run post-briefing and before the afternoon PAB training. The AWACS-PAB and all paper and pencil tests are described in Nesthus et al. (1991). The Mood II was taken after the last performance test of the afternoon training. Subjects ingested 1 Benadryl placebo and 1 Seldane placebo at 2230 or prior to going to sleep.

On Tuesday morning, the first day of testing, teams reported to the AESOP facility at 0600 for breakfast. After breakfast, teams ingested 2 placebos, completed sleep and symptom surveys, and responded to the Mood II (see Table 1). Although teams were instructed to plan by themselves for the morning simulation scenario, a prebriefing was given by the SD to clarify the objectives of the mission, give out information, and answer specific questions. Approximately 5 minutes before the start of the simulation, each subject completed a USAFSAM Fatigue scale. Teams performed their WD tasks during a 3.5-hr scenario. At the completion of the simulation, subjects completed another USAFSAM Fatigue scale and indicated the subjective level of workload by giving SWAT ratings. After a post-briefing session, subjects took a light lunch and ingested 2 more placebos, followed by two consecutive cognitive performance testing sessions. The 50-minute AWACS-PAB sessions were separated by a 10-minute rest. Thereafter, the WD team had time to plan its next mission for the evening simulation. Subjects were not allowed to sleep or rest at any time other than after the final simulation of the day. The events of the evening simulation were identical to the morning. After the post-briefing, the subjects took the Mood II survey before leaving for dinner.

Table 1 also shows an event time-line of the dose administration and experimental event schedule for each 16-hr session. Drugs were administered 1 hour before the beginning of any performance testing. All groups ingested placebos only during the testing schedule for Tuesday, Day 2 (Figure 2). Starting on Tuesday evening, a randomly assigned team ingested the recommended therapeutic dose of either Benadryl plus lactose placebo, Seldane plus lactose placebo, or both lactose placebo preparations. Total antihistamine/placebo ingestion for each group consisted of either 8 Benadryl 25-mg tablets and 10 placebo preparations; 4 Seldane 60-mg tablets and 14 placebo preparations; or 18 placebo preparations.

In order to keep the experiment double-blind, dosing regimens for all groups followed the same regimen as for both Benadryl and Seldane. Benadryl and Seldane have different appearances, hence the concurrent schedules under all test conditions.

Each medication and its placebo looked identical to prevent the identification of the drug by appearance. Therefore, each subject, regardless of group, consumed 18 capsules.

At 0630 Wednesday, depending on the group assignment, each team member ingested the second dose of the drug treatment with the other placebo (or 2 placebos) after the normal breakfast meal. All events of Tuesday were repeated on Wednesday and Thursday. Subjects did not take a drug or placebo Thursday at 2230.

The subjects' only free time was in the evening. They were expected to eat lightly, limit alcohol consumption, ingest the assigned capsules, and retire by 2230. Caffeine intake was prohibited throughout the testing session. Decaffeinated sodas, herbal tea, and water were available periodically during the off-task times. Smoking was allowed in designated, outside areas, during off-task periods only. Meals were low in protein to prevent the slower absorption of drug into tissue due to plasma protein binding.

Air Defense Commander's Perspective

In the present report, only the Mission Effectiveness level measures are analyzed. Other reports describe the results at the Performance Capabilities and Strategies level (Nesthus, 1991). Reports at the System/Team and Individual Performance levels will be published later. Correlations of Performance Capabilities and Strategies measures with those of the simulation tasks will provide data to assess the feasibility of predicting complex "real-world" performance from laboratory tasks under the same medications. The Mission Effectiveness level was analyzed first because of the need to show Tactical Air Command (TAC) the capabilities and realism of the system. The higher level measures are also more easily interpreted.

At the Mission Effectiveness level, the viewpoint of the Air Defense Commander (ADC) is taken. The ADC is interested in 2 basic questions. Did the aircrew "win the war?" And at what cost? From the ADC's point of view, the DCA mission overshadows the other supporting specialized tasks of: 1) Intelligence, 2) Weather Service, 3) Aerial Refueling, 4) Search, Rescue and Recovery, and 5) Warning, Command, Control, and Communications. It was assumed that some level of efficiency was achieved before effectiveness was reached. The relationship of effectiveness to efficiency is considered throughout our interpretation of the data.

In articulating specific questions, a model of operator behavior was assumed:

Detect → Identify → Intercept → Destroy = Assets Protected

Asset protection results from intercepting and destroying enemy aircraft, which is based on prior identification and detection. Working backwards from this model at the upper level of the performance measures hierarchy, 11 questions were developed that an ADC would ask to evaluate performance. Most of the ADC questions have a quantifiable answer.

Variables

For each of the ADC questions with a quantifiable answer, the numbers were identified in the database by the following variables:

- Session Number
- Week
- Scenario Name
- Drug
- Day of Week
- Time of Day
- Scenario Difficulty

Questions

1. What were the number of "get throughs" or strikes completed by the enemy against friendly bases and assets?

The measurement, Protection of Assets, operationally defined the question of winning the air battle. Since the end point of the DCA mission was asset protection, mission success would be degraded if a hostile bomber successfully bombed friendly ground targets, such as airbases. Although no bombing accuracy was recorded in the scenarios, the system did record when a hostile aircraft overflew a friendly base. Since the system recorded the position of every aircraft each minute, a hostile strike completion was defined as a hostile flight within 5 miles of a friendly airbase. Airbases were represented by Airbase objects or by Special Point objects. Only Airbase objects could be the target of a strike completion. Data recorded for each Hostile Strike Completion included:

- Track designator of striker
- Track team of striker
- Strike objective
- Simulation Time (in minutes)
- Ordinal Strike Counter of Track.

2. What was the ratio of assets lost by category, enemy to friendly?

The concepts of intercept and destroy are closely intertwined by the design of the scenarios. In theory, a hostile aircraft must be intercepted before it is destroyed. Each friendly and hostile fighter had "JUDY" (contact) and weapons firing parameters. When the parameters were met, the simulation software took over control of the flight and fired the weapons, thus destroying misidentified hostile or friendly aircraft. Loss was defined as the difference between the number of assets at the beginning of the simulation and the number of assets at the end of the simulation. Data were tabulated for the following categories:

Hostile

- Airbase
- Aircraft (All)
  - Bombers
  - Fighters
  - Reconnaissance
  - Jammers
- Armaments
  - Radar Missiles
  - Infrared Missiles
  - Guns
- Souls on Board

Friendly

- Airbase
- Aircraft (All)
  - Tankers
  - Fighters
  - Strikers
  - CCC Platform
  - SAR
- Armaments
  - Radar
  - Infrared Missiles
  - Guns
- Souls on Board
- SAMs
- SAM Sites
- Fuel

Loss Ratios for airbases, aircraft, surface-to-air missile (SAM) sites, and pilots were calculated by the formula:

Assets Lost = Total Available - Total Remaining

$$\text{Loss Ratio} = \frac{\text{Hostile Assets Lost}}{\text{Friendly Assets Lost}}$$

Enemy loss ratios, friendly loss ratios, and enemy-to-friendly loss ratios are an attempt to quantify assets lost on both sides. The loss ratios are expressed as percentages as is the custom within the operational community. The ratios were devised for several categories within the groupings of friendly and hostile assets.

3. What was the percent of friendly assets lost, by category?

The totals from question 2 were used to compute the percentages. The following formula was used to calculate Percentage Loss for each category.

$$\text{Percentage Loss} = \frac{\text{Assets Lost}}{\text{Total Available}} \times 100$$

4. What were the kill ratios for all the friendly fighters combined and what were the kill ratios of only the Air Defense Fighters (ADFs) alone?

The kill ratio has historical and operational significance. With a favorable kill ratio, an air defense commander eventually achieves victory, assuming an equivalent amount of assets as the enemy. The ratio has the benefit of combining two quantitative effectiveness numbers into a quantitative efficiency measure. The first quantitative effectiveness measure is the total number of hostile aircraft destroyed. The second is the total number of friendly fighters destroyed by hostile fighters. Since our scenarios included a friendly fighter/bomber strike force, we further subdivided the kill ratio into two groups. One group included only the ADFs, while the other included both the ADFs and the friendly strikers. The categories for kill ratios included:

- Hostile Aircraft destroyed by Friendly Fighter Aircraft
- Friendly Fighter Aircraft destroyed by Hostile Aircraft
- Hostile Aircraft destroyed by Friendly non-Strike Fighter Aircraft (ADFs)
- Friendly non-Strike Fighter Aircraft (ADFs) destroyed by Hostile Aircraft

Separate kill ratios were calculated for all fighter aircraft and for all non-Strike fighter aircraft by the formula:

$$\text{Kill Ratio} = \frac{\text{Hostile Aircraft Destroyed}}{\text{Friendly Aircraft Destroyed}}$$

5. <u>What tactics did the enemy use?</u>

This question requires an explanation of the scenarios used in the study. We developed 7 scenarios, 1 training, 3 low-difficulty, and 3 high-difficulty scenarios. Each scenario was based on a standard enemy attack of 4 waves. The first wave was a reconnaissance probe and had only 3 enemy aircraft. It occurred during peacetime Rules of Engagement (ROE). The second wave had 12 enemy aircraft. Most of the attackers were bombers escorted by fighters or fighter/bombers. It happened under intermediate ROE. The third wave had 12 enemy aircraft also. Again, it consisted mostly of bombers escorted by fighters or fighter/bombers. It started out under intermediate ROE and escalated into wartime ROE. The last wave had a mass of 16 enemy aircraft; most of them were grouped as 1 bomber escorted by 2 fighters or fighter/bombers. The last wave occurred under wartime ROE.

The course of each attacker was laid out on an xy-coordinate plane. A latitude/longitude map was overlaid on the xy map so each (x,y) corresponded to a lat./long. point. To make each scenario appear unique, the xy plane was rotated a number of degrees, and matched with a lat./long. center from a different geographical region. To get a good fit of the geographic points for the enemy and friendly bases, some of their xy coordinates were slightly changed.

The training scenario was a low-difficulty scenario with the distance between way points doubled. The low-difficulty scenarios had at most two turns, two tracks that crossed over a single lane, and no zig-zags, crosses, or weaves within a lane. The hostile aircraft courses of the high-difficulty scenarios were more evasive. They zig-zagged, crossed, weaved, and crossed over between two WDs' lanes.

In both the training and low-difficulty scenarios, the enemy aircraft flew at 40,000 feet. In the high-difficulty scenarios the enemy flew at many different altitudes, adding greatly to the complexity of the hostile threat. The hostile aircraft had no fuel limitations.

The hostile fighters and fighter/bombers had some automated attack parameters. Hostile aircraft changed their altitudes only to attack friendly aircraft, once a set of engagement parameters were met. The parameters were the following: for radar, a cone 30 degrees right or left of the nose and 15 degrees up or down from the nose. If, for example, at 21 nm the parameters for a radar-equipped aircraft were met, the computer system took control of the fighters and fighter/bombers and flew the final attack phase for weapons launch. At 5 nm, the same occurred for the visual bubble. The priority for carried weapons was: 1) radar missiles; 2) infrared missiles; 3) guns. Weapons launches were at 10 nm for radar missiles, 2 nm for infrared missiles, and 1 nm for guns. The weapons loads carried by the hostile forces depended upon their aircraft types. All hostile aircraft and armament were Soviet-made. The weapons loads and performance characteristics of the airframes conformed as specified in open literature sources. This parameter ensured that the hostile forces emulated a real world threat of engaging and destroying friendly aircraft, even though model constraints did <u>not</u> allow for a highly sophisticated emulation.

To fully appreciate the enemy threat, the constraints on the friendly ADFs must be understood. Like the hostiles, the friendly aircraft had weapons loads and performance characteristics of U.S.-made aircraft from open sources of information. They were further constrained with a limited fuel supply, i.e., they could run out of fuel and fall from the sky.

The ADFs also had automated parameters. When the parameters were met, the computer system took over and executed the final attack phase. The parameters were the following: for radar, a cone 30° right or left of the nose and 15° up or down from the nose; and a 10-nm visual bubble for all fighters. If, for example, at 21 nm the parameters for a radar-equipped aircraft were met, the computer system took over control of the fighters and flew in the final attack phase for weapons launch. The same occurred for the visual bubble at 10 nm. The priority for carried weapons was the same as for hostile aircraft: 1) radar missiles; 2) infrared missiles; 3) guns. Weapons launches were at 2 nm for infrared missiles, and 1 nm for guns. Radar missile range varied by aircraft type, but all were greater than 10 nm.

The friendly weapons were lethal only when certain parameters were met. These parameters were based on the friendly fighters Heading Crossing Angle (HCA). For radar missiles, the HCA had to be greater than 120°. For infrared missiles, the HCA had to be 120° or less. Guns are all-aspect. Weapons launch parameters had to be met first or a "No Joy" situation exists. Once weapons are launched, the weapons success parameters had to be met for a kill to occur, otherwise a "Heads Up" situation occurred.

6. <u>How would an ADC interpret and summarize each team's performance for the week?</u>

This question precipitated the development of a composite scoring scheme to provide a standard quantitative measure of a WD team's performance. Using a composite score, changes can be noted and measured by comparing results from the different scenarios each team completed. The term *composite* denotes that multiple performance outcomes are measured and used in the computation of an overall score. All the component performance outcome measures were drawn from the results of the previously described ADC questions.

To derive the mission effectiveness composite score, the overall DCA mission model was used. All the components came from the intercept through asset protection portions of the model. None of the measures for the supporting specialized tasks, e.g., refuelings, were incorporated. The composite score is derived from the following terms:

- The negative square of the number of hostile strikes completed,
- <u>Plus</u> the ADFs' kill ratio <u>times</u> the number of hostiles killed by the ADFs,
- <u>Minus</u> the friendly to hostile aircraft loss ratio <u>times</u> the number of hostiles <u>not</u> killed,
- <u>Plus</u> the square root of the friendly strikers kill ratio <u>times</u> the number of hostiles the friendly strikers killed,
- <u>Minus</u> the total friendly kill ratio <u>times</u> the number of friendlies lost due to friendly fighter fire <u>and</u> hostile fire,
- <u>Minus</u> the friendly aircraft loss ratio <u>times</u> the number of friendlies lost by SAMs <u>and</u> Fuel.

The above can also be expressed in the following scoring algorithm:

$$CS = -(HS)^2 + (KR_A \cdot HD_A) - (LR_H \cdot HND) + (KR_S \cdot HD_S)^{1/2} - [KR_F(FL_F + FL_H)] - [LR_F(FL_G + FL_S)]$$

Key

CS	Composite Score
HS	Hostile Strikes Completed
KR_A	Kill Ratio of ADFs
HD_A	Hostiles Destroyed by ADFs
LR_H	Friendly to Hostile A/C (aircraft) Loss Ratio
HND	Hostiles Not Destroyed
KR_S	Kill Ratio of Friendly Strikers
HD_S	Hostiles Destroyed by Friendly Strikers
KR_F	Kill Ratio of Friendly A/C
FL_F	Friendlies Lost by Friendly Fire
FL_H	Friendlies Lost by Hostile Fire
LR_F	Friendly A/C Loss Ratio
FL_G	Friendlies Lost by Friendly Fire
FL_S	Friendlies Lost by SAMs

A few notes about the above scoring algorithm are in order. The purpose of the composite score was to measure the primary goal of the DCA mission, asset protection. Since the number of hostile strikes was a negative measure of asset protection, it was given a minus sign. Because asset protection was the goal, a means of making it the most important contributor to the composite score was necessary. We squared its value, and placed a minus sign in front. Asset protection is achieved primarily through the destruction of hostile attackers. The ADFs' kill ratio assesses the efficiency of using the ADFs to accomplish the destroy portion of the DCA mission. Subtracting the number of hostiles not destroyed is a factor in the composite score. The friendly-to-hostile loss ratio represents the negative aspects of the DCA mission operations. It accounts for a loss in future combat power. Since the friendly strikers can be used as ADFs in a secondary role, but are not a primary means of carrying out destruction of attackers, the square root of their similarly computed component was taken to give it a lower value. Because of the number of times they were not killed by the hostiles, their kill ratio was then arbitrarily assigned as the total number of hostile attackers for the scenario plus one; hence, zero would not be in the denominator. The remaining components were derived to better account for all losses. The total friendly kill ratio was used to attribute those losses directly related to combat of friendlies shot down by mistaken identification and bad tactics. The friendly loss ratio attributes include losses caused by operator error not actually involving combat of friendlies, for example, those destroyed by friendly SAMs and fuel depletion.

7. Where were the hostile aircraft destroyed?

Another way of examining the efficiency of asset protection is to see how deep the attackers were able to penetrate the friendly assets before being destroyed. This strategy involved noting the position of the hostile aircraft destroyed and referencing it from a common reference point. The data collected include:

- Position of Hostile Aircraft Destruction (in xy-coordinates)
- Track Designator
- Destroying Agent
- Simulation Time (in minutes)

The common reference points used for statistical evaluation were the average position of 1) hostile bases, 2) Combat Air Patrol (CAP) points, 3) friendly bases, and 4) total fixed friendly asset points. A reference distance was created by using the average position of the hostile bases as one end point and the other 3 average positions for the other end point. The destruction position then formed an end point for calculating the distance from each of the other end points. These distances for each hostile track in each scenario were then statistically evaluated. Since the start and destruction times for each hostile track were known, the total time a hostile track was in the system is also statistically evaluated. Each of the hostile tracks was further categorized by the wave in which they were generated.

By creating these reference distances and evaluating the total time of hostile tracks, the 7th ADC question can not only be answered, it can be statistically evaluated. By rotating the hostile track positions of each scenario to a common north/south axis, the positional data can be compared directly for each hostile aircraft destruction location, across all scenarios. The data were compiled and presented by video taping computer displays that showed visually distinct differences in performance.

8. What was the total number of misidentifications, if any?

Rather than try to make a direct accounting of misidentifications, a method of scoring each WD team's ability to perform the task of identification was developed. Since each misidentification has meaning only in relationship to the total identification task, Table 3 illustrates a method of categorizing the elements of the identification task, weighting them, and then adding the components for an overall score on the identification task.

TABLE 3. OVERALL SCORES ON THE IDENTIFICATION TASK

Track Designated as:	Friend	Hostile	Unknown
Track ID'd as:			
Friend	1	-3	-1
Hostile	-2	3	0
Unknown	-1	1	2
Not ID'd	0	-1	1

Each WD team, by its actions, places each discrete track in every scenario into one of the above cells. Multiplying the number of tracks in each cell by its corresponding cell weight, a score for each cell is established. Summing all the cell scores in the matrix gives an overall identification score for the WD team for each scenario.

9. Where were the hostile aircraft first detected?

An ADC needs to know where the enemy aircraft are first detected. This information helps define the beginning of the WD decision task and the solutions available using the air defense forces. Also, the circumstances surrounding the detection task may provide the explanation for any hostile "get throughs."

Detection includes both human and machine system elements and must be objectively measurable. Detection does not occur simply when the track enters the system, but when a track enters the system and a human recognizes its presence. Because human recognition is a mental process, the exact moment of detection cannot be ascertained objectively by an observer. However, it is possible to note the overt human behavior in response to the recognition of a track's presence. Thus the moment of detection can be defined as the time at which a WD performs an action that implies a track has entered the system and has been recognized.

To be detected, a track must exist and be airborne in the system. Several conditions for detection were established:

- When a track is detected, all other tracks in the flight are also considered detected.

- A track may be detected when it joins a flight that has been detected.

- A track is detected when track symbology is placed *near* that track. A position is *near* a track under the following conditions.

  (a) The computer correlates symbology placed at the position of that track. The computer will have selected the closest track, up to 5 nm from the position.

  (b) Symbology placed (either at the initiate or reinitiate WD switchaction) at the position is not correlated to another track, and the position is within 10 nm of the track. The track must be visible at the console specifying the position. When multiple tracks are within 10 nm of the position, the position is only near the closest track.

- A track is detected when another track, under Sim Pilot (SP) control, commits against it, on either an ID or Destroy mission. This circumstance may occur either through SP switchaction or by computer decision (i.e., "Best" mode).

- A track is detected when an arrow, initiated by WD switchaction, is sent near a track.

- A track is detected when a WD manually points out that track to another WD. (This method was not included in the results.)

- A track is detected when a WD verbally points out that track to another WD. (This method was not included in the results.)

No consideration was given to loss of detection. Once a track or tracks met the conditions for detection, it remained in the detected category.

This question required the following data:

- Position of Hostile Aircraft at First Detection
- Track Designator
- Method of Detection
- Simulation Time (in minutes)

10. <u>What were the friendly losses due to friendly fire?</u>

During the confusion inevitably present in battle, losses due to friendly fire are likely. By tracking the types and occurrences of these losses, an ADC can determine the probability of these events causing changes in the efficiency of accomplishing the mission, or in effectiveness of the forces.

This question required the following data for friendly aircraft destroyed by friendly fire:

- Track Designator
- Destroy Agent
- Destroy Agent Type (1 = aircraft, 2 = SAM site)
- Simulation Time (in minutes)

Separate totals based upon type of destruction agent were calculated and labeled by session, week, scenario, drug, and scenario difficulty.

11. <u>What were the friendly losses due to fuel depletion?</u>

Assistance in fuel management is always a critical aspect of air power. Failure to properly provide sources of fuel as they are needed results in lost aircraft assets, breakdowns in airspace coverage, and inefficient use of air refueling assets to support the DCA mission. Noting the numbers and types of friendly aircraft, where they went down, and how far they were from their recovery bases helps identify if and where a problem exists in fuel management and tanker deployment.

This question required the following data for friendly aircraft destroyed by fuel depletion:

- Track Designator
- Friendly Aircraft Category
  (Fighter, Tanker, Striker, SAR, CCC Platform, Other)
- Destroy position
- Simulation Time (in minutes)

Separate totals were calculated by session, week, scenario, drug, and scenario difficulty.

DATA ANALYSIS

The data were evaluated using an analysis of variance (ANOVA) with two repeated measures (difficulty and day) and one grouping factor (drug group). Each item of each question was analyzed separately using the SAS statistical package. Question 5 was not amenable to analysis. It is presented in narrative form. Five hypotheses were tested at the .05 level of significance.

1. Was there a day effect?
2. Was there a drug-by-day interaction effect?
3. Was there a difficulty effect?
4. Was there a day-by-difficulty interaction effect?
5. Was there a drug-by-day-by-difficulty interaction effect?

Because the drug variable changed across days, it cannot be interpreted independent of its interaction with the day variable. Any "real drug effect" will show up in the day-by-drug interaction or the day-by-drug interaction with difficulty.

ALTERNATE DATA ANALYSIS

Since subjects experiencing a high-difficulty scenario before one of low difficulty could give different results, the order of difficulty was counterbalanced within each drug group. If an analysis of the order variable was found to be significant, the other treatment effects could be questioned. Because of this potential problem, an ANOVA of the order and difficulty variables was conducted on the day 2 data for all questions. Unknown to the subjects, all groups were treated with a placebo on day 2. None of these analyses was statistically significant for the order variable.

An alternative approach to processing the data is to analyze for a morning/evening effect instead of scenario difficulty. This approach is possible because the degree of difficulty and the order of administration are counterbalanced for morning and evening scenarios. Unfortunately, it is not orthogonal to the difficulty variable and cannot be assessed in the same design. Since such an analysis is of interest in the area of sustained operations, all dependent measures were analyzed substituting AM/PM for difficulty. This approach with two repeated measures (time and day) and one grouping factor (drug group) did not result in any significant AM/PM effects, but gave similar day effects as the analysis using the difficulty independent variable.

RESULTS

From a behavioral research perspective, only some of the Mission Effectiveness and ADC measures directly measured performance. Outcome measures with excellent face validity (for team performance assessment) are:

- strike completions,
- kill ratios,
- friendly losses by friendly fire, and
- losses by fuel depletion.

Some outcome measures were affected by multiple conditions, which rendered them impossible to interpret behaviorally. For example, the loss of radar missiles by friendly fighters could result from firing at and hitting a target, firing at and missing a target, or when radar missiles were attached to a fighter that was shot down or ran out of fuel.

Presented in this section are the results for each question asked by a typical ADC. The SME's observation on data trends that have operational significance are included even when they did not reach statistical significance.

1. What were the number of "get throughs" or strikes completed by the enemy against friendly bases and assets?

   Table 4 shows the number of "get throughs" or penetrations completed by the enemy against friendly bases and assets, by day and scenario difficulty.

   There were no statistically significant results for this dependent measure. However, two trends were determined by the SME. The first, and strongest, was a difference in difficulty. The high-difficulty scenarios had more hostile strike completions than in the low-difficulty scenarios. The next trend showed a learning effect. There was a general improvement by WD teams in preventing hostile strike completions over days.

TABLE 4. STRIKE COMPLETIONS BY DAY AND SCENARIO DIFFICULTY FOR ALL TEAMS

Condition	Penetrations
Day 2	
High Difficulty	22
Low Difficulty	5
Day 3	
High Difficulty	13
Low Difficulty	3
Day 4	
High Difficulty	6
Low Difficulty	6

2. What was the ratio of assets lost by category, enemy to friendly?

The loss ratios, enemy to friendly, were divided among four categories:

(1) Airbases,
(2) Aircraft,
(3) SAM Sites, and
(4) Souls on Board.

The loss ratios were logarithmically transformed to obtain data that were not significantly different from a normal distribution. The natural logarithm of aircraft and Souls on Board showed significantly higher enemy to friendly loss ratios (more effective performance) for low-difficulty compared to high-difficulty scenarios, $F(1,9) = 23.8$, $p = .0009$ and $F(1,9) = 43.5$, $p = .0001$, respectively.

A significant day-by-difficulty interaction, $F(2,18) = 5.4$, $p = .0150$ for aircraft and Souls on Board showed improvement across days under low difficulty, but not under high difficulty. In Figure 3, Aircraft Loss Ratio--Hostile to Friendly by Difficulty, shows the improvement in the low-difficulty group on days 3 and 4 for the natural logarithm of the aircraft loss ratio. The loss ratios are shown in Table 5.

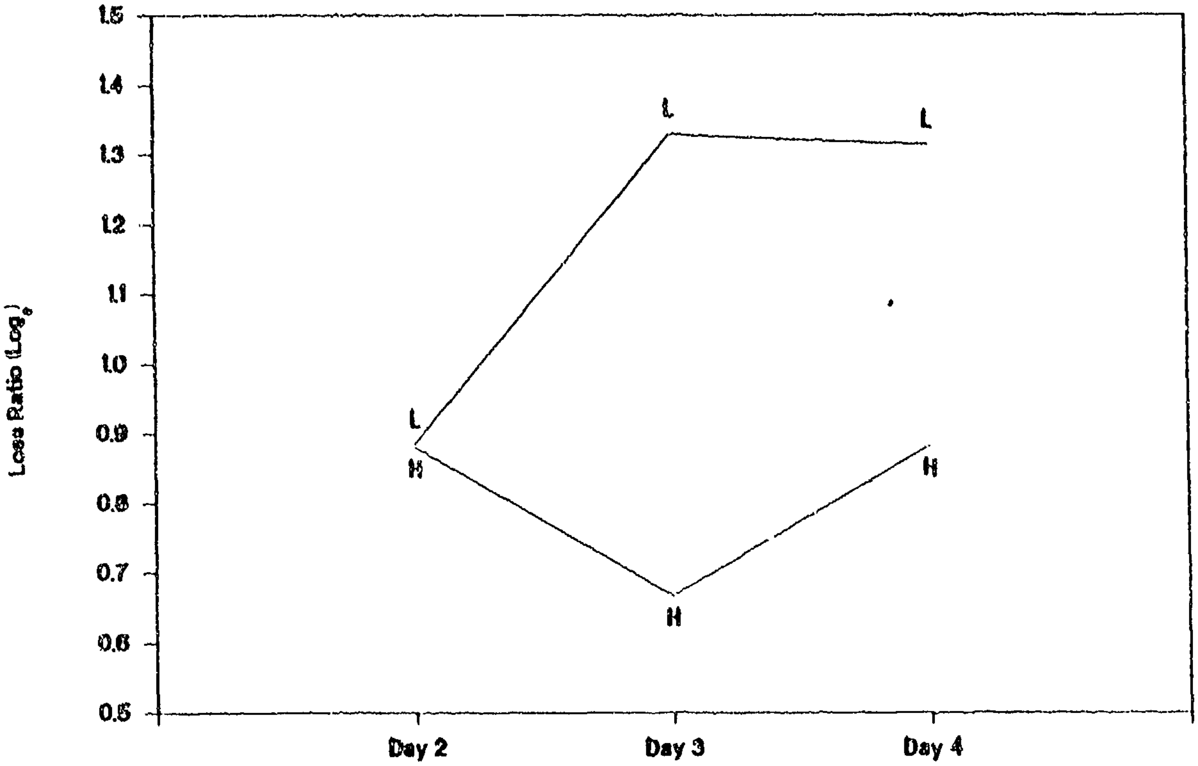

Figure 3. Aircraft Loss Ratio--Hostile to Friendly by Difficulty (H = High Difficulty, L = Low Difficulty).

TABLE 5. LOSS RATIOS FOR ALL AIRCRAFT BY DAY AND SCENARIO DIFFICULTY

Difficulty Level	Day 2	Day 3	Day 4
Low	2.64	5.07	4.09
High	2.71	2.17	2.70

3. <u>What were the percent of friendly assets lost, by category?</u>

Several of the percent of *friendly assets lost* categories showed statistical significance for the scenario difficulty variable. For the percent of all aircraft lost, the difficulty variable approached significance, $F(1,9) = 5.07$, $p = .0508$, but definitely interacted with the day and drug variables, $F(2,18) = 4.36$, $p = .0286$, and $F(2,18) = 4.56$, $p = .0430$, respectively. In Figure 4, Loss of All Aircraft by Difficulty, shows that in the day-by-difficulty interaction, performance on the third day was impaired in the high-difficulty scenarios (Least Squares difference test, $p = .0057$). Although in the low-difficulty scenarios fewer aircraft were lost on days 3 and 4 compared to day 2, these differences were not significant.

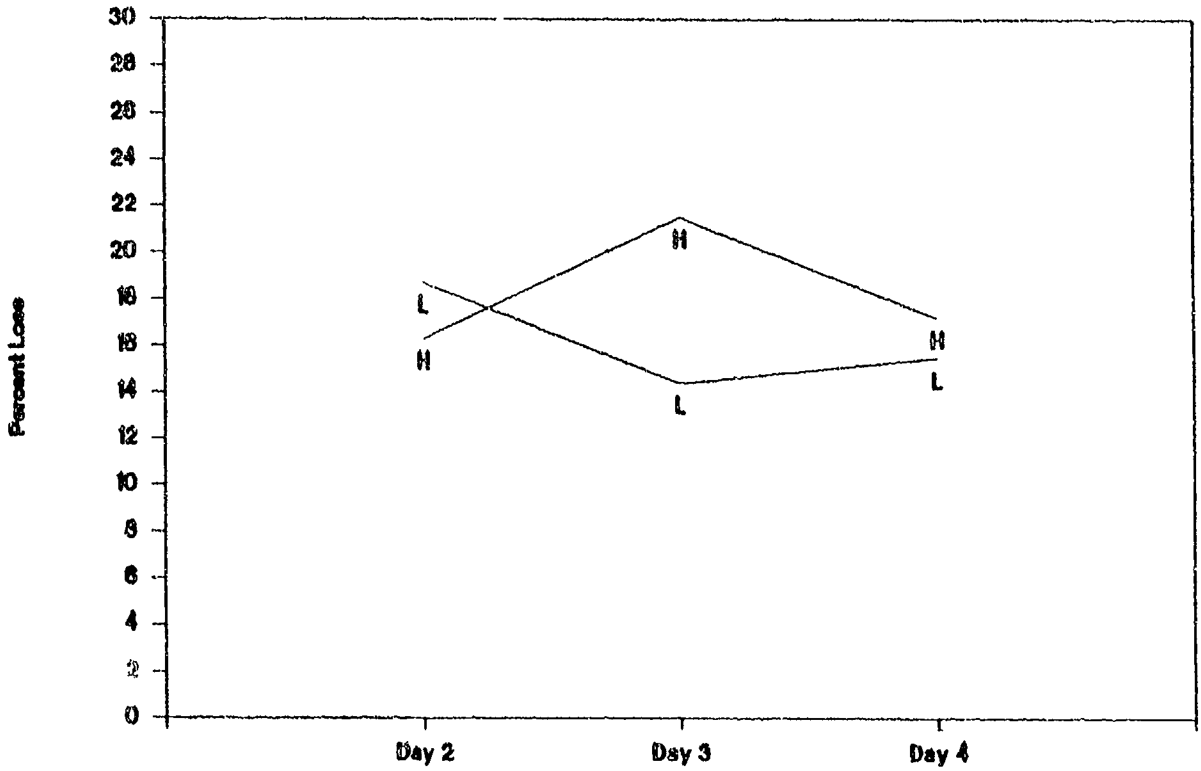

Figure 4. Loss of All Aircraft--by Difficulty
(H = High Difficulty, L = Low Difficulty).

The drug by difficulty interaction indicated that the Benadryl group performed better in low-difficulty rather than in high-difficulty scenarios. However, these results are confounded with the day variable since on day 2 the Benadryl group was under a placebo. Since the three-way interaction was not significant, the drug-by-difficulty interaction is not interpretable.

For percent loss of *friendly airbases,* three statistically significant results were present. The first was an improvement over days; fewer airbases were lost as the week progressed, $F(2,18) = 7.4$, $p = .0045$.

Since hostile strike completions make up half of the definition of what constitutes the loss of an airbase, this result helps provide a better understanding of the strike completion trends. The second result was a difference in difficulty, $F(1,9) = 6.8$, $p = .0236$. Losses of friendly airbases occurred more frequently under high-difficulty conditions. A significant day-by-difficulty interaction, $F(2,18) = 4.0$, $p = .0373$, showed that teams improved their performance more under high-difficulty scenarios than under low-difficulty scenarios (comparison of first and last days using Least Squares difference *t*-tests, $p = .0010$). See Figure 5, Loss of Friendly Airbases--by Difficulty.

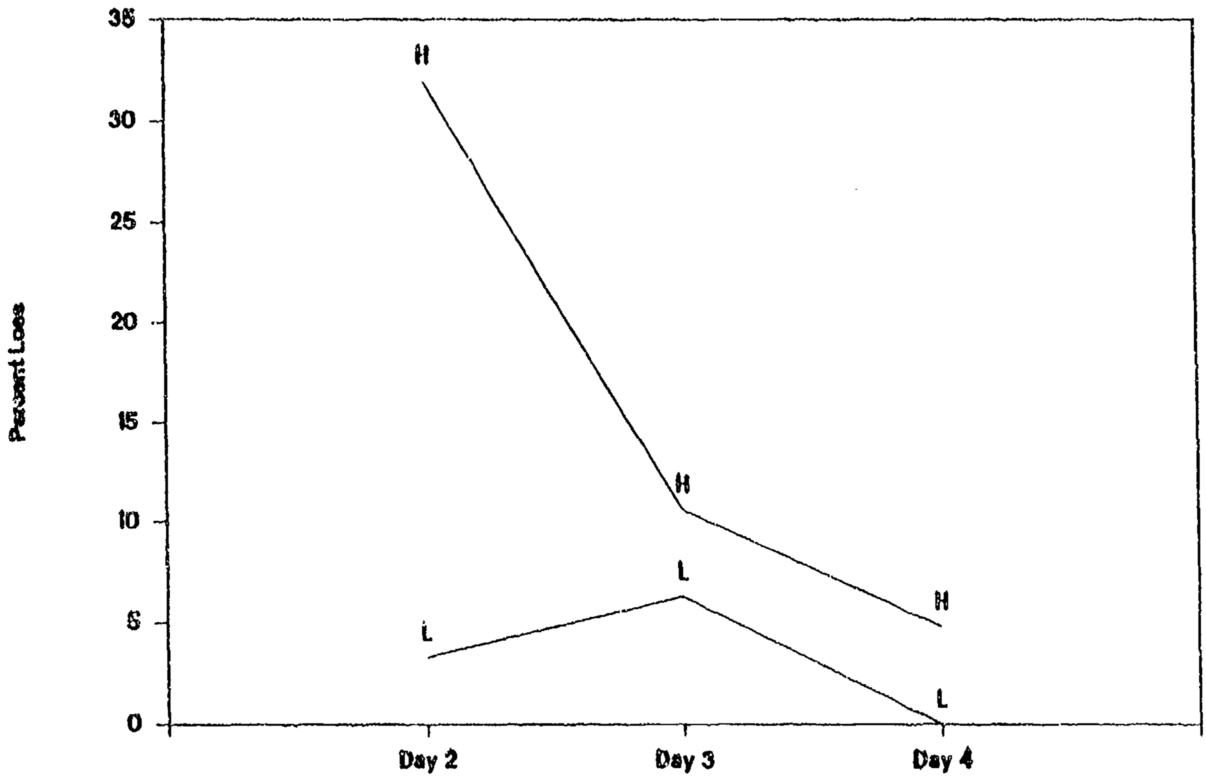

Figure 5. Loss of Friendly Airbases--by Difficulty
(H = High Difficulty, L = Low Difficulty).

For percent loss of *friendly tanker aircraft*, the only statistically significant result was a difference in difficulty, $F(1,9) = 6.6$, $p = .0307$. More tankers were lost under high-difficulty conditions than under low, 19.4% and 6.7% respectively.

The percent loss of *fighter aircraft* was sensitive to the day and difficulty treatments as demonstrated in their interaction, $F(2,18) = 4.85$, $p = .0206$. Figure 6, Loss of Fighter Aircraft--by Difficulty, shows that a significant number of fighter aircraft were lost in the day 2 low-difficulty scenario (Least Squares difference test $p = .0365$).

The drug-by-difficulty interaction, $F(2,18) = 6.1$, $p = .0216$, indicated that the performance of the Benadryl group was best under low-difficulty scenarios compared to the placebo group, Least Squares difference means test, $p = .0041$ and that their performance dropped precipitously under high difficulty, Least Squares difference means test $p = .0136$. These results are confounded, however, with the day variable. See the explanation with similar results for percent loss of all friendly assets.

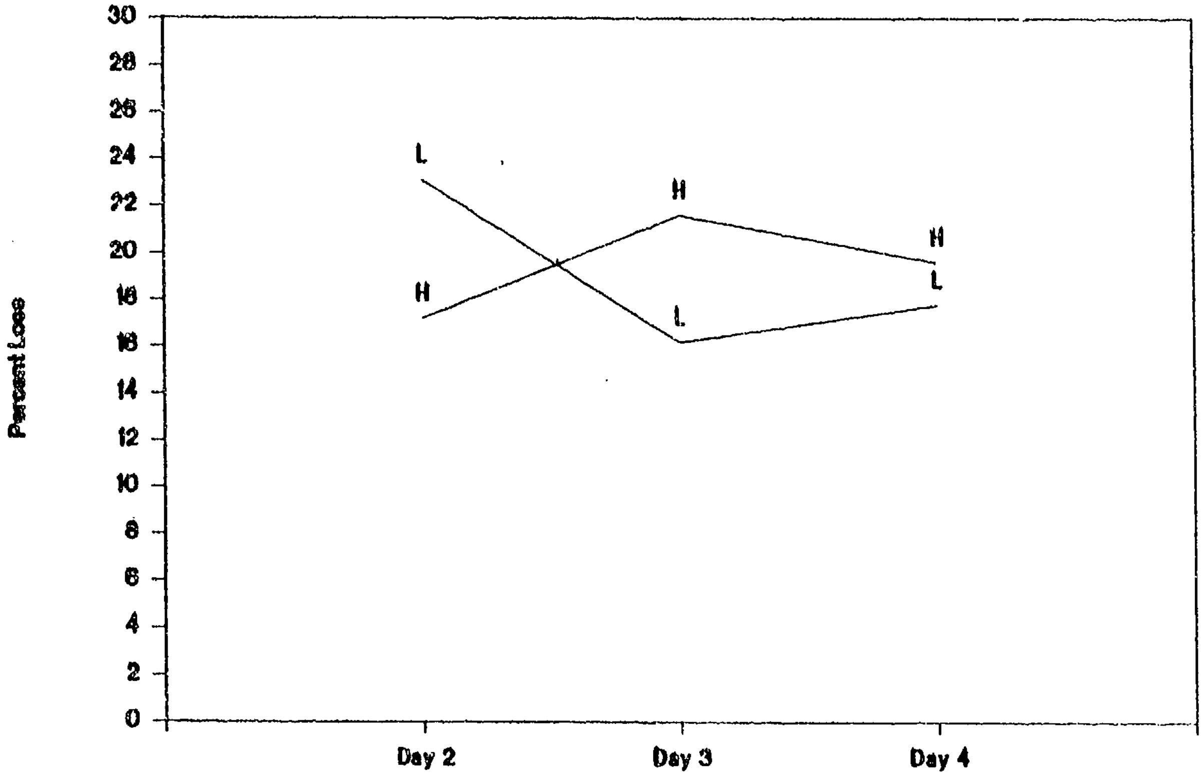

Figure 6. Loss of Fighter Aircraft--by Difficulty
(H = High Difficulty, L = Low Difficulty).

For Airborne Command, Control, and Communication *AB CCC platform aircraft,* the significant day-by-difficulty interaction, $F(2,18) = 8.3, p = .0028$, showed no consistent trends. In 72 scenarios, 7 C^3 platforms were lost. All of these losses came from either the Canaan or Thebes scenarios. Canaan, a high-difficulty scenario conducted on Day 2, had two C^3 platforms. One was an escorted Airborne Command, Control, and Communications (ABCCC) simulated C-130 aircraft and the other was the AWACS simulated E-3 aircraft, common to all scenarios. Operationally, the ABCCC platform was an escort mission and the protection of the AWACS C^3 platform was a self-defense activity. Under Canaan, three teams lost a C^3 platform, all ABCCCs, due to hostile air attack.

Thebes, a low-difficulty scenario conducted on Day 3, had only the AWACS C^3 platform. In Thebes, 4 teams lost their C^3 AWACS platform through fratricide by the friendly SAM site. In each of the instances, the E-3's orbit overflew the friendly SAM site's missile engagement zone (MEZ). Initially, and for most of the scenario, the SAM site remained inactive. When the SAM site became active, two of the teams experienced unscripted equipment failures just prior to the E-3 entering the active SAM site's MEZ.

These interrupts caused a temporary suspension of the simulation that lasted approximately five minutes. Shortly after resumption of the simulation, the C^3 platforms were shot down by the friendly SAM site. In the SME's opinion, these interrupts impaired the team's situational awareness of the tactical flow of events making the conditions of their test unique and not comparable to the other teams. Thus, two of the four C^3 platforms lost under the Thebes scenario were not included in the analysis.

In evaluating the remaining five C^3 losses, no trends were observed by the SME. For example, one was lost due to lack of attention to the radio message announcing the activation of the SAM site. Another was lost because the team had not drawn a circle around the SAM site and didn't know its location. These losses were in a placebo and Seldane group. An ANOVA with only five events was inappropriate.

For percent loss of *SAR aircraft,* there were three statistically significant effects. There was a day effect, $F(2,18) = 6.0$, $p = .0100$, a difference in difficulty, $F(1,9) = 18.6$, $p = .0020$, and a day-by-difficulty interaction, $F(2,18) = 6.5$, $p = .0077$. The teams lost more aircraft under high difficulty, but the interaction showed that on day 3 under high difficulty they lost over 60 percent of all SAR aircraft. The Least Squares difference test showed this day and difficulty different from all others, $p < .0006$. See Figure 7, Loss of SAR Aircraft--by Difficulty.

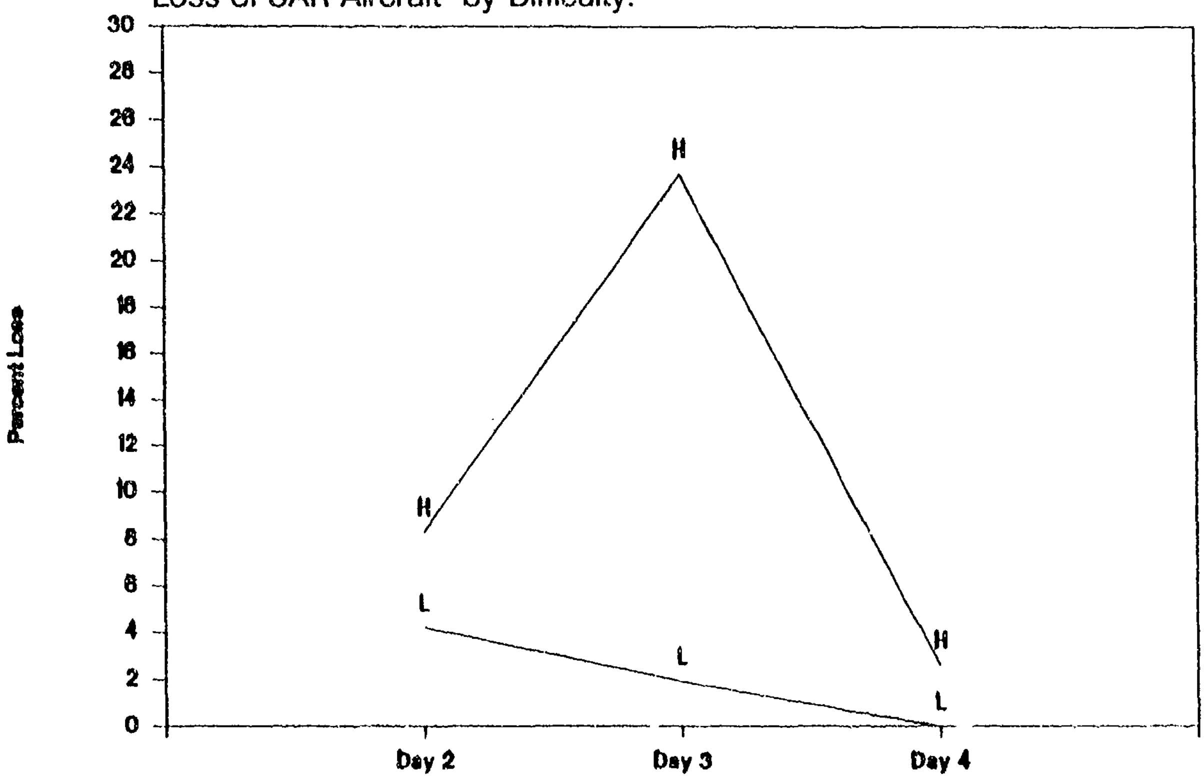

Figure 7. Loss of SAR Aircraft--by Difficulty (H = High Difficulty, L = Low Difficulty).

The percentage of *Infrared missiles* used or lost was significantly affected by the interaction of days and scenario difficulty, $F(2,18) = 4.15$, $p = .0329$. On day 3, more infrared missiles were used in the high-difficulty scenario compared to the low-difficulty scenario (Least Squares difference means test $p = .0250$). Also under the low-difficulty condition, more missiles were used on day 2 than day 3, Least Squares difference means $p = .0230$.

For percent loss of infrared missiles, a statistically significant drug-by-difficulty interaction, $F(2,18) = 12.3$, $p = .0027$, was uninterpretable since drugs are partially confounded with days.

For percent *gun ammunition* used or lost, the only significant effect was for the drug-by-difficulty interaction, $F(2,18) = 6.57$, $p = .0174$, which was uninterpretable because of the partial confounding of drugs and days.

Percent loss of *souls on board* showed a statistically significant difference in difficulty, $F(1,9) = 7.1$, $p = .0258$. The higher difficulty scenarios had higher casualties (14.6% vs. 11.1%). This result was expected since souls on board is correlated with aircraft, and aircraft showed the same effects.

4. What were the kill ratios for all the friendly fighters combined and what were the kill ratios of the Air Defense Fighters (ADFs) alone?

The only statistically significant result found with the natural logarithm of the kill ratios of fighters was a day effect. For all fighters the $F(2,18)$ was 5.0, $p = .0186$.

For the natural logarithm of the non-striker fighters the $F(2,18)$ was 5.1, $p = .0173$.

The WD teams showed an improvement in their kill ratios as the week progressed. The means for the natural logarithm of all fighters were 1.3, 1.4, and 1.7 for days 2, 3, and 4. The means, without including the strike package, were 1.3, 1.6, and 1.8, respectively. The Benadryl teams had lower kill ratios for the first day of the drug, under high difficulty only. However, this trend was not statistically significant.

5. What tactics did the enemy use?

Enemy tactics were previously defined and did not differ across independent variables except for difficulty.

6. <u>How would an ADC interpret and summarize each team's performance for the week?</u>

The performance of each WD team was summarized and interpreted by creating the composite score. Statistical analysis of the composite score resulted in no significant effects. The scores did show two trends. The first trend was an improvement over days. All teams improved their composite score as the week progressed; the means for each day were 122, 173, and '28, respectively. The second trend related to the Benadryl teams only. hey showed a marked degradation in their scores on the first day of exposure to the the drug (Day 3) under high-difficulty conditions. No trends were observed in the low-difficulty scenarios. See Figure 8, Composite Score--High Difficulty Scenario.

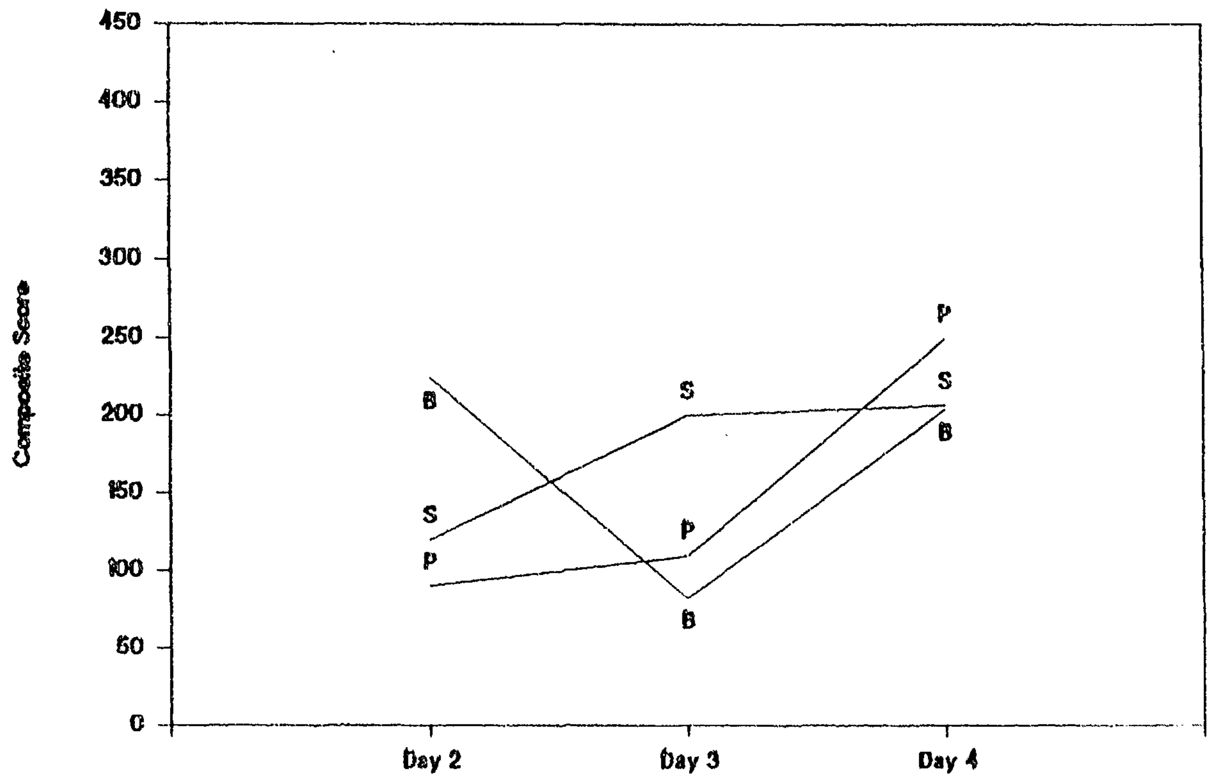

Figure 8. Composite Score--High Difficulty Scenario
(P = Placebo, S = Seldane, B = Benadryl).

7. <u>Where were the hostile aircraft destroyed?</u>

Inspection of graphical representations of the location of destroyed hostile aircraft can be summarized as follows:

- Hostile aircraft penetrated further under high-difficulty conditions.
- Penetration distances of enemy aircraft were different for each wave, reflecting the different ROE effective at the time of the wave.

However, statistical analyses of these data revealed no differences that could not be accounted for by chance.

Wave 1 showed the greatest penetration due to a peacetime Air Defense Warning Level (ADWL). The least penetration of friendly air space occurred during the other three waves under increased ADWL ROE. Figure 9 graphically illustrates these data. The hostile aircraft move from bottom to top. The hostile destruction points tend to form lines along their flight paths because of the similarity among scenarios.

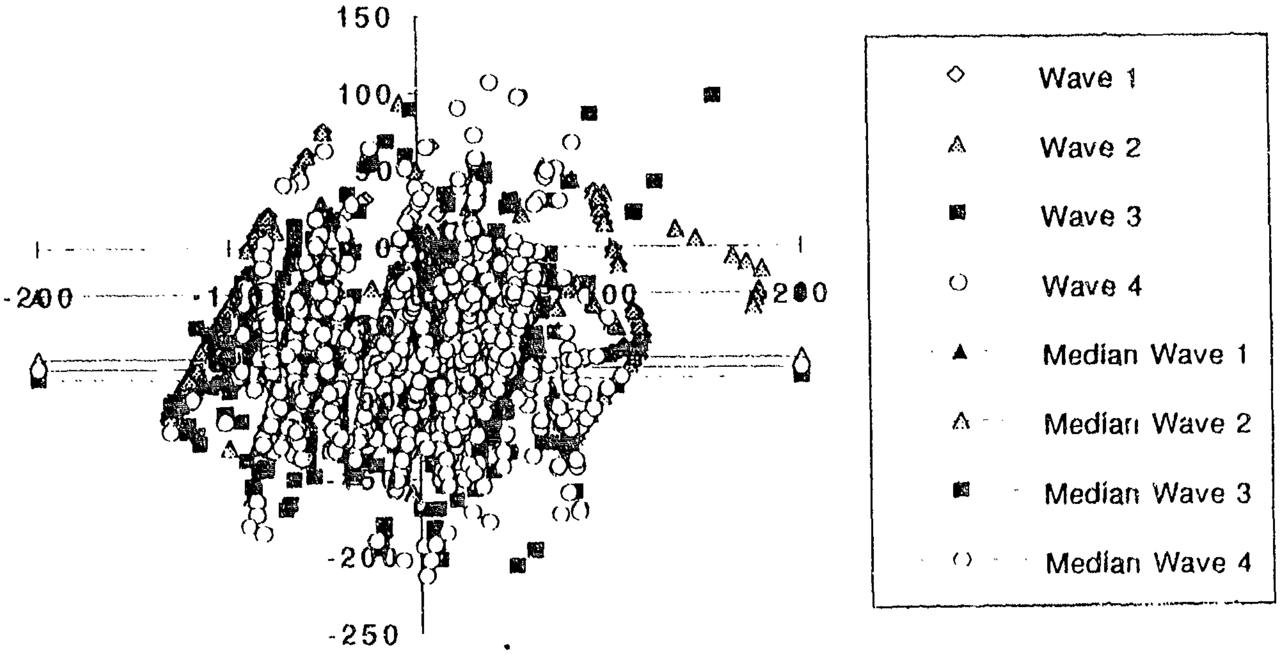

Figure 9. Hostile downed positions by wave.

Figure 10 shows the same data as Figure 9, but identifying the destroy points as coming from either the high- or low-difficulty scenarios. The distance between the medians of the two conditions along the ordinate is approximately 10 nm.

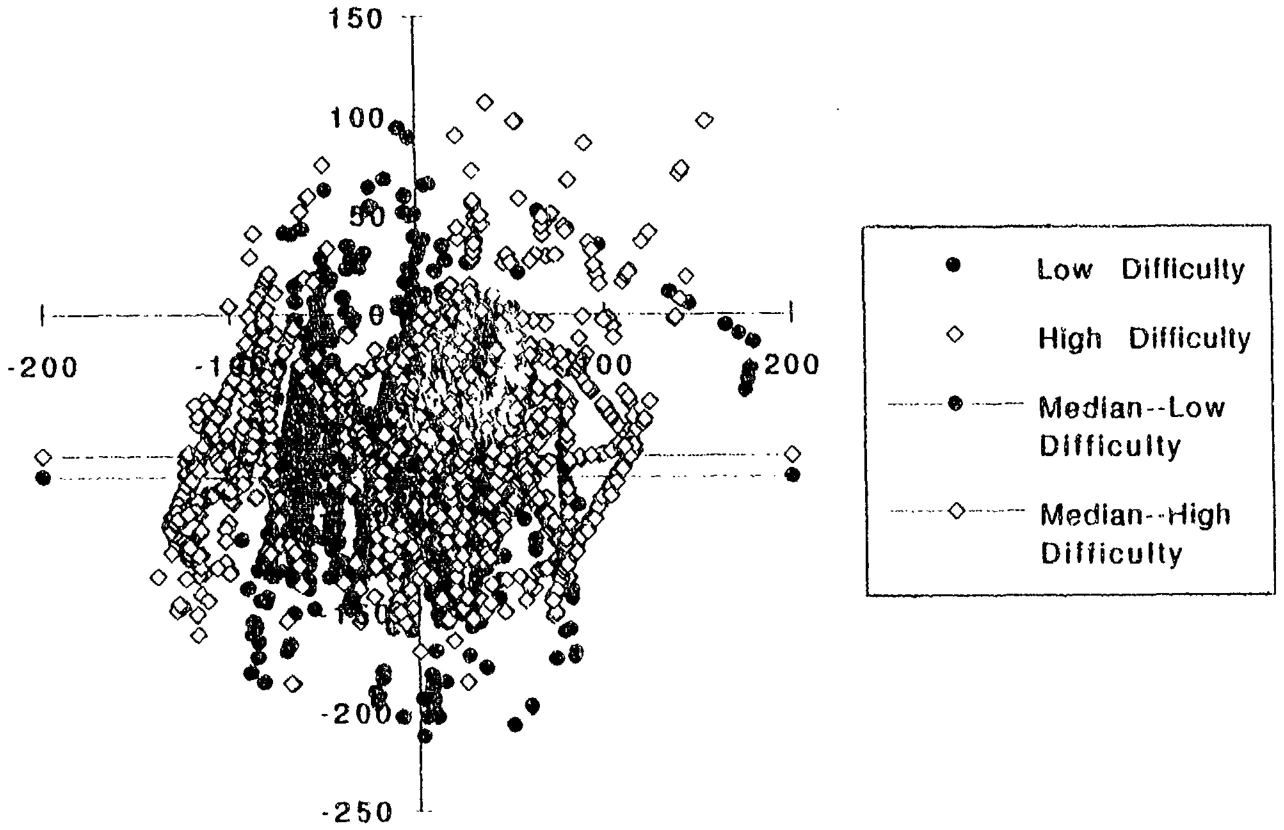

Figure 10. Hostile downed positions--low vs. high difficulty.

8. What was the total number of misidentifications, if any?

Teams misidentified a number of aircraft under each scenario. The most common type of misidentification was the failure to positively identify hostile tracks. Most were identified as unknowns. Using the weighted identification matrix to arrive at an identification score for each WD team under each scenario, a statistical analysis showed only a significant difference in scenario difficulty, $F(1,9) = 13.6$, $p = 0005$. The means for the high- and low-difficulty scenarios were 71.8 and 56.8 respectively.

To understand why the teams were better at identifying aircraft when they were under more stress and had more difficult problems to solve, the numbers of each type of aircraft were examined. It was hypothesized that teams countered the more difficult scenarios by using more interceptors, which could have had two effects. More interceptors would provide more opportunities to identify aircraft and would increase the load on memory and attentional processes. Since interceptors were requested by the WD team, they would know what base they would depart from and when and where they would appear in the scenario. This knowledge would make interceptors easy to identify and as a result would increase their scores. The second effect of more aircraft is an increase in workload. The logical way to reduce workload is to identify more tracks and use the WD computer's symbology as an external memory aid.

Both of these hypotheses were confirmed in examining the number of aircraft identified in each category. Table 6 shows that more aircraft of all categories were identified under high workload, 98.03 compared to 84.77, and that higher scores were the result. Friendly aircraft were identified 15.9% more under high than under low difficulty increasing their score by 4.78, which supports the first hypothesis that more opportunity leads to higher scores. Also 11.5% more hostiles were identified under high than under low difficulty leading to a score increase of 9.61. This finding implies that WDs do understand and use their workstations to prevent and reduce cognitive overload.

TABLE 6. AIRCRAFT MISIDENTIFICATION SUMMARY

Category	Workload	Average Identifications	Average Score	High-Low Score Difference
Friendly	High	54.00 (15.9%)	39.89	4.78 (13.6%)
	Low	46.61	35.11	
Hostiles	High	40.67 (11.5%)	29.39	9.61 (48.6%)
	Low	36.47	19.78	
Unknowns	High	3.36 (98.8%)	2.50	0.56 (28.9%)
	Low	1.69	1.94	
All	High	98.03	71.78	14.95
	Low	84.77	56.83	

9. Where were the hostile aircraft first detected?

In the AWACS, the surveillance section has primary responsibility for detecting and identifying tracks. To simulate this activity for the WD team, most tracks appeared in the display with symbology. The few hostile tracks appearing without symbology had, in all cases except one, symbology placed on them by the system within the first minute of their existence. Therefore, with so few opportunities to respond, these data could not be effectively scored in the scenarios.

10. What were the friendly loses due to friendly fire?

Friendly losses due to friendly fire did not show any statistical differences for the independent variables.

11. What were the friendly losses due to fuel depletion?

Friendly losses due to fuel depletion did not show any statistical differences for the independent variables.

DISCUSSION

At the Mission Effectiveness level, six of the ADC Report questions, 33 dependent measures, were amenable to statistical analysis. Of these measures, six showed a scenario difficulty effect, four showed a learning effect (days), and eight showed a day-by-difficulty interaction. In no case did Seldane or Benadryl differ from the placebo group. Loss ratios showed that high-difficulty scenarios were more difficult than low-difficulty scenarios and that performance improved across days. These performance results for scenario difficulty were supported with subjective estimates of workload (difficulty).

Benadryl was included in the study as a positive control to assess the sensitivity of the dependent measures. Plots of many of the dependent measures appeared to show its degrading effect on performance, but statistically it failed to achieve significance. With only four teams per group, wide variabilities in scores, and non-interval data in some cases, the power of parametric tests for detecting differences is marginal. Reducing these measures to the level of individuals will increase the number of subjects, and may then show statistically significant performance degradation for the Benadryl subjects.

OPERATIONAL INTERPRETATION

After reviewing the data, group means, and statistical analyses, the SME who served as SD during all simulations developed a scoring system weighting 12 of the dependent measures. With the resultant composite score and the SD's subjective evaluation from an operational perspective, the SME determined four distinct findings.

First, Seldane did not affect the WD team's performance. It closely mirrored the performance of the placebo control group. Second, Benadryl decreased WD team performance, but only for the first day that it was administered, and only under high difficulty. Third, the high- and low-difficulty scenario manipulations were successful. High-difficulty scenarios caused an increase in performance errors. Fourth, all WD teams showed a learning effect. As they learned from each scenario, their performance improved.

Scientific Interpretation

Drug Effects

Seldane had no effect on Mission Effectiveness measures, compared to placebo. Benadryl had no statistically reliable effects on Mission Effectiveness measures, but did impact operational effectiveness as determined by an SME with 8 years experience as a WD instructor. This determination resulted from a review of the trends in the data and subjective impressions of the team's performance during the simulations. All of the observed trends failed to reach statistical significance because of the data's high variability within the teams. Without knowing which teams received which drug, the SD correctly identified 3 out of 4 teams on Benadryl. This judgment received further support from effects on cognitive skills and abilities as measured by the AWACS-PAB, especially on the first day of Benadryl administration, day 3 (Nesthus, 1991). The Benadryl subjects' subjective assessment of fatigue was greater on day 3.

Scenario Difficulty Effects

Performance generally degraded under the high difficulty scenarios. This trend was true for three variables across all three days, for five variables on two days, and for six variables on one day. One variable showed performance degradation under low difficulty on day 2. No explanation was uncovered.

One variable, misidentifications, showed an increase under high difficulty. This finding was explored in the **Results** and found to be the result of more friendly interceptors and of an attempt by the WDs to reduce workload by using the workstation as a memory aid. Since the enemy penetrations, aircraft losses, and other outcome measures of the overall scenario showed degradation under high difficulty, an interesting hypothesis arises. If weapons directors are time-limited and can spend time either identifying or directing aircraft, it could be that under high-difficulty scenarios they are making the wrong tradeoff. Individual subject data will be assessed specifically to answer this question.

Learning Effects

End-of-the-week debriefings confirmed that subjects viewed each scenario as unique. Mission Effectiveness measures generally improved across days showing a learning effect. Four variables showed significant improvement across days. Kill ratio measures, percent loss of airbases, and percent loss of SAR operations all improved.

Time-of-Day Effects

Performance on the morning simulations did not differ from that in the evenings, with difficulty balanced, even though subjective fatigue measures were higher during the evening simulation.

FUTURE DIRECTIONS

Data Analysis

The next step in data analysis involves developing rules for assigning individual WD responsibility within each scenario. These rules or definitions of areas of responsibility follow from the WDs' training and practice. Once developed, each individual's role in "winning the war" can be assessed. This assessment will include how well WDs control their own area of responsibility (AOR), how they assist others, and how they request assistance from the WD team. Through this approach the team's performance can be understood as a combination of individual efforts that either support or block the attainment of team goals. After the outcome measures of individual performance are obtained, process measures on the WD tasks and subtasks that produce the outcomes will be assessed. These measures will assess how well the individuals and teams accomplish such tasks as committing interceptors to targets, passing information to pilots, conducting intercepts, maintaining coverage of CAP points, maintaining situational awareness, etc.

We anticipate that the performance of individual WDs will show degradation with the Benadryl antihistamine and with the difficult scenarios when compared with placebo condition. We do not anticipate any performance degradation with the Seldane antihistamine.

RECOMMENDATION

Seldane appears to have little effect on aircrew performance related to mission effectiveness of non-flight deck personnel and should be considered for use under operational conditions as an aid in the reduction of seasonal allergies or nonallergic rhinitis symptoms. From a performance standpoint, the prohibition of Benadryl and other centrally active antihistamines should continue.

REFERENCES

Betts, T., Markman, D., Debenham, S., Mortiboy, D. and McKevitt, T. (1984). Effects of two antihistamine drugs on actual driving performance. British Medical Journal, 288, 281-282.

Bhatti, J. Z. and Hindmarch, I. (1989). The effects of terfenadine with and without alcohol on an aspect of car driving performance. Clinical Experimental Allergy, 19, 603-611.

Boggs, P. B. (1987). Clinical Experience with Terfenadine Worldwide Trials. Proceedings of Current Pharmacotherapy for Allergic Rhinitis and Urticaria. The Journal of Respiratory Disease, 7, 36-46.

Clarke, C. H., and Nicholson, A. N. (1978). Performance studies with antihistamines. British Journal of Clinical Pharmacology, 6, 31-35.

Cohen, A. F., Hamilton, M. J., & Peck, A. W. (1987). The effects of acrivastine (BW825C), diphenhydramine and terfenadine in combination with alcohol on human CNS performance. European Journal of Clinical Pharmacology, 32, 279-288.

Dyer, J. L. (1986). Annotated Bibliography and state-of-the-art review of the field of team training as it relates to military teams. ARI Research Note 86-18, ARI Field Unit at Fort Benning, Georgia.

Eddy, D. R. (1989). Selected team performance measures in a C^3 Environment--An Annotated Bibliography. Technical Report USAFSAM-TR-87-25, USAF School of Aerospace Medicine, Brooks AFB, Texas.

Eddy, D. E. (1990). Performance-based measures in individual and team complex decision making. Proceedings of the Seventh Annual Workshop on Command and Control Decision Aiding. Air Force Institute of Technology (AFIT). Wright-Patterson AFB, Dayton, Ohio.

Eddy, D. R., and Shingledecker, C. A. (1988). Synthesis of performance measures for the C^3GW. In: Delivery Order 0001, CLIN 0002, CDRL Sequence 14 of Contract No. F33615-87-D-0601: Interim Report on Research and Development. Brooks AFB, TX.

Fink, M., and Irwin, P. (1979). CNS effects of the antihistamines diphenhydramine and terfenadine (RMI 9918). Pharmacopsychiatry, 12, 35-44.

Gaillard, A. W., Gruisen, A., and de Jong, R (1988). The influence of antihistamines on human performance. Journal of Clinical Pharmacology, 35, 249-253.

Goetz, D. W., Jocobsen, J. M., Murnane, J. E., Reid, M. J., Repperger, D. W., Goodyear, C., and Martin, M. E. (1989). Prolongation of simple and choice reaction times in a double-blind comparison of twice-daily hydroxyzine versus terfenadine. Journal of Allergy and Clinical Immunology, 84, 316-322.

Hartel, C. (1988) Expanded complex cognitive assessment battery (CCAB): Test descriptions. AAC-UM-33221, Systems Research Laboratory, U.S. Army Research Institute, VA: Alexandria.

Intano, G. P., Howse, W. R., and Lofaro, R. J. (1991). Selection of an experimental test battery for aviator cognitive, psychomotor abilities and personal traits. ARI for the Behavioral and Social Science's interim report, ARI-RN-91-21, Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Intano, G. P. and Lofaro, R. J. (1990). Initial Validation of the Army Aviator Candidate Classification Algorithm. Human Factors Society Bulletin, 33(4), 1-3.

Kubula, A. L. (1978). Problems in measuring team effectiveness (Report No. HUMRO-PP-2-78) Alexandria, VA: Human Resources Research Organization (DTIC No. AD A049 560).

Kulshrestha, V. K., Gupta, P. P., Turner, P. and Wadsworth, J. (1978). Some clinical pharmacological studies with terfenadine, a new antihistamine drug. British Journal of Clinical Pharmacology, 6, 25- 29.

Mann, K. V., Crowe, J. P., and Tietze, K. J. (1989). Nonsedating histamine H1-receptor antagonists. Clinical Pharmacology, 8, 331-344.

Meltzer, E. O. (1990). Antihistamine- and decongestant-induced performance decrements. Journal of Occupational Medicine, 32, 327-34.

Moskowitz, H. and Burns, M. (1988). Effects of terfenadine, diphenhydramine, and placebo on skills performance. Cutis, 42, 14-18.

Nesthus, T. E., Schiflett, S. G., Eddy, D. R., Whitmore, J. N. (1991). Comparative effects of antihistamines on aircrew performance of simple and complex tasks under sustained operations. Technical Report AL-TR-1991-104, Armstrong Laboratory, Brooks AFB, Texas.

Nicholson, A. N., Smith, P. A., and Spencer, M. B. (1982). Antihistamines and visual function: Studies on dynamic acuity and the pupillary response to light. British Journal of Clinical Pharmacology, 14, 683-690.

Nicholson, A. N., and Stone, B. M., (1986). Antihistamines: Impaired performance and the tendency to sleep. European Journal of Clinical Pharmacology, 30, 27-32.

Perez, W. A., Masline, P. J., Ramsey, E. G., and Urban, K. E. (1987). Unified tri-services cognitive performance assessment battery: Review and methodology. Technical Report AAMRL-TR-87-007, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio.

Reeves, D. L., Thorne, D. R., Winter, S. L., and Hegge, F. W. (1989). The Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB). II. Hardware/Software Design and Specifications. JWGD[3] MILPERF REPORT NO. 89-1, NAS, Pensacola, FL; WRAIR, Washington, D.C.

Reid, G. B., and Nygren, T. E., (1988). The subjective workload assessment technique: a scaling procedure for measuring mental workload. In Hancock, P. and Neshkati, N. (Eds.) Human Mental Workload, The Netherlands: North Holland.

Samet, M. G., Marshall-Mies, J. C., and Albarian, G. (1987). Expanded complex cognitive assessment battery (CCAB): Final test administrator user guide. AAC-UM-33212, Analytical Assessments Center, CA: Los Angeles. Deliverable under Contract MDA903-84-C-0449 to U.S. Army Research Institute, Alexandria, VA.

Schiflett, S. G., Strome, D. R., Eddy, D. R., and Dalrymple, M. A. (1990). Aircrew evaluation sustained operations performance (AESOP): A triservice facility for technology transition. Technical Paper USAFSAM-TP-90-26, USAF School of Aerospace Medicine, Brooks AFB, Texas.

Schilling, J. C., Adamus, W. S., and Kuthan, H. (1990). Antihistaminic activity and side effect profile of epinastine and terfenadine in healthy volunteers. International Journal of Clinical Pharmacology, Therapy and Toxicology, 28, 493-497.

Spector, S. L. (1987). The pharmacology of antihistamines. Proceedings of Current Pharmacotherapy for Allergic Rhinitis and Urticaria. The Journal of Respiratory Diseases, 7, 24-28.

Strome, D. R. (1990). A unique command and control research facility. In: Proceedings of the Seventh Annual Workshop on Command and Control Decision Aiding. Air Force Institute of Technology (AFIT). Wright-Patterson AFB, Dayton, Ohio.

Thorne, D. R., Genser, S. G., Sing, H. C., and Hegge, F. W. (1985). The Walter Reed Performance Assessment Battery, Neurobehaviorial Toxicology and Teratology, 7(4), 415-418.

White, J. M. and Rumbold, G. R. (1988). Behavioural effects of histamine and its antagonists: a review. Psychopharmacology, 95, 1-14.

Woodward, J. K. (1990). Pharmacology of antihistamines. Journal of Allergy and Clinical Immunology, 86, 606-612.

*U.S. GOVERNMENT PRINTING OFFICE 1992-0-661-049/60032