Technical Report Documentation Page

1. Report No. 18 19 FAA-AM-78-39	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle 6 Air Traffic Controller Health Change Study		5. Report Date 11 December 1978
		6. Performing Organization Code
7. Author(s) 10 Robert M. Rose, [redacted] C. David Jenkins, Ph.D.; Michalel W. Hurst, ~~Ed. D.~~		8. Performing Organization Report No. 12 819p.
9. Performing Organization Name and Address Boston University School of Medicine Boston, Massachusetts		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 15 DOT-FA73WA-3211 new
		13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591		
		14. Sponsoring Agency Code FAA

15. Supplementary Notes
This research was conducted under Contract No. DOT-FA-73WA-3211

16. Abstract
The main goal of the Study was to determine the nature and extent of health changes in the Air Traffic Controllers and by what charactesistics these health changes might be predicted.

17. Key Words	18. Distribution Statement
Air Traffic Controller Health Change	Document is available to the public through the National Technical Information Service Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	809	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Ø61 250

JB

AIR TRAFFIC CONTROLLER

HEALTH CHANGE STUDY:

A Prospective Investigation of Physical, Psychological and Work-Related Changes

*Robert M. Rose, M.D.
C. David Jenkins, Ph.D.
Michael W. Hurst, Ed.D.

Boston University School of Medicine

August, 1978

A report to The Federal Aviation Administration on research performed under CONTRACT No. DOT-FA73WA-3211 awarded to Boston University. The contents of this report are the responsibility of the authors and do not necessarily reflect the views of Boston University, the Department of Transportation, or the Federal Aviation Administration.

*Presently Professor and Chairman, Department of Psychiatry and Behavioral Sciences, The University of Texas Medical Branch, Galveston, Texas

PERSONNEL OF THE AIR TRAFFIC CONTROLLER
HEALTH CHANGE STUDY

PRINCIPAL INVESTIGATORS

Rose, R.M., M.D.
Jenkins, C.D., Ph.D.
Hurst, M.W., Ed.D.

PSYCHOLOGISTS

Abplanalp, J., Ph.D.
Hefele, T., Ph.D.
Kavanagh, M., Ph.D.
Lomastro, J., Ph.D.
Lynch, J.P., Ed.M.
Suzman, R., Ph.D.

PHYSICIAN

Kreger, B.E., M.D.

CONSULTANTS

Barrett, J., M.D.
Herd, A., M.D.
LeClair, N.
Miller, N., M.A.
Rich, E., M.S.
Samloff, I.M., M.D.
Singer, M.T., Ph.D.
Sturdevant, R., M.D.
Zyzanski, S., Ph.D.

FIELD TECHNICIANS

LaFleur, A., Head
Brown, R.
Donohoe, S.
Jensen, K.
Kennedy, K.
Reiss, R.
Sansone, M.

LABORATORY

Livingston, L., Head
Acebedo, R.
Courtney, M.
Fleishman, M.
Grady, C.
Gutzmore, J.
Hoffman, J.
McDonnell, M.J.
Nelson, F.
Parker, R.
Sandwisch, D.
Seidel, M.
Simundza, S.
Stechowych, J.
Tacey, R.
Whelan, M.

COMPUTER/PROGRAMMING STAFF

Morris, K., Head
Anderson, L.
Cohen, J.
Koolkin, L.
McDonnell, M.J.
Vaughan, T.
Vasak, T.
White, H.

STATISTICIAN

DiScala, C., Ph.D.

EDITOR

Levin, M.

ADMINISTRATIVE

Cermak, J.
Morgenthal, E.

SECRETARIAL

Anckle, M.
Bourneuf, S.
Brelvi, P.
Cotton, F.
Nance, B.

DATA ENTRY STAFF

Black, A.
Davis, C.
Friedman, S.
Geary, J.
Hingston, B.
McCleary, C.
McGonagle, M.
Moore, K.
Stein, M.

AIR TRAFFIC CONTROLLER HEALTH CHANGE STUDY REPORT CONTENTS

I. SUMMARY

Acknowledgment

The inspiration and support for the Air Traffic Controller Health Change Study came from many sources. The need for such a study was clearly one of the recommendations of the Corson Report (1970). The questions regarding the implications of air traffic control work on employee health were, and remain, an important issue for the Office of Aviation Medicine of the FAA. Factors leading to health change in air traffic controllers were of great interest to many individuals in the agency. Administrators responsible for the maintenance of the air traffic control system and those in the Office of Personnel were vitally interested in the problems of air traffic controllers and the factors that might represent a risk for their health change.

We received a great deal of support from many individuals in many different parts of the FAA, both nationally and at different facilities, ranging from their backing for the development of such a study to their crucial suggestions for design of the study and their help in its execution.

We are also much indebted to the administrative officers of the Professional Air Traffic Controller's Organization (PATCO). Without their support and endorsement, it would have been impossible for us to complete this study.

We are perhaps, however, most indebted to the air traffic controllers who participated at a cost to themselves, but yet with enthusiasm, openness and sustained dedication. Without their cooperation, investment, perserverance, and honesty, there would have been no study. Many individual controllers also participated in making suggestions about the various questions that might be asked, helped educate us about some of the intricacies of air traffic control work and helped us to overcome the difficulties in scheduling. Because our study required controllers to be away from their job at least once every nine months, we are also indebted to the various facility chiefs and their assistants who helped us schedule the controllers' visits to Boston.

An enormous amount of information was collected on the 416 men that we studied for three calendar years. Our participants were an unusually cooperative and dedicated subject group, representing an 80% volunteer rate from a stratified sample of ATCs in New York and New England. They averaged a 95% return rate on the monthly mood and health check lists. Although, of the 416 men who started the study, many were either transferred to other facilities or promoted, 388 men, or 93%, of the original group returned for a final examination visit. They came from across the

country and as far away as Guam and Hawaii. Only 24 men dropped from the study for personal reasons during the entire three years, and even some of them came back for a final visit.

We are grateful for the excellent cooperation we received during our visits to the facilities to study the men while they were actually controlling traffic. We were able to complete 2,626 men-days of observation during the course of the study, each day lasting approximately 5 to 6 hours. As collection of data on a man took place every 20 minutes during this time, we made almost 40,000 individual observations on controllers while at work.

Nature of the Study

The main goal of the study was to determine the nature and extent of health changes in the air traffic controllers and by what characteristics these health changes might be predicted. We were interested in both physical and psychological health changes. As documented in the body of the report, we established rigorous criteria for both classes of health changes, and defined health change operationally in such a way that the reader could understand clearly what was classified as an illness episode.

Many variables were included in the study as possible predictors. We selected men with as broad a range of experience and years of age as possible. We selected men from two major centers, the Boston Center in Nashua, New Hampshire and the New York Center in Islip, New York, and from Boston Tower (Logan Airport) and the Common IFR Room (Kennedy Airport), providing radar control for the entire New York metroplex area. We also studied in smaller facilities located at Bradley International Airport (Windsor Locks, Conn.), and at two enroute radar centers (Quonset Naval Air Station at Quonset Point, Rhode Island and Otis Air Force Base at Cape Cod, Mass.) and also included a few men from Providence Airport in Rhode Island.

The data used to determine health change were gathered in several ways. Men filled out a medical history questionnaire reviewing personal and family health histories, health habits and symptomatic complaints and had an interview and physical exam with the study physician five times over the course of the study. During the same visit, chest X-rays and electrocardiograms were made, blood chemistries were measured and urinalyses performed. We learned about most of the mild and moderate illnesses and injuries by sending each man a monthly health checklist on which to report illness episodes, continuing health problems or injuries that occurred during the past month, as well as whether he saw a physician or was hospitalized. Psychiatric health change was assessed primarily in two ways: first, by use of a widely accepted structured psychiatric interview, the Psychiatric Status

Schedule (PSS), and second, by use of a monthly checklist designed to detect troubling episodes of depression or anxiety.

From these sources of information, we were able to diagnose old and new cases of hypertension, along with the occurrence of major illnesses such as peptic ulcer, diabetes, etc. The vast bulk of the physical health changes, however, related to more short term respiratory illness and accidents which were associated with moderate degrees of disability.

The major clusters of predictor of risk variables included stable personality characteristics to determine whether or not particular traits raised one's risk for developing physical or psychological problems. We were particularly interested in men's attitudes and perspectives about work, such as their feelings about the "costs" of working as a controller in terms of its interference with the rest of their lives, their feelings of investment in the work, concerns about burnout, ability to bounce back after periods of being away from working the boards, anxiety at work, and ability to cope with the problems of being a controller. We also employed a number of standard questionnaires to assess worker satisfaction and perception of on-the-job relationships in comparison with other groups. We were interested in how competent, amicable and "ideal" as a work team member each of our participants was as perceived by his coworkers. We also inquired about life changes individuals experienced throughout the study inasmuch as major events such as deaths, marriages, and divorce have been shown to place individuals at increased risk for physical and psychiatric illness.

The other major category of predictors assessed were responses to air traffic controlling itself. While they were on the job actually separating aircraft, we measured the controllers' blood pressure and endocrine responses at work and rated their overt behavior or arousal while controlling. Work load variables were also recorded, including the peak number of planes controlled, complexity of the controlling task in terms of planes transitioning, type of sector worked, type of position worked (radar, hand off, data, training or other) and number of observation points on the boards during the study day. Thus, we were able to compare men with respect to workload on the days they were studied, as well as on their subjective responses to work, changes in overt behavior, blood pressure or hormone levels.

The design of this study also allowed comparison of those who were promoted, those who were disqualified for medical or psychiatric reasons, and those who remained as active journeymen controllers to see if any predictor variables could discriminate among these outcomes. We also defined burnout operationally and determined what characteristics accompanied it.

Our main focus was to search among this rather large array of potential predictors for what was indeed associated with health change.

Most of these variables were studied in a predictive fashion using data collected before the person manifested health problems. Some of the data, however, was collected over the course of the study, and for this data, any relationships found with health changes were concurrent associations.

Health Changes

The largest single chronic illness condition among air traffic controllers was hypertension. The findings of this study parallel previous reports which documented the increased prevalence and incidence of hypertension in controllers compared to other individuals. Regardless of the criterion used to define hypertension, the ATCs had a significantly increased prevalence and incidence of hypertension over comparison groups. Using a rigorous definition the average of two readings over criterion levels on two successive visits to the physician, we found that about 32% (135) of the men entering the study already had borderline or definite hypertension. In addition, 36 men developed hypertension during the course of the study. When other criteria were used such as that of the Framingham Heart Study, the controllers showed a substantially increased rate of hypertension compared to similar age males in this comparison population. We concluded that there was an increased risk for hypertension among the air traffic controller population. Some of this increased risk was associated with (but not necessarily caused by) cardiovascular responsivity to work, as will be discussed later.

A second major category of health change relates to the annual rate of mild to moderate illness. The average number of acute health change episodes for the controller population was approximately 2.5 per year. However, individuals varied significantly with approximately 20% of the men averaged more than five episodes per year and another 20% having had less than one episode per year. These illness episodes were primarily accounted for by upper respiratory infections, minor or moderate injuries, non-specific viral disorders, and acute gastrointestinal syndromes. Although the occurrence of these mild or moderate illnesses did not present a significant risk for future mortality, they were associated with significant numbers of days of restricted activity and could be expected to be a major source of absenteeism.

Although numbers were not large enough to facilitate significant statistical comparisons, we did find that there was a slightly increased risk for developing peptic ulcer in our population. We were unable to study new cases of ulcer disease systematically in terms of what variables were predictive of onset.

Although the methodology for determining psychological health change is not as standardized as that for diagnosing physical health change, nevertheless we employed a methodology which provided objective assessments by psychology professionals of the prevalence and incidence of psychiatric problems among the men studied, so that we could contrast men who did and did not develop significant psychiatric problems during the course of the study.

About half of the men in our study had at least one psychiatric problem, as defined by our criteria, although most of these did not receive professional treatment. The most prevalent psychiatric difficulty was impulse control disturbances, reflecting the relatively high number of men who manifested an inability to control overt anger, anti-social impulses, or illicit drug use. Alcohol use was quite high, but alcohol abuse was about the same as derived from other survey data. Anxiety and depression were experienced at levels equal to or less than that experienced in general populations of non-patients. The experience of psychiatric problems was not particularly related to FAA performance criteria. Mate role impairment because of psychiatric problems was experienced at approximately the same rate as among an urban community group, and therefore, we would conclude that mate role difficulties caused by psychiatric problems were not more common among air traffic controllers than among others.

Predictors and Correlates of Health Change

Hypertension

We tested many variables to establish their ability to discriminate men who developed hypertension during the course of the study from those who remained normotensive.

One of the best discriminators was the systolic blood pressure responses to work. The men who showed significantly elevated systolic and diastolic blood pressure on the job during their first field study had higher risk of developing hypertension. This was prior to the diagnosis being made by the internist. We also found that future hypertensives were characterized by having <u>less</u> hard driving, more Type B personality, and by having experienced less life change in the past nine months than the individuals who remained normotensive. As a group they were somewhat lower in marital resources. We also found that although their blood pressures in the clinician's office were still normotensive during their first visit, they did show significantly elevated blood pressures compared to those who remained normotensive.

Not only at the first field study, but at field studies throughout the three-year surveillance, the individuals who developed hypertension during the course of the study showed blood pressure responses to work that were greater than the responses of those who remained normotensive. Their average systolic and diastolic blood pressure at work was significantly elevated. Their blood pressure responses to and at work were just as high as those who already were hypertensive when they came into the study. When the men did more work, those who developed hypertension also exceeded their own average blood pressure level by a greater amount than those who remained normotensive.

When we took into account all factors that were significant predictors of hypertension, i.e., psychological variables, blood pressure levels in the physician's office on intake examination, as well as blood pressure responses to work, we could correctly predict 86% of new cases of hypertension and 82% of those who remained normotensive.

Although these results were clearly highly significant, some qualifications must be added. It is inappropriate to conclude that hypertension is caused by air traffic control work per se. Our interpretation of these findings is that for individuals who are predisposed to developing hypertension by reason of a host of possible genetic and biological factors, exposure to air traffic control work increases the risk or perhaps hastens the rate of their developing hypertension. Given the increased prevalence and incidence of hypertension among air traffic controllers and the evidence that those who develop hypertension have greater physiological responsivity to the work environment, an appropriate conclusion might be that air traffic controlling represents a risk factor interacting with other risk factors in development of hypertension.

It is also important to note several factors that were not associated with the risk for developing hypertension. Controllers who developed hypertension were considered neither more nor less competent than those who did not. They did neither more nor less work than those who remained normotensive, nor did they spend either more or less time on position during the days they were monitored. With respect to the quality of their work, they were neither more nor less likely to receive awards, nor promoted more or less than those who remained normotensive.

However, it should be noted that developing hypertension carried with it a risk for future disqualification. The issue of whether or not controllers should be disqualified for being hypertensive or for being treated for hypertension should be reconsidered. We found that the significant number of men who were hypertensive did as much work on study days and received as high a frequency of competency nominations as those who remained normotensive. We concluded that hypertension per se did not interfere with a person's abilities to control traffic.

Mild and Moderate Health Changes

Although the illnesses and injuries that comprise mild and moderate health changes represent a much less serious risk to future health compared with hypertension, they do account for a very significant amount of time lost from normal activities.

Four diagnostic categories accounted for almost two-thirds of all mild and moderate health changes - respiratory, acute gastrointestinal, non-specific viral syndromes and injuries. As increased incidence of these different mild and moderate health changes was associated with many similar factors, we combined their incidence into an average annualized rate of mild and moderate health change and focused on determining those factors that predicted differences in this rate. We found that there was a tendency for men in the high illness group to be younger and to have less experience with the FAA for their ages. Men in the older groups tended to have less minor illness compared to those in the younger group.

The men who experienced more mild and moderate health change were more invested in their work, and they stated that ATC work "cost" them a lot subjectively (expenditure of efforts to adjust). They reported increased symptoms of anxiety before, during and after work. As with the other variables discussed, their ratings of group morale, made before the collection of illness data, was lower. The high illness group also reported least satisfaction with management and consistently rated their supervisors as showing less than adequate amounts of consideration for others.

Men in the highest illness group were least often chosen as the most amicable or as ideal team members by their co-workers. However, they were not considered less competent by peers, indicating that competence *per se* is not associated with the risk for more illness.

Men who had higher illness rates were found to be more Type A, i.e., more competitive, hurried, and achievement-oriented than those with lower illness rates, who were more Type B.

One of the strongest predictors of total mild and moderate illness rates was the amount of life change and life stress encountered in the period of time immediately preceding the collection of illness data. Those with higher illness rates had more life changes that distressed them.

The men with higher illness rates also differed significantly in terms of their cardiovascular and endocrine responses in the field. The high illness group showed significantly lower levels of plasma cortisol at work and were more likely to drop in plasma cortisol levels when there was an increased amount of work, compared to those in the lower illness group. Those men with more illnesses also had lower blood pressure variability at work and, similar to cortisol, showed a diminished responsivity in terms of cardiovascular changes with varying levels of work load. Both of these findings, however, were not predictive in that the lower cortisol and lowered blood pressure variability were observed in the same periods of time as the occurrences of illness. However, men who developed more illness had a lower level of cortisol during the first field study before illness data were collected. Thus, it is likely that the diminished cortisol responsivity is a predictive as well as a correlational finding.

It is also of note that men who showed lower cortisol and blood pressure variability during work and an increased rate of mild to moderate illness also tended to minimize their estimate of the amount of work performed on a given day during field study. In contrast, those with higher levels of cortisol and higher blood pressure variability tended either to report accurately or to exaggerate the amount of work done compared to objective indices of work load.

The fact that individuals who showed lower physiological responses also tended to minimize the amount of work they were doing, and were often among those with more frequent illness, suggests that this increased frequency of illness was not primarily a reflection of a tendency to exaggerate or complain. Other checks that were made supported this conclusion.

We found no relationship between a "faking bad" scale on the California Psychological Inventory (CPI) and various measures of job attitude. This finding suggests that the ATCs responded honestly to questions concerning their job satisfaction and their answers were not part of an attempt to exaggerate difficulties as measured by this scale on the CPI.

We also found that when we compared reported gastrointestinal symptoms with measurements of serum pepsinogen level, individuals who reported continuing gastrointestinal problems had significantly higher levels of serum pepsinogen than others. This supported the interpretation that they were not exaggerating their symptoms, but were experiencing greater GI distress. With respect to upper respiratory infections, we found that among those persons who had greater frequency of this illness, there was increased use of medical care and increased proportion of episodes involving 4 or more days lost from usual activities. In addition, reports of URI symptoms followed known seasonal variations. There was no relationship between the average annual rate of illness and scores on the MMPI hypochondriasis and hysteria scales, indicating that greater frequency of symptoms reported was not a reflection of a psychological propensity for complaint behavior. Actually the air traffic controllers in our study had a lower score on the hypochondriasis scale than the MMPI standardizing population.

To summarize those variables most related to ATC work, men in the highest illness group tended to be more invested in their work but felt that it cost them a lot, had lower satisfaction with management, had lower group morale, and were more dissatisfied with their supervisors.

Psychiatric Problems

Slightly over one-half of the men experienced one or more significant psychiatric problems during the three years. This finding was not anticipated in light of the information gathered at intake when controllers were not found to have a great excess of psychiatric problems except for the relatively high rate of impulse control difficulties. However, as the study continued we found that more men developed problems and although a relatively few number of men showed chronic problems throughout the entire study, a significant number did show some problem at one time or another. These data suggest that the development of a psychiatric problem was not limited to a very small percentage of controllers, but that over time a larger number of the men did develop some difficulty, even though it may have been resolved by the time of a later examination. This may not be a finding specific to air traffic controllers, but we cannot make comparisons with other groups because no other occupational group has, to our knowledge, been studied so intensively over so long a period with the PSS.

Five areas of psychological problems were assessed systematically. The first was subjective distress, involving significant problems of depression, anxiety, disturbed daily function, disturbed eating and sleeping. Over the three years, only 12.5% of the men had, or developed, a problem in this area. Another major type of psychiatric problem assessed was impulse control disturbances: the inability to control overt anger, anti-social impulses, and illicit drug use. Although illicit drug use was not serious among controllers, there were significant difficulties in controlling anger and anti-social impulses. We found that 12.7% of the controllers had significant impulse control disturbances at the first exam at intake, and over the three years, 30% of the controllers developed significant impulse control disturbances at least once. We also evaluated work role disturbances associated with psychiatric problems and assessed difficulty in mate role relationship. Twenty percent of the controllers did experience some mate role problems over the three years in the study.

Alcohol use was high. Over 50% of the controllers were classified as heavy drinkers, a higher rate than that derived from survey data. However, only 7.5% of the men at intake had significant problems in alcohol abuse despite the high use of alcohol often associated with social drinking after work. Only four men had physiological symptoms of alcoholism, a low rate of 1%.

Although these data clearly support the conclusion that there was a significant amount of emotional distress in our controller population, perhaps the impact of this area of difficulty can best be appreciated by the relationship between the findings on our psychiatric evaluations and the rate of medical disqualifications. We found that of all the men who completed all 5 of the psychiatric evaluations in this study, 20 were subsequently medically disqualified by the FAA. Of these 20, 19 were found to have had significant psychiatric problems. Thus, the risk for disqualification was greatly enhanced in those individuals who developed psychiatric problems. Some of those who were disqualified had psychiatric problems at intake and hence, this was a correlational and not predictive finding. However, those who developed psychiatric problems during the course of the study also had a significant risk of being disqualified later and hence a predictive relationship was also suggested.

There were many variables that predicted psychiatric problems. Although the predictors assessed at intake were not uniformly associated with all five problem areas (subjective distress, impulse disorders, alcohol abuse, mate and work role pathology), there were a sufficient number common to more than one area so that a general pattern emerged.

The men who developed psychiatric problems were significantly lower in their satisfaction at work, their satisfaction with co-workers and their ability to discharge tensions following work. Similar to the finding for mild to moderate illness, these men had high investment in work and reported high subjective cost associated with being air traffic controllers. They also reported an increased tendency to cope with stress on the job by drinking after work and showed lower amounts of coping by other activities such as physical exercise.

From the perspective of stable personality characteristics, they also showed more Type A behavior - more competitive and hard driving, and experienced more life change in the period of time prior to the development of problems. Thus, in many ways the men who developed psychiatric problems showed significant similarities with those who had higher rate of mild to moderate illness. One might ask, therefore, if these were the same men. Were difficulties in the psychiatric area and medical area showing up in the same individuals? Was there one healthy and one sick group. This turned out not to be the case. The men who developed psychiatric problems did not have significantly more mild or moderate illnesses, nor did they have more hypertension. We found essentially no relationship between these three major categories of health outcome. That is to say, the men who developed hypertension were no more likely than others to have mild and moderate illnesses and no more likely than others to develop psychiatric problems.

This finding supports the specificity hypothesis in psychosomatic medicine which, simply stated, emphasizes that different individuals will develop different problems despite the fact that they may be exposed to similar difficulties or similar problems in their psychosocial environment. For psychiatric illness and mild to moderate illness, several common themes clearly emerge. The work environment seems to be more implicated in the risk for developing health change than the work activity itself. Men who developed more mild or moderate illness or who developed psychiatric problems did neither more nor less work when observed during the field studies. They spent neither more nor less time on position nor did they control more or fewer planes. It was also noted that they were not considered less competent by their peers, although they may have been less frequently chosen as amicable or less likely to be nominated as an ideal team member. Because of these findings, one may not conclude that individuals who develop more mild or moderate illness such as flu or psychiatric problems were poorer controllers or did less work.

One is led to the conclusion that these individuals are more at odds with their work environment. As noted, they are often highly invested in being controllers although they find themselves less able to discharge tensions associated with work. Hence they find that functioning as controllers is associated with an increased cost in terms of their personal lives and psychological health. This finding of "It's not so much what they are doing as the context in which they are doing it" holds definite implications for changes that might be considered in the work environment to reduce the risk for future morbidity.

Job Outcome

We analyzed our findings to see if we could predict job outcomes that clearly involved important career changes. We looked at individuals who were disqualified for medical reasons during the course of the study, those who were promoted, and those who developed signs or symptoms of burnout.

We compared the disqualification rate for our controllers with that of controllers from the same facilities who were not studied, and found significant differences. It was clear that our controllers experienced a lower disqualification rate as compared to those not studied, although we were not able to control statistically for age and years of experience in the two groups. This finding suggests that at least our men were not sicker or in worse health than those who were not studied and also suggests that our findings are on the conservative side. If we included those other men who were not in our study, we might have found a higher disqualification rate. We also found that the men whom we studied had 1.7 times the promotion rate of those not studied. This finding also suggests that we may well have studied controllers who were as good as, if not more competent, than those in the same facilities who were not studied.

Men in our study group who became disqualified were nominated less often for ideal team choices and were also nominated less often for competence. They showed lower investment in their jobs compared to those who remained active journeymen. They also showed lower marital resources and were different on four scales of the CPI. They had diminished feelings of well being, lower responsibility, lower tolerance, and lower intellectual efficiency. As some of the men were disqualified following the first round, and were already in disqualification procedure, their answers at the entry exam were no doubt in some ways affected by the proceedings underway.

We were able to compare the 49 men who received promotions to supervisor during the course of the study with those who remained in the study and were not promoted. Many factors that predicted promotion were consistent with expectation. The controllers who were promoted were slightly older and had slightly more years of experience. They were highly invested in their job and they were also chosen more frequently by their peers for increased competence as controllers. They showed higher self and group morale and more satisfaction with management policy, specifically satisfaction with promotion and pay. They were higher on the dominance scale of the CPI. They were also more Type A, with more job involvement and more hard driving behavior. However, the controllers who were promoted also manifested increased concern for burnout during the entry examination. They reported that they had less ability to bounce back to full controlling capabilities when the workload changed from light to heavy and were more likely to burnout than those who remained active controllers. This increased concern with burnout occurred at about the same time that they received increased competence nominations. Therefore, one might conclude that although their peers rated them higher than those who were not promoted, the men themselves reported increased concern about burning out and less ability to bounce back on the job.

In summary, these data support the view that the controllers who were promoted were doing a good job as judged by their peers, regarded as more competent, were highly involved in their work, were more satisfied with management, were slightly older and had more experience and were more hard driving. However, it is also noteworthy that they showed increased concern about burnout despite the fact that they were considered more competent by their peers and were chosen for promotion by management.

We were interested in knowing whether or not we could discriminate between those who developed burnout and those who did not by means of data collected in the first two rounds of exam. W found that those individuals who later developed burnout, by our definition, scored significantly higher on a number of positive factors or variables early in the study. They had more vigor, friendliness, and elation on the Profile of Moods State (POMS). They showed less anxiety regarding work and a higher tension discharge rate. They coped by drinking less often and showed less anxiety. They also showed more assertiveness (interested in doing a good job as a controller) and they showed more coping by physical activity. It thus appeared that individuals who later developed burnout showed more psychological health early in the study than the comparison group. They did not start out as dissatisfied or having more anxiety but rather were committed and, from their own estimation as well as that of peers, were functioning as well or better than others.

However, the men who developed burnout by Round 5 showed an increased incidence of subjective distress. It is apparent that although they started out feeling good about themselves and scoring low on psychological distress in the beginning, during the course of the study they showed increased psychological problems.

It is notable that individuals who developed burnout did not show any differences in the average amount of work done or in time spent on position during field studies. Furthermore, they showed no differences in physiological measures taken at work, i.e., cortisol levels or blood pressure responses. Nor did they show a higher incidence of hypertension or higher incidence of mild to moderate health change. From these data it is apparent that their physical health, at least during the three years that they were studied, did not show deterioration. However, as their burnout status was associated with increased incidence of psychological problems, if they were followed for a longer period of time, the effect of increasing psychological problems might place them at risk for future medical disqualification.

In some ways the men who developed burnout showed some of the characteristics observed in the men at intake who were later promoted. This suggests that competent controllers can and do experience concern about burnout and as a matter of fact, air traffic controllers with such concerns may frequently be represented among individuals who are promoted. These findings suggest that burnout is not necessarily characteristic of individuals who are considered less competent controllers, but may be related to an apprehension that one will not be as good as one has been, and will fall in the estimation of one's peers. The anticipation of this fact may function in some ways as a self-fulfilling prophecy such that some of the men may convince themselves that they will not function as well and, in time, it turns out to be correct.

Implications

Our study was oriented toward the task of discriminating between men who showed a deterioration in health and those who remained essentially without symptoms or illnesses in any given category. For some of the illnesses we observed, such as hypertension, we could compare controllers to other groups. With respect to psychiatric illness, this was more difficult as prevalence data is not as readily available.

We did find that the controllers had more hypertension than other groups, and possibly some forms of psychiatric problems were also more prevalent. However, the implications of our findings are based primarily on the differences we observed between men who developed health changes in a given area and those who did not.

Our findings are in many ways different from our expectations, especially with regard to the predictors of psychiatric illness and mild to moderate physical health problems. However, despite the fact that different men had different illness problems and relatively few had no problems at all, several themes emerged from our data. One cannot help but be impressed that controllers who perceived their work environment negatively, who were dissatisfied with work, with their co-workers, or with the FAA, showed a significantly increased risk for developing either psychological disorders or mild medical illness. We did not find that those men who developed more physical or psychological problems spent less time working, nor were they rated less competent by their peers. We did find that they felt estranged or alienated from their work and this occurred despite the fact that they said they enjoyed air traffic controller work. They were usually more invested in doing a good job, even though they could not discharge their tensions well after work, and too often they used alcohol as a way of coping.

These findings suggest that it was not so much what they were doing but the context in which they were doing it and the attitudes and feelings they had about their situation that influenced their risk for health change. We were surprised by the relative importance of attitudes about work in predicting health change. We had expected that the work load itself during field studies would have stronger predictive power than it turned out to have. The consistency of these results, which might be summarized as the alienation controllers experience from their work environment, suggests that changes should be made in this environment and in the way it is experienced. We believe that dissatisfaction with FAA management policies are a significant problem and represent part of the negative set associated with an increased risk for health change.

We also believe that some of the divisiveness that the controllers experience may in part derive from union-management interactions. Thus, despite controllers' perceptions of many positive benefits derived from the growing strength of PATCO, some of the alienation and divisiveness controllers often experience may be an unexpected and unwanted side effect of the adversary relationship between union and management. One of the recommendations for change coming from this study is that attempts be made to improve the work environment, to diminish the degree of dissatisfaction or alienation that controllers feel. Our interpretation is that this is not solely a matter of working hours or of pay, but that there is a need to improve the communication between management and individual controllers and to attempt to limit the adversary nature of their relationship. It is our view that this could be accomplished by a cooperative effort between the FAA and union management and that individual controllers would significantly benefit from such a reduction of controversy in their work environment.

The predictors of hypertension are in a different cluster and have different implications from those predicting psychological problems and mild or moderate illnesses. Air traffic control work *per se* has a closer relationship to development of hypertension than for other illnesses. Individuals who developed hypertension did not report as high a degree of work dissatisfaction or general alienation with management as those who developed other problems. It does appear that for individuals who, for a variety of genetic and as yet poorly understood biological factors are predisposed to develop hypertension, increased workload adds to the risk for developing hypertension. These men do show greater blood pressure responses to increased work and this increased cardiovascular responsiveness to work is a major predictor of developing hypertension. This finding, taken along with the observation that hypertension is generally more prevalent among controllers than among others, suggests the possibility that interventions might be attempted on an experimental basis to try to minimize this increased cardiovascular responsiveness. Perhaps controllers could be taught procedures to minimize cardiovascular responses to work, by using relaxation or biofeedback techniques. Various interventions might be employed on a trial basis to see whether any might be effective.

Other interventions of more proven benefit should be provided on a voluntary basis to attempt to reduce psychiatric morbidity. For example, provision might be made for counseling of individuals who are at increased risk for developing psychological or physical health.

By the time an individual becomes a journeyman controller, he has undergone three to five years of training as a developmental. There has been a considerable investment of time and money to help him acquire his controlling skills. Our data support the notion that controllers are concerned with burning out and being less able to control aircraft efficiently. The period of maximum productivity as controller is a limited one, perhaps 10, 15, but not more than 20 years. Medical disqualification is a problem for the individual controller and represents a significant burden financially to the agency and the government. It is one of our conclusions that interventions be attempted to maximize the period of time that controllers can function productively. A variety of interventions might be considered, and we can do no more than suggest them as our study was not primarily oriented toward a trial of interventions but toward a determination of the nature of the problems and the factors that might predict them. We know that individual controllers who are at increased risk for psychiatric problems can be identified early and provided counseling or treatment. We feel that some selection process might be employed to select controllers who are more mature, with a lower propensity for impulsiveness. Some interventions to reduce blood pressure variability in those controllers who experienced it at work are worth trying in an effort to delay the onset of hypertension. These are strategies designed to help the individual controllers.

Our data also suggest another intervention not oriented toward the individual but toward the relationship of the controller to management and the agency. There seems to be considerable general dissatisfaction among the controllers with FAA management. Those who showed the most dissatisfaction towards work, co-workers, and the FAA had a significantly increased risk for health change. We feel that some union activities may inadvertently have contributed to the alienation and divisiveness many controllers experience in the ongoing adversarial relations with the agency. We therefore suggest that an organizational development program be undertaken with joint union management cooperation to improve work life. If successful, such a change might be most beneficial for reducing the risk for future health change among controllers.

II. INTRODUCTION

A. Background of Problems and Rationale for the Study

This study of health change in air traffic controllers was preceded by years of controversy over alleged difficulties and effects of the work on the health of the controllers. Although the popular press gave dramatic attention to the struggles of air traffic controllers to perform their job through manifold difficulties, emphasizing the negative effects on the health of these men, no comprehensive study of the physical and psychological health changes in controllers had been made. Several events occurring in close proximity, however, focused on a need for such a study and led eventually to the initiation of the present study in 1973.

Although air operations in the United States had increased 100% between 1960 and 1970, the number of qualified controllers increased only 10% during that same period, according to Federal Aviation Administration (FAA) data (Walton, 1977). This discrepancy arose in part because of the length of training required and in part because of a moratorium on hiring during the previous several years. As a consequence, journeyman controllers were required to work extra shifts, long hours, more aircraft per controller, and often without even a lunch break. By 1968-1969, the increase in difficulties resulted in a series of work slowdowns, in which controllers enforced air traffic regulations strictly, regardless of the inconvenience caused. In addition large numbers of men called in sick over a period of several months, exacerbating the effect of the slowdowns. When the labor-management disputes were settled sufficiently to get air traffic moving again, a commission was set up to investigate many of the issues that had been raised. The commission's findings were published in the Corson Report (Corson et al., 1970), and among a multitude of recommendations was one which suggested that the FAA needed to document not only the degree of job stress in air traffic control work, but also the impact this stress might have on the physical and psychological health of controllers.

The Office of Aviation Medicine at the FAA and the Professional Air Traffic Controllers Organization (PATCO) subsequently sponsored several studies in an attempt to answer some of the questions raised in the Corson Report.

One of these studies (Cobb and Rose, 1973) was a retrospective comparison of illness rates among controllers with those of other men involved in the aviation industry. The retrospective comparison was made by examining the medical records of 4,325 air traffic controllers and those of 8,435 other men involved in the aviation industry. Comparable records were available because both groups of men were required to obtain yearly medical examinations from the same group of designated physicians (Aeromedical Examiners) for relicensure in commercial or general aviation. This medical record study revealed a number of important differences between the controllers and the others in disease prevalence (existing cases) and disease incidence (new cases).

The largest difference between controllers and the other aviation group occurred for hypertensive disease. The controllers were found to have 5.6 times the incidence of new cases of hypertension in 1969 to 1970 that the other group of aviation men had. The prevalence of diagnosed hypertension in the air traffic controllers was found to be 4 times that in the control group. Some of the difference in the prevalence rate could have been attributable to a preselection factor since 17% of candidates for relicensure in aviation jobs other than controller had been disqualified because they had pre-existing hypertension, and were not, therefore, in the control group. In contrast, only 3% had been excluded because of pre-existing hypertension from the air traffic controller group. However, the six-fold difference in the incidence of hypertension could not be explained by the possible selection bias. That is, the occurrence of new cases of hypertension was independent, by definition, of any possible bias before relicensure, all men in both groups starting off after relicensure with a clean bill of health.

Controllers were found to have ulcer disease and diabetes more frequently than the control group, although the comparative rates of these diseases were less dramatic than hypertension. These findings strongly suggested that a prospective study should be undertaken to clarify what factors might be relevant to the increased rate of illness from these diseases among controllers, as well as to determine risks for other health changes.

As a result of the Corson Report, the several retrospective studies, and the limitations inherent in existing findings, a five and one-half year contract was awarded to Boston University School of Medicine in 1973 to conduct a biomedical, prospective study to determine the kinds and degree of health changes occurring in air traffic controllers and the extent to which these health changes might be predicted by other factors.

B. General Hypotheses and Specific Issues to be Investigated

The study was designed to assess the individual and interactive relationship between health change and the following major factors, whose relevance was suggested by available information: work environment, years on the job, physiological differences in response to stress, occurrence of significant life events, work attitude and morale, availability and usefulness of psychosocial supports, job commitment, job performance and enduring personality traits and characteristics. Moreover, it was specified that the factual data collected and the interpretation of the data should be in a form readily addressable to the following questions and issues:

1. What is the nature, derivation, extent and significance of health changes among air traffic controllers? How do these health changes affect performance and career longevity?

2. Do controllers experience stress? If so, what is the nature and extent of the stress? How much is related to the job? What causes it? To what job conditions, including social and psychological forces or factors, does it relate?

3. To what extent do management, supervision and manpower management policies, practices, programs or procedures contribute to health changes and especially to the degree of stress present in the air traffic controller job; e.g., does how work is defined, work schedules, hours on duty, and the length of work periods, etc., relate to stress, and if so, how?

4. Are there predictive factors or measures that could be applied before employment of air traffic controllers which would indicate to management officials which candidates might be more prone to deleterious health changes?

5. Are there ways by which individuals can be identified who need counseling or help due to their increased potential risk for illness? Is this risk increased because of their particular personalities, their family problems or the nature of the working environment they are in?

6. How do the stress levels in the air traffic occupation compare with other occupations, e.g., pilots, business executives, etc.? Although the data collected in this study may not provide for direct comparison of air traffic controllers with other individuals in other occupations,

attempts should be made to compare the results that are obtained with those from other investigators studying other occupations.

7. Is there a "burnout" phenomenon? If so, what is it?

   a. Can the individuals who are more prone to early physical deterioration be identified?

   b. Can we identify work situations that are more prone to produce widespread early physical deterioration than other work situations may be, so that early counseling and other career possibilities can be identified and considered?

Although specific questions and issues of concern to the FAA were framed for our information, we were directed to proceed with the study of factors we considered to be relevant. We attempted to heed, however the advice that our data should be presented in a way that would allow these questions and issues to be answered by the sponsors (FAA) and collaborating supporters (PATCO) of the study.

C. Overview of Methodology

Previous Research Findings Providing a Focus for the ATC HCS

The concepts that social, psychological, vocational, environmental and other events may influence health is an ancient one, but more intense interest and increasingly rigorous scientific studies emerged only relatively recently. Flanders Dunbar and other clinicians and researchers who were interested in what is now known as psychosomatic medicine reported numerous clinical case histories from which clues emerged regarding potentially relevant variables to be studied systematically (Alexander, 1934, 1939; Cobb, 1943; Dunbar, 1954; Hinkle, et al., 1957, 1958; Engel, et al.,1956). Based on the many suggestive findings of these earlier investigators, recent rchers have selected and rigorously pursued certain areas of investigation which may be crucial in determining the extent to which social and psychological factors influence health (Weiner, 1977). Included among these crucial areas individual differences in hormonal (Wolff et al., 1964; Mason, 1968; Rose et al., 1968) and cardiovascular (Murawski et al., 1968; Riess et al., 1968; Herd et al., 1969; Reiser et al., 1962) responses to stressful stimuli, the impact of life events (Rahe et al., 1964; Rahe, 1968; Dohrenwend and Dohrenwend, 1974), individual differences in ability to adapt to changes in environmental demands (Carlestam, 1971; Stokols, 1972; Henry, 1956; Sandberg and Bilding, 1976), and individual differences in behavioral life styles (Paffenbarger, 1968; Jenkins, 1971; Jorgensen and Gyntelberg, 1976). Most of these areas have, besides important theoretical underpinnings, an associated methodology for reliable measurement. There is evidence that each of these areas may be related to the risk for future health change. Given reliable methodology and evidence of their relevance to health change, these areas served to focus our selection of variables for this study.

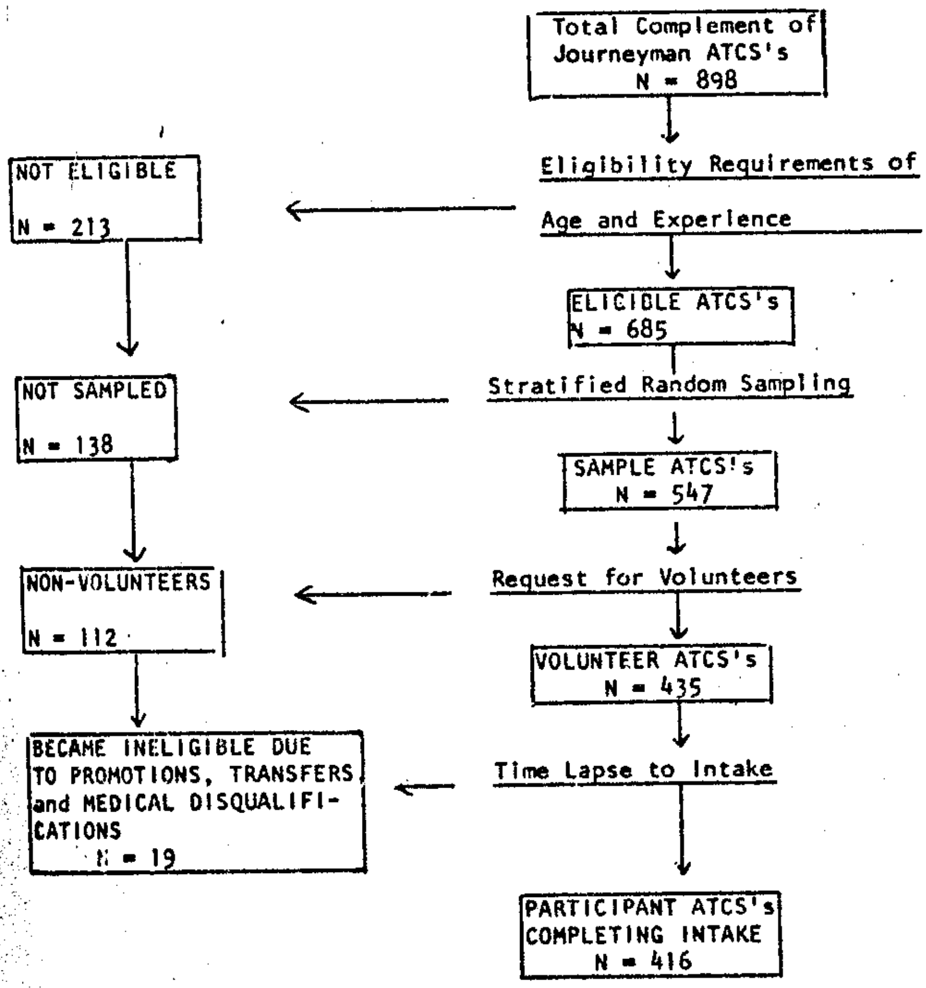
Epidemiological Model - Population at Risk

In order to assess the contribution of biological, psychological and social variables to risk for future illness among air traffic controllers, an epidemiological model was chosen in planning this study. Implementation of the research model would require that the study be prospective and longitudinal in design, with concurrent measurement in all of the participants of all variables considered to be relevant and that the data to be collected would be specified in advance. In addition, all subjects would receive equal surveillance for the detection of any illness.

The study was to be based on a probability sample of journeyman controllers from various facilities in New York and New England. The overall sampling procedure is shown in Table 1. Since we ...

TABLE 1

Overall Sampling for
the Air Traffic Control Specialist (ATCS)
Health Change Study



wished to assess the possible independent influence of age and years of experience on the job upon the risk for becoming ill, we attempted to dissociate age and years of experience, although the two are highly correlated. We therefore sought to include as many younger men as possible with many years of experience, as well as older men with fewer years of experience. In addition, we needed to study men who would continue on the job over the course of the study. Therefore, trainees (who might be washed out) and men who could retire before the end of the study (by becoming 55 years of age or by having more than 20 years of experience) were excluded. Journeyman controllers assigned to positions without continual responsibilities for actively separating aircraft (supervisors, data specialists, and so on) were excluded because they were not exposed sufficiently to the main risk factor of the study: controlling aircraft. Finally men on medical waivers were not eligible because of their pre-existing conditions.

After we applied these criteria, 685 men were eligible for the study. Stratified random sampling was applied to maximize the age and experience dissociation. A sample of 547 journeyman air traffic controllers in selected sites in New England and New York was identified. After extensive briefing on the nature of the study, 435 men volunteered to participate and signed informed consent forms, a 79.5% volunteer rate from the 547 men invited to participate. For various reasons (see Table 1), 19 of the 435 volunteers dropped out and 416 (95.6%) completed their intake examinations. The average age of the 416 participants was 36.2 years with a range of 25 to 49 years. In spite of the differences in sampling, this figure compares well with the average of 36.6 years reported by Karson and O'Dell (1974) for a national sample of 11,047 center and tower controllers. The average years of experience of our sample was 11.2 with a range of 3 to 20 years.

Although it would have been more desirable to study a national sample of controllers, the logistics of the study dictated that facilities from which men were drawn should be restricted to the New York and New England regions. The participants were drawn to represent both major categories of facilities where most controllers work — centers and towers (airports). Among the Air Route Traffic Control Centers (ARTCCs) responsible throughout the country for separating aircraft en route from one airport to another, the Center located in Nashua, New Hampshire controls most of the air space over New England. One hundred and seven men (26% of our sample) volunteered from this facility to participate in our study. The other Center from which men were selected to participate is responsible for the air space between the New England and Washington Centers, including New York, most of Pennsylvania and New Jersey and parts of other states. The New York ARTCC is located in Islip, Long Island, and is one of the busiest Centers in the country, averaging approximately 40% more operations per year than in New

England. The New York ARTCC provided 41% of our sample, or 172 men. The remaining 33% of the sample (137 men) were selected from the Boston Tower at Logan Airport (34 men), from the New York Common IFR Room (66 men), which provides radar control for the entire New York City area, and from a number of smaller facilities with less dense air traffic, all located in New England. A total of 37 men were drawn from the Hartford-Springfield Airport (Bradley Field), Otis and Quonset radar facilities and Providence Tower (Green Airport). The total selection represented facilities with higher and lower traffic density, including centers, airports and small radar en route facilities (TRACONs). A summary of the sampling results by facility, age, and experience groupings is given in Table 2.

Throughout the study there was a continuing gradual attrition of participants. The attrition after intake into the study is summarized in Table 3. Most of the men who discontinued their participation were forced to do so because of promotion to supervisory or data systems specialist jobs which were not included in the study, or because of transfer to other facilities not represented in the study. Of the 133 men who had withdrawn between Rounds 1 and 5 of the physical examinations, only 24 did so for personal reasons, while the others were required to withdraw by the eligibility criteria. However most of the men who withdrew at various times after the intake examination agreed to return for the final examination three years after intake. The report therefore includes data on their major health changes but not on more minor transient health changes.

In order to assure that the continuing attrition did not bias unduly the sample remaining an analysis was made of the age and years of service of those who dropped out. Table 4 displays this data. Although there are apparent differences in the age and experience between the participants and those who had been disqualified from the study by any given examination, the differences were not significant statistically ($p > .10$). There were significant differences in the age and experience of men from different facilities ($p < .02$), with younger and less experienced men at smaller facilities. This was due to the composition and stratified sampling procedure. However, this difference between facilities was maintained in spite of the disqualified volunteers at each examination. Consequently these data clearly indicate that no bias was introduced in our original age and experience sampling by subsequent disqualifications from the study.

Later in this report we compare the disqualified men on a number of characteristics. These comparisons are made to investigate differences between those promoted and those not promoted; those transferred and those not; those medically disqualified and those

TABLE 2

Summary Of Volunteer Rate As Of September 1974

FACILITY	STATUS	I	II	III	IV	TOTAL
	GROUP:	I	II	III	IV	
	AGE:	25-37	25-37	38-49	38-49	
	YEARS EXPERIENCE:	1-7	10-15	10-15	16-20	TOTAL
ISLIP	# eligible	64	59	39	48	210
	# volunteered	55	57	34	42	188
	% of eligible who volunteered	86%	97%	82%	90%	90%
NASHUA	# eligible	33	29	57	32	151
	# volunteered	28	23	35	23	109
	% of eligible who volunteered	85%	79%	61%	72%	72%
LOGAN	# eligible	17	15	13	8	53
	# volunteered	14	9	8	4	35
	% of eligible who volunteered	82%	60%	62%	50%	66%
OTHER	# eligible	23	6	16	10	55
	# volunteered	17	5	9	5	36
	% of eligible who volunteered	74%	83%	56%	50%	65%
IFR ROOM	# eligible	25	23	18	12	78
	# volunteered	22	21	14	10	67
	% of eligible who volunteered	88%	91%	77%	83%	86%
TOTALS	# eligible	162	132	143	110	547
	# volunteered	136	115	100	84	435
	% of eligible who volunteered	84%	87%	70%	76%	80%

TABLE 3

ATTRITION after INTAKE
for PARTICIPANTS in the
ATCS Health Change Study

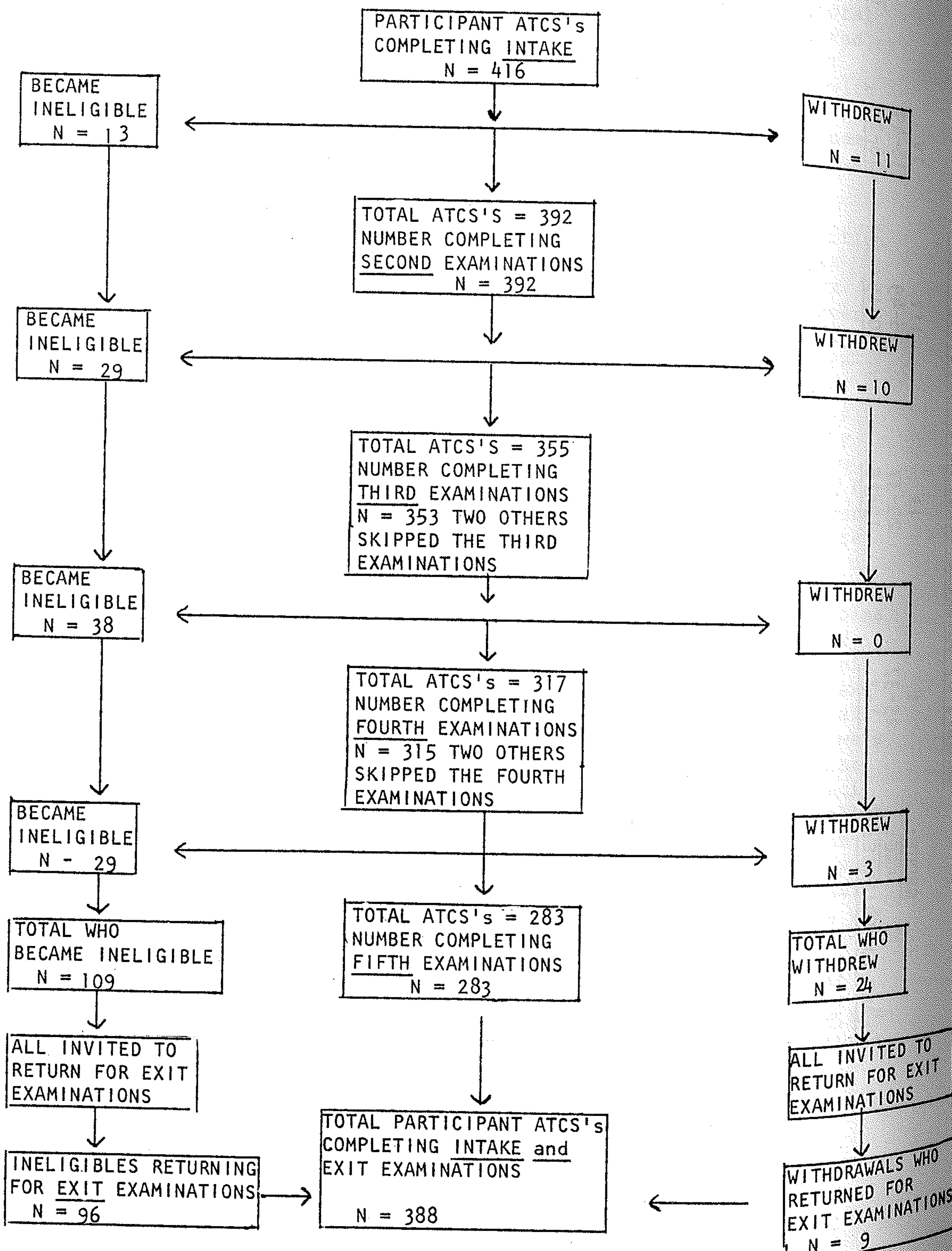


TABLE 4

Comparison Of Age And Experience Between Participants And Disqualified Volunteers Cumulatively Across Examinations

FACILITY

	Examination Number	Subject Status	Boston ARTCC	New York ARTCC	Logan	Others	Common IFR Room	Total Number	Overall Mean
	2	Participants	37.5	35.9	34.9	36.1	35.1	392	36.1
	2	Disqualified	37.2	39.6	36.0	32.3	37.4	24	37.3
	3	Participants	37.4	35.7	34.6	36.2	34.6	353	35.9
	3	Disqualified	37.7	38.0	36.7	32.0	38.3	63	37.5
YEARS OF AGE	4	Participants	37.3	35.6	35.7	36.5	34.9	315	36.0
	4	Disqualified	37.9	37.8	33.9	33.3	36.3	101	36.6
	5	Participants	37.3	35.4	35.5	35.9	35.0	283	35.9
	5	Disqualified	37.9	37.8	34.4	35.4	35.9	133	36.8
	2	Participants	12.0	11.3	10.0	10.0	10.3	392	11.1
	2	Disqualified	11.0	13.8	12.5	7.3	14.2	24	12.2
	3	Participants	11.9	11.2	9.6	10.1	9.9	353	10.9
	3	Disqualified	12.3	13.1	12.5	6.8	13.5	63	12.5
YEARS OF EXPERIENCE	4	Participants	11.7	11.0	10.7	10.4	10.2	315	11.0
	4	Disqualified	12.7	13.2	9.3	7.5	11.5	101	11.7
	5	Participants	11.6	10.7	10.5	10.0	10.3	283	10.8
	5	Disqualified	12.6	13.4	9.7	9.3	11.1	133	11.9

not; and those who withdrew for personal reasons. As these comparisons are made on the criterion variables, physical health change, psychological health change, and job outcomes, they are discussed in the sections on study findings.

Domains of Data

The predictor variables selected for their suggested relevance to health changes were classified into the following ten domains for data collection and analysis:

1. Socio-demographic characteristics, including social background and early life, as well as current life variables,

2. Health history and health-related behavior, including family history of disease, personal health, illness at intake, health-related behaviors;

3. Personality characteristics measured by such tests and surveys as the California Psychological Inventory, the Jenkins Activity Survey, the Cattell 16 Personality Factors Questionnaire, and scales for measurement of such characteristics as anomie and "connectedness";

4. Marital, family and social supports, each assessed by several different coping scales;

5. Job-related characteristics, including specific ATC variables such as attitudes about skills and investment in the job, affective discomfort, including work anxiety, subjective costs, burnout-bounceback, shift change costs, training anxiety, incident anxiety and adaptive styles including coping by drinking, work avoidance, tension discharge rate, coping by physical activity, and attitudes toward the FAA including satisfactions with management, training and promotions. Other job-related variables included are general satisfactions such as self morale or group morale, and specific satisfactions including work, co-worker and pay satisfactions and supervisory descriptions including initiation of structure, consideration of others and tolerance of freedom;

6. Life changes in work, relocation, education, legal, financial, marital, family, children, personal and interpersonal areas of life,

7. Measurement of air traffic controller work load and special problems such as weather, traffic blockages and equipment failure;

8. Psychological and behavioral measures at work;

9. Cardiovascular measures at work;

10. Endocrine measures at work;

Table 5 lists the predictor variables and the instrument or technique used to measure them as well as the frequency of observation. Each variable is described in turn in Section III, Descriptive Findings, and the rationale for its selection is reported along with the descriptive findings. Those variables found to have predictive value are discussed in this respect in Section V, Predictive Findings.

Plan for Data Analysis

The prospective, repeated-measures design of this study suggested five basic models for analysis using both parametric and non-parametric techniques. These included (1) predictions from intake status, (2) predictions from early dynamic changes, (3) retrospective group comparisons, (4) combinations of these three, and (5) what we called "sliding interval analyses".

The predictions of health change from predictor variables measured at intake were based on parametric forms of statistical analysis such as multiple regression and canonical analysis where outcomes were continuous or multi-variate, or the predictions were tested by simple chi-square analyses where predictors and outcomes were both categorical. The strength of this model lay in its simplicity for clarifying relationships between independent and criterion variables in a time-dependent, though not necessarily causal, sequence.

The second model for analysis was based on the fact that we had repeated measures of our independent variables. The prototypical question to be raised with this model was: Does early change predict later change? The same types of statistical analyses that were applicable for the first model applied to this second model as well. However, since changes were analyzed with respect to other changes, adjustments for initial levels and the reliability of assessments were important. The importance of such adjustments stems from the heightened probability of correlating chance fluctuations with other chance fluctuations in dynamic predictions. We did not emphasize this model of analysis for these reasons.

The third model for analysis was based on criterion group comparisons of prospective intake data. In this type of analysis we attempted to answer such questions as: What was different three years ago between those who have become ill and those who have

TABLE 5

Predictor Variables

Generic Variable	Instrument or Technique	Frequency of Observation: Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Yearly: Twice	Yearly: Four times
1. Socio-demographic characteristics (e.g. social background and early life, current life situation)	Biographical Questionnaire	X						
2. Health history and health-related behavior (e.g. family history of disease, personal health history, health-related behavior)	Medical Questionnaire							
	Medical history (family and personal)	X						
	Update on medical history		X	X	X	X		
	Smoking eating, drinking habits, medications and review of systems in the interim	X	X	X	X	X		
	SMA - 12, Rd.1,	X						
	Audiological Exam			X				
3. Personality characteriscics, including anomie, "connectedness"	CPI, JAS		X					
	Boston Health Study Questionnaire			X				
	MMPI subscales			X				
	Interview ratings			X				
	ATC Questionnaire "anomie" scale	X						
4. Marital, family and social supports including marital resources, social coping, biographical items	ATC Questionnaire	X	X	X	X	X		

TABLE 5 cont'd.

Predictor Variables

Generic Variable	Instrument or Technique	Frequency of Observation						
		Visit					Yearly	
		1	2	3	4	5	Twice	Four times
5. Job-related characteristics, including:								
a. Attitudes, investment;	ATC Questionnaire	X	X	X	X	X		
b. Affective discomfort, including anxiety factor, subjective costs, burnout-bounceback, shift change costs, training anxiety, incident anxiety;	Job Description Inventory, including Leadership Behavior Scale, Kavanagh Life Attitude Profile, Satisfaction with FAA Policies scale	X	X	X	X	X		
c. Adaptive styles, including coping by drinking, work avoidance, tension discharge rate, coping by physical activity;	Sociometric Questionnaire	X	X	X	X	X		
d. Attitudes toward FAA including satisfactions with management, training and promotions;								
e. General satisfactions including self-morale or group morale;								
f. Specific satisfactions including work, co-worker and pay satisfactions. Supervisory descriptions including initiation of structure, consideration of others tolerance of freedom;								
g. Work competence - peer competence, amicability, ideal team choice								
6. Impact of life changes	Review of Life Experiences (ROLE), life change inventory with standardized and individual adjustment and distress ratings	X	X	X	X	X		

TABLE 5 cont'd.

Predictor Variables

Generic Variable	Instrument or Technique	Frequency of Observation: Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Yearly: Twice	Yearly: Four times
7. Air traffic controller work load including work, pace, and time loads, and special problems of weather, traffic blockages, equipment failure. Exposure to ATC work	Measures of workload made by trained observers.							X
	Subjective Difficulty (of day's work) Questionnaire							X
	Questionnaire on assignment to tasks other than ATC				X			
8. Psychological and behavioral measures at work including relationship of workload to self-report of difficulty	Profile of Mood State Questionnaire (before start of work)							X
	Behavior Rating Scale							X
	Subjective Difficulty Questionnaire							X
	Measures of workload							X
9. Cardiovascular measures at work	BP and HR responses each 20 min. for 5 hrs. while working						X	
10. Endocrine measures at work	Cortisol and growth hormone levels each 20 min. for 5 hrs. while working						X	

remained disease-free? In this model, the groups of interest were defined at the end of the study and then their data at the beginning of the study were compared by means of analysis of variance, chi-square, or discriminant/classification analysis. This model of analysis was a primary one since it yielded more easily presentable results.

The fourth model for analysis was an amalgam of those presented thus far: that is, two prospective groups were formed, one composed of those subjects who were relatively healthy at intake and one composed of those who were ill in some way. The two groups were then studied using the predictive intake and retrospective criterion models. The results from the study of the healthy group indicated what variables may have an etiological role as risk factors in the development and onset of more acute illnesses. Since both groups were part of the same initial study population and since both would have been exposed to essentially the same occupational environment, variables that showed significant relationship to health change in both groups should be of highest priority for further research and action.

Probably our most important and robust model for analysis was what we called "sliding interval analysis." This model was needed for several reasons. Major illnesses and injuries are relatively infrequent in healthy persons ages 25 to 50 years. This fact, coupled with the size of the study sample in the ATC HCS meant that there were relatively few cases of severe illness or major injury occurring in any one interval (the approximate 9-month span between examinations at Boston University Medical Center) of the three-year surveillance period of the ATC HCS. It was necessary, therefore, to combine new health change events over several intervals to obtain a sufficient sample of cases for an adequate statistical search for predictors of the health changes. In addition, it was necessary to use measurements made as close in time as possible to the onset of illness or injury while still avoiding data which was collected after the subject was aware of the presence of a health change. These methodological necessities were resolved by use of a technique standard to epidemiologic studies of acute conditions. This technique has often been used to trace the specific cause of common-source infections. In this method, the time of the onset of symptoms is designated Time 0. Even though the actual onset of symptoms may differ widely for different individuals, for purposes of study the onset is designated as Time 0 for each. Antecedent behaviors and exposures for each case are gathered by interview or other means, and placed in a time frame, which for each individual counts backward from Time 0.

Using the same approach, we designated the interval in which a serious illness or injury occurred as Time 0 for each man, and then used as predictors all information gathered at the preceding physical and psychological examination.

The question then arose as to which data were to be selected as comparable predictors for the control group of ATCs who remained healthy. One possible approach was to use predictor data from the intake examination for all healthy persons, and compare it to the data for the illness cases. This introduced time biases inasmuch as the predictor data for the illness cases were derived from examinations 1 through 4. Another approach was to select all data for healthy persons from the second or third examination inasmuch as it was midway in the series. This also introduced a number of biases, such as the familiarization effect which occurs with repeated visits to the examination site, changes in interviewers and modification of the test forms over time, and secular changes at local facilities and in agency policies. In many facilities and for many controllers, conditions were different at Examination 2 from those at Examination 1 or 4.

A procedure that eliminated all of these problems -- and the methodological approach that we finally selected -- was to determine the proportion of cases for whom the pre-onset examination was the first, second, third or fourth round of examinations in the study. We then applied these same proportions to the control sample which remained free of disease, and selected the relative frequency of data from Rounds 1, 2, 3 and 4 for the healthy controls that would provide a match for the rounds contributing data for the men with health changes. Thus, if 10%, 20%, 40%, and 30% of cases had Rounds 1, 2, 3, 4, respectively, as their pre-onset examination, the healthy control group would be identified and a random 10%, 20%, 40%, 30% would have their "predictor data" drawn from examinations 1, 2, 3, 4, respectively. This procedure controlled for learning and familiarization effects of repeated visits to the examination center and repeated administration of the same and similar test forms. It also controlled for such changes in interviewers as occurred over the course of the study, the different contexts in which questions and scales were imbedded at different examination periods, and the changes that occurred over the three years of the study at local facilities and in agency policies. In addition, there were substantial changes in amount of traffic being handled over the course of the study, as exemplified by the year of the fuel shortage in which aircraft traffic was much diminished. This, in turn, might have affected either the answers to the predictor questions, or the probability of psychological or physical health change if, as hypothesized, work-load conditions were in some way related to these changes.

The technique of sliding interval analysis was used in all instances where the onset of disease could be clearly specified and the frequency of health change in the average interval was not sufficiently great to permit data analysis for single intervals.

Within all five general models for analysis we also studied moderating variables such as social and marital supports. In this way we were able, for example, to ascertain whether those who experienced equal amounts of physiological arousal at work but who had more social or marital support experienced less health change.

This report has been organized to facilitate rapid comprehension of the main points. Tables and figures required to convey understanding or clarity are included with the narrative in each section. Other tables lending supportive evidence, and background to the narrative, are included in Appendix I. Copies of instruments, questionnaires, forms, etc., used in the study are presented in the Exhibits appended. Copies of papers published under auspices of this study and contract are presented in Appendix II.

SUMMARY

Sociodemographic Characteristics of Sample

Responses on the ATC Biographical Questionnaire revealed the following sociodemographic characteristics of the sample:

1) The majority grew up in cities or suburbs;
2) More grew up in average neighborhoods (47%) than in "better" (35%) or "poorer" (10%) ones;
3) Most came from a similar socioeconomic background as their wives;
4) Most (66%) came from a 2-parent home,
5) Fathers and fathers-in-law were, for a majority, skilled, semi-skilled, manual or service workers;
6) At least half of the mothers and mothers-in-law were occupied mainly as homemakers;
7) Parental education was considerably lower than that of the ATCs,
8) About half of the ATCs and 27% of their wives had some college or a degree;
9) The majority (56%) were reared in the Roman Catholic religion,
10) Most (79%) were currently living in a "good" or "one of the best" neighborhoods;
11) Most (89%) were married and had children;
12) Most (almost 60%) of the wives were fully occupied as homemakers;
13) Almost all (99%) of the men had served in the Armed Forces, 96% of them as enlisted men, 69% worked in Air Traffic Control during military service.

III. DESCRIPTIVE FINDINGS

1. Sociodemographic Characteristics of Sample

The Air Traffic Controller Biographical Questionnaire (ATC BQ) administered at the intake examination yielded the following information on the general characteristics of our sample: Cities, or suburbs of these cities, provided the community setting for about 61% of the participating ATCs while they were growing up; another 27% grew up in small towns and 12% in rural areas. Forty-seven percent of the men said that they grew up in "about average" neighborhoods, about 35% in "better than average" neighborhoods, while only 10% reported that they came from "poorer" areas. However, about half said that their families were limited to necessities only while the other half reported that they were able to live comfortably. It appeared that growing up in "about average" neighborhoods meant being able to afford the necessities only. About 17% of the ATCs came from a higher socioeconomic background than their wives, and 19% from a lower one.

The family backgrounds of our sample were varied. Two-thirds of the study subjects lived with both natural parents until at least age 18. For 10%, a parent died and for about 12%, a divorce or separation (greater than a year) occurred. The average age at the time of such parental losses was 8.1 years. The fathers of 60% and the fathers-in-law of 62% of the men were either skilled or semi-skilled, manual or service workers. Eight percent of the fathers were professionals and 14% were managers and proprietors, and approximately the same total percentage of fathers-in-law were professional or business men. Among the mothers, slightly more than half were fully occupied as homemakers and 63% of the mothers-in-law were reported to be housewives. Forty percent of the mothers and 33% of the mothers-in-law were divided among clerical, manual and service occupations, usually as part-time employment shared with homemaking duties.

Parental education was considerably lower than that of the ATCs, reflecting the generally higher educational levels of the younger age cohort. About 50% of the ATCs had some college or a degree, while only 13% of their fathers, 9% of their mothers and 27% of their wives reached this level of education. Only 4% of the ATCs were not high school graduates, while 57% of their fathers and 44% of their mothers did not graduate from high school. Thirty-seven percent of the ATCs had more education than their wives, and 20% had less.

The parental family size differed greatly for our sample, ranging from 11 children to an only child (the ATC). The study subjects tended to be in the middle of their parents' family, having on the average 1.3 older siblings and 1.3 younger siblings. These were equally divided between the two sexes, but on the average these men were more likely to have an older sister and a younger brother than the reverse.

The majority of our sample (56%) were reared in the Roman Catholic religion. Thirty-three percent were reared in a Protestant religion, 6% in the Jewish religion, and 5% in either no religion or one not included in our listing. Somewhat the same current religious identification was reported, except that the organized religions were endorsed by a slightly lower percentage in each case and no religion at all was claimed by 14%. It should be noted, however, that 55% of these men reported going to religious services twice a year or less, so the religious identification was nominal in many instances. Only 20% of the group reported attending church services once a week or more.

Less than 1% of our sample reported that they currently were living in a section of town described as "one of the poorer ones," 20% were currently living in average neighborhoods, and 79% were living in either a "good" or "one of the best" neighborhoods. The ATC study group was upwardly mobile. Not only did they have higher occupational and educational levels than their parents (91% were at a higher educational and occupational level than their fathers), but they were also more likely to be living in a better section of town than they did when they were children, 59% of them reporting that they currently lived in better neighborhoods than they did as children, and only 10% reporting that they now lived in worse neighborhoods. The ATCs were also quite stable in their residence. Although we did not ask for the number of times they had changed their residence, we did collect information on how long they had lived in the county of their present address. The mean length of time was 10.2 years, and 76% had lived for five or more years in the county of their present address.

The majority of the sample were married and had children. Eighty-eight percent had been married once, 8% more than once, and 4% had not been married. At entry into the study, 89% were married, 4% were unattached, and 7% were separated, divorced or widowed. Eighty-seven percent had children living with them (an average of 2.3 children at time of entry into the study).

Among the wives living with their husbands, 30% had full or part-time jobs, but almost 60% were fully occupied as homemakers (another 12% were engaged in volunteer and other activities). The employed wives were largely in clerical and sales positions (84% of those employed); another 16% were pursuing a "Level 2 Profession" such as school teacher, technician (with college degree), social worker, registered nurse.

More than 99% of the men had served in the Armed Forces. Their tours of duty ranged from two to fifteen years with median and mode both at four years. Ninety-seven percent of the men were in enlisted ranks throughout their careers and only 3% were officers. For 69% of those who served in the Armed Forces, air traffic control was included in their military services; 2% had been pilots or navigators, and another 18% engaged in other activities related to aviation. Eleven percent were in military positions other than aviation-related ones. The history of a military career was the most uniform of all the life history information in our study; most were non-commissioned officers doing ATC work.

SUMMARY

Comparability of ATC HCS Subjects and ATCs in Other Studies

To assure credibility and applicability of results of this study, comparisons were made with other available studies: a nationwide testing, beginning in 1966, of all operational controllers in ARTCCs and airport towers using the 16 Personality Factor Questionnaire (16 PF), and a study that included 125 controllers from large and small airports in the vicinity of the University of Michigan, using scales measuring boredom, job future ambiguity, social supports and role conflicts.

The comparisons showed that:

1) The 16 PF profiles of the ATC HCS controllers matched the national sample profiles reasonably well, suggesting that the subjects in this study were reasonably representative of controllers nationwide in terms of this test;

2) ATC HCS controllers had more boredom than the Michigan area controllers, but were not greatly different in job ambiguity, social supports or role conflict, suggesting that Northeastern ATCs were reasonably similar to Midwest controllers in these respects.

2. Comparability of ATC HCS Subjects and ATCs in Other Studies

To assure credibility and applicability of the results of this study, it was important to determine whether or not the participating controllers, all of whom were from the New York and New England region, were similar to controllers around the country or to the controller subjects in previously published studies. There were two major sources of data for comparison.

Karson and O'Dell (1974) published the results of a nation-wide testing of all operational controllers in Air Route Traffic Control Centers (ARTCCs) and airport towers beginning in 1966. They used the Sixteen Personality Factor Questionnaire (16PF), 1962 Form A, of Cattell (1962) and obtained data from 6,231 ARTCC controllers and 4,816 tower controllers.

The second source of comparative data was the work of Caplan, Cobb, French, Van Harrison, and Pinneau (1975). They reported a large number of findings for 82 controllers from large airports and 43 controllers from small airports. These controllers were a convenience sample (as opposed to either a total or random sample) of controllers in the vicinity of the University of Michigan and therefore represent a sample for regional comparison. The 82 controllers from large airports were most comparable to our tower controllers in age, experience and Civil Service Grade (GS13), and therefore this particular sub-group provides the best regional comparison.

Although we could not expect that our controllers would be identical in all respects to either the national or regional comparison groups, a reasonable similarity would suggest that the findings of the present study might be applicable to controllers not in the region or study.

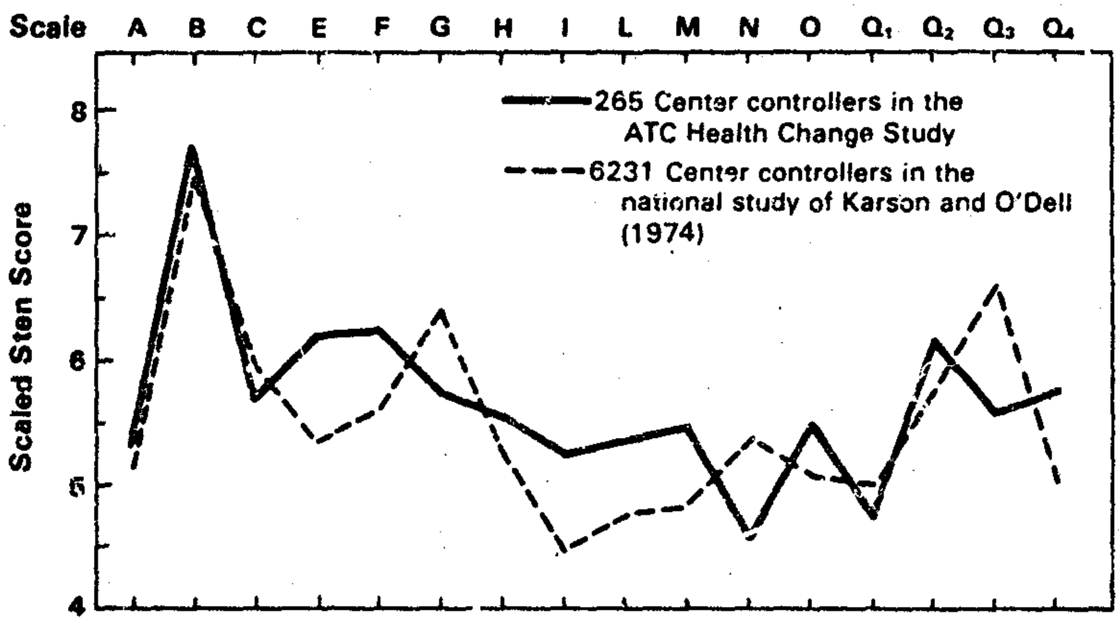
Comparison of 16PF Results To National Sample

Figure A displays the scaled score profiles of 265 ARTCC controllers in the present ATC Health Change Study (ATC HCS) and of 6231 ARTCC controllers in the national study of Karson and O'Dell (1974). Figure B displays the profiles of the 42 ATC HCS tower controllers and the 4816 tower controllers from the Karson and O'Dell (1974) study. In both cases the profiles appear quite similar in spite of several differences.

The degree of similarity in profiles can be expressed statistically by using the weighted occupational group-to-group coefficient of similarity for the 16PF (Cattell, Eber, and Tatsuoka, 1970, p. 311-312). The profile of the ARTCC controllers was significantly related to the national sample profile as expressed by a coefficient of similarity of .65 ($p<.01$). The tower controller profile also was significantly related to the national tower controller profile with a similarity coefficient of .76 ($p<.01$). These statistical results indicated that

FIGURE A

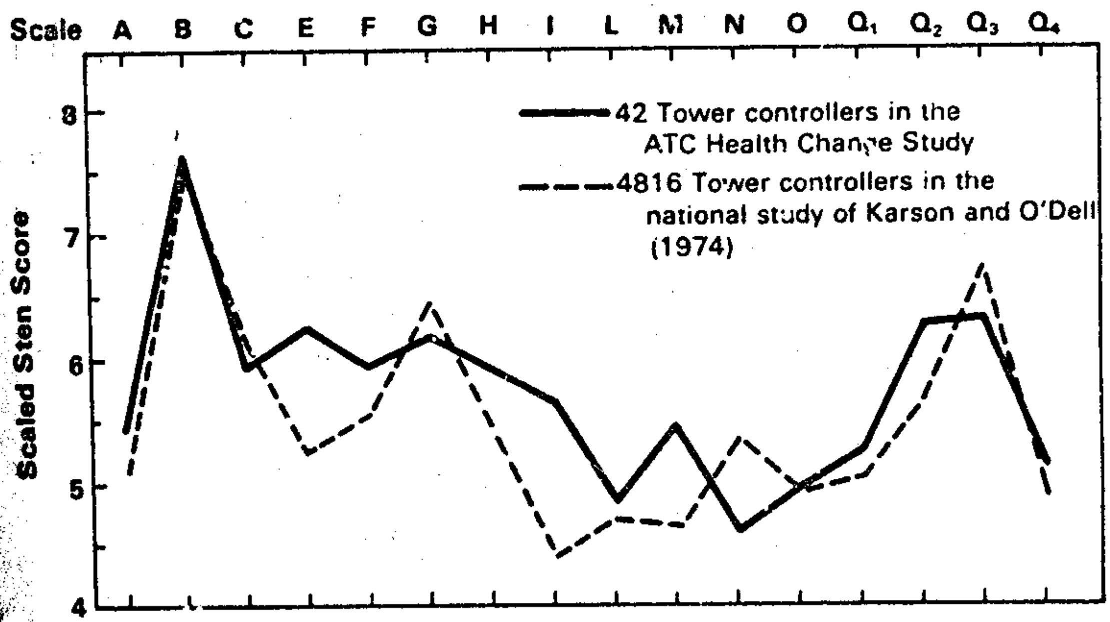
COMPARISON OF 16 PF PROFILES OF CENTER CONTROLLERS: ATC HEALTH CHANGE STUDY VS NATIONAL SAMPLE



A = WARMTH
B = INTELLIGENCE
C = EGO STRENGTH
E = DOMINANCE
F = ELATION
G = GROUP CONFORMITY
H = BOLDNESS
I = EMOTIONAL SENSIBILITY
L = ANXIOUS INSECURITY
M = BOHEMIANISM
N = SOPHISTICATION
O = GUILT PRONENESS
Q_1 = CRITICALNESS
Q_2 = SELF-SUFFICIENCY
Q_3 = COMPULSIVITY
Q_4 = FREE FLOATING ANXIETY

FIGURE B

COMPARISON OF 16 PF PROFILES OF TOWER CONTROLLERS: ATC HEALTH CHANGE STUDY VS NATIONAL SAMPLE



A = WARMTH
B = INTELLIGENCE
C = EGO STRENGTH
E = DOMINANCE
F = ELATION
G = GROUP CONFORMITY
H = BOLDNESS
I = EMOTIONAL SENSIBILITY
L = ANXIOUS INSECURITY
M = BOHEMIANISM
N = SOPHISTICATION
O = GUILT PRONENESS
Q_1 = CRITICALNESS
Q_2 = SELF-SUFFICIENCY
Q_3 = COMPULSIVITY
Q_4 = FREE FLOATING ANXIETY

the 16PF profiles of the ATC HCS controllers matched the national sample profiles quite well. Overall, the results meant that one would be right somewhat more than 99% of the time in saying that the ATC HCS controllers were reasonably representative of controllers nationwide, in terms of the 16PF profile of personality characteristics.

On individual scales of the 16PF, there were differences between the ATC HCS controllers and the national sample. Controllers in this study had more warmth (A), intellectual orientation (B), dominance (E), elation (F), boldness (H), emotional sensibility (I), anxious insecurity (L), bohemianism (M), guilt proneness (O), self-sufficiency (Q2) and free-floating anxiety (Q4). They also had less ego strength (C), group conformity (G), sophistication (N), and compulsivity (Q3).

Although there were statistically significant differences between the ATC HCS controllers and the national sample on individual scales, the high similarity of profile coefficients with respect to personality characteristics suggests that data and findings of this study may have application to the broader group of air traffic controllers.

Comparison to Caplan, et al. Regional Sample

During the fifth and final examination, a number of scales used in the Caplan et al. study (1975) were administered to the ATC HCS controllers: Boredom; Job Future Ambiguity; Supervisor Support, Social Support-Co-Workers, Social Support-Friends, Relatives, and Family; and Role Conflict Scales. One man refused to complete these scales, therefore, data was available on 387 of 388 men seen at the last examination.

After grouping our controllers into ARTCC (Center), tower, and TRACON (Traffic Radar Control) facility options, we compared their scores with the 82 controllers from large facilities and the 43 controllers from small facilities reported by the Caplan group. Table 6 displays the results.

The results indicated that the ATC HCS controllers had significantly higher ($p<.001$) Boredom scores than the Caplan study controllers. However, the ATC HCS controllers were not dramatically different on the remaining possible comparisons for Job Ambiguity, Social Supports or Role Conflict.

These results were important because they showed the basic similarity of the ATC HCS controllers from the Northeastern facilities in this study to those from the Midwest in the Caplan study.

TABLE 6

Comparison of Caplan, et al. (1975)
Regional Controllers with ATC Health Change Study Controllers

Comparison Variable	Caplan, et al. ATCs-Large Airports (N=82)	Caplan, et al. ATCs-Small Airports (N=43)	ATC Health Change Study Center Controllers (N=265)	ATC Health Change Study Tower Controllers (N=46)	ATC Health Change Study Tracon Controllers (N=76)	
Job Dissatisfaction	3.09	2.85	1.43	1.51	1.46	Mean
	.66	.69	.44	.54	.56	S.D.
Boredom	1.49	1.40	2.32	1.81	1.88	Mean
	.68	.65	.67	.85	.79	S.D.
Job Future Ambiguity	2.21	2.02	2.48	2.64	2.45	Mean
	.66	.72	.97	.88	.87	S.D.
Supervisor Support	2.69	2.87	2.77	2.59	2.65	Mean
	.72	.64	.85	.86	.91	S.D.
Social Support-Coworkers	3.12	3.37	2.84	2.92	2.95	Mean
	.54	.46	.57	.60	.68	S.D.
Social Support-Friends, Relatives and Family	3.43	3.58	3.30	3.16	3.19	Mean
	.60	.51	.70	.95	.76	S.D.
Role Conflict	1.56	1.36	1.48	1.59	1.61	Mean
	.56	.50	.51	.55	.47	S.D.

We made one other comparison to the other samples by grouping the various characteristics into those which might be socially desirable or undesirable for controllers in this study, from the motivational standpoint and tested to see if the ATC HCS subjects had a propensity to present themselves in any better or worse light than the other samples. We found that the ATC HCS subjects were reasonably representative of the other controller population.

SUMMARY

Comparability of ATC HCS Subjects to Non-Participating ATCs in The Same Regions

To assure that the sampling procedures and definition used in this study did not select a group of subjects very different in illness proneness from those who either were not selected or did not volunteer, a comparison was made of medical disqualifications among participants and those among non-participants.

It was found that participants had been medically disqualified at a rate approximately half that of non-participants. Differences between participants and non-participants in disqualification for psychiatric reasons were not as great as differences between the two groups in disqualifications for somatic reasons.

Not only was the rate of medical disqualification (including both participants and non-participants) for the New England facilities higher than that for the New York facilities, but there was also a marked difference between participants and non-participants in medical disqualification rate.

These findings indicated that the rates of health change reported in this study were probably conservative estimates of health change among the larger population of air traffic controllers.

Among 95 men (both participants and non-participants) medically disqualified from New England facilities and the New York Common IFR room, 78% were granted Second Career and/or OWEC options.

Participant ATCs were promoted at a rate approximately 1.7 times that of non-participant controllers.

3. Comparability Of ATC HCS Subjects To Non-Participating ATCs In The Same Regions

The preceding section reported that the air traffic controllers in the Health Change Study were reasonably comparable to controllers around the country in terms of their personality profiles. Furthermore, they were reasonably comparable to samples of midwestern controllers in terms of several psychological and job-related variables. The participating controllers in the Health Change Study were also compared to non-participating controllers from the same facilities, in terms of their medical disqualifications for somatic or psychiatric reasons. This comparison was important to assure that the sampling procedures and sample definition that we used did not result in a group of men who had medical disqualification rates or problems very different from men who were not participants. To obtain the necessary data we asked the Eastern Regional Flight Surgeon, John P. Skelly, M.D. and the New England Regional Flight Surgeon, U.A. Garred Sexton, M.D., to provide us with certain data.

We asked for the total number of fully rated journeyman controllers on every facility roster, the attrition frequencies for reasons other than medical, the number of medical disqualifications, the disqualified man's initials, date of birth, entrance-on-duty date, the date qualified as a controller, and the diagnosis if it were a medical disqualification. These data were requested for each month beginning January, 1974 through October, 1977. The men's initials, birthdates, and entrance-on-duty dates allowed us to identify those disqualified controllers who had been participants in the study. However, that same data did not allow us to obtain the actual identity of the disqualified non-participants, but allowed us to identify how many non-participants in each facility had the characteristics described. We were deeply appreciative of the efforts put forth by Drs. Sexton and Skelly and their staff to provide us with this data in a relatively short period of time.

Although these data also represented occupationally related outcomes which are discussed in Section III E, the purpose of the analysis in this section was to ascertain the extent to which the subjects in this study appeared to match their non-participating colleagues in illness proneness, as measured by number of medical disqualifications in each group.

Table 7 displays the comparison of somatic and psychiatric disqualifications of non-participants and participants in the health change study according to age groupings. The figures at the top of the table are simple counts of the number of men disqualified for the listed reasons in the various categories. To make comparisons between the disqualification rates for participants and non-participants, we calculated the number of man-months at risk.

TABLE 7

Overall Age-Grouped Comparison of Medical Disqualification Rates For ATCs Participants And Non-Participants

MEDICAL DISQUALIFICATIONS AT ALL FACILITIES IN STUDY

	NON-PARTICIPANTS		PARTICIPANTS		
Age as of 9/73	somatic	psychiatric	somatic	psychiatric	Total
25-37	20	15	5	5	45
38-49	20	12	7	6	45
50+	30	11	0	0	41
subtotals	70	38	12	11	131
Total Non-participant/participant	108		23		

I. Overall man-months at risk = 43674

A. Participant man-months at risk = 12780
B. Non-participant man-months at risk= 30894

II. Overall medical disqualification rate - 3.00 per 1000 man-months

A. Participant medical disqualification rate = 23/12780 = 1.80 per 1000 man-months.
B. Non-participant medical disqualification rate = 108/30894 = 3.50 per 1000 man-months.

Total somatic disqualification rates

A. Participants = 12/12780 = .94 per 1000 man-months.
B. Non-participants = 70/30894 = 2.27 per 1000 man-months.

Total psychiatric disqualification rates

A. Participants = 11/12780 = .86 per 1000 man-months.
B. Non-participants = 38/30894 = 1.23 per 1000 man-months.

For participants, the number of man-months at risk was calculated from the date of entry into the study until the date of exit from the study. The date of exit from the study was the date on which a man became ineligible for participation for any one of several reasons, or the date of his fifth and final examination at Boston University.

Since we did not have the same depth of data regarding the men who were non-participants in the study, the man-months at risk for non-participants were calculated by first summing the number of fully rated journeyman controllers in each facility during each month from January, 1974 through October, 1977, and then subtracting from this number the man-months at risk for the participants in this study. In all of the facilities in the study, the monthly roster of individuals changed due to medical disqualifications and other reasons, such as transfers and retirements. The overall man-months at risk for fully rated journeymen controllers in all of the facilities in the study was equal to the sum of all men on the facility roster each month from January, 1974 to October, 1977, made up of participants and non-participants.

With these calculations it was possible to compare the disqualification rates across all facilities as well as for regions. Table 7 shows that the overall medical disqualification rate for any reason was 3.00 per thousand man-months. The participant medical disqualification rate was 1.80 per thousand man-months, whereas the non-participants were medically disqualified at a rate of 3.50 per thousand man-months, almost twice the participant rate.

Part III of Table 7 shows that the participants had somatic disqualification rates of .94 per thousand man-months, while the non-participants were disqualified for somatic reasons at more than double that rate, 2.27 per thousand man-months. Disqualifications for psychiatric reasons at a rate of .86 per thousand man-months among participants, and 1.23 per thousand man-months among non-participants, were not so greatly different.

We decided to examine these data in more detail by region. Table 8 displays a distribution of all of the medical disqualifications by primary somatic and psychiatric diagnoses, in the Eastern (New York) region facilities in this study (New York Common IFR Room, New York Air Route Traffic Control Center). The same diagnoses are listed so that the reader may compare them with the diagnoses made in the study reported in Section III C 3. The diagnoses for psychiatric disqualifications are listed for comparison with those reported in Section III D 3. For the latter comparison it is important to note that the primary psychiatric diagnoses were neurotic and not psychotic disorders.

Table 8 also shows the comparative rates of medical disqualifications in the New York region for somatic diagnoses and for psychiatric

TABLE 8

Distribution of Medical Disqualifications
From January 1974 Through October 1977
New York Region Facilities
Journeyman ATCs Only

1. Number of ATCs who were medically disqualified = 48

A. Primary somatic diagnoses

	Non-Participants[1]	Participants
Gross obesity and hypertension	1	0
Bronchitis and hypertension	1	0
Hypertension	3	0
Myocardial infarction	1	1
Cardiovascular disease	3	0
Epilepsy	1	0
Paraesthesia	0	1
Spinal disc injury	1	0
Back injury	0	1
Hearing deficit	1	1
Nephritis	1	0
Ulcer and hiatus hernia	1	0
Diverticulitis and colitis	1	0
Ulcer and colitis	2	0
Post-concussion syndrome	1	0
Multiple phorias	2	0
Deceased (1-cancer, 1-accidental)	1	1
Subtotals by participation status	21	5

Subtotal of primary somatic diagnoses = 26

[1] Includes those subjects who were ineligible, randomly excluded and randomly selected but non-volunteers.

B. Primary psychiatric diagnoses

	Non-Participants	Participants
Anxiety and/or depression[2]	12	6
General psychiatric	2	0
Psychophysiological disorder of G.I. Tract	1	0
Alcoholism	0	1
Subtotals by participation status	15	7

Subtotal of primary psychiatric diagnoses = 22

[2] ...des two cases with secondary diagnosis of hypertension

TABLE 8 conc'd.

II. Comparative rates of medical disqualifications

A. Total man-months at risk = 20730

1. Non-participants man-months at risk = 13279
2. Participant man-months at risk = 7451

B. Overall medical disqualification rate = 2.32 per 1000 man-months

1. Non-participants = 36/13279 = 2.71 per 1000 man-months
2. Participants = 12/ 7451 = 1.61 per 1000 man-months

C. Total somatic disqualification rates

1. Non-participants = 21/13279 = 1.58 per 1000 man-months
2. Participants = 5/ 7451 = .67 per 1000 man-months

D. Total psychiatric disqualification rates

1. Non-participants = 15/13279 = 1.13 per 1000 man-months
2. Participants = 7 / 7451 = .94 per 1000 man-months

III. Frequencies of Medical Disqualifications by type of Disqualifications and Age Group: New York Facilities

	Non-Participants		Participants	
Age as of 9/73	Somatic	Psychiatric	Somatic	Psychiatric
25-37	11	7	4	2
38-49	6	5	1	5
50+	4	3	0	0
	21	15	5	7

diagnoses. The comparative rates indicated that the New York facilities' participants in the study experienced somatic health change at a rate equal to only 42% that of non-participants, but were disqualified for psychiatric reasons at a rate equal to 83% of that for non-participants.

Similar comparisons were made between participants and non-participants for the New England regional facilities in the study. Table 9 displays the results of these comparisons as well as the primary diagnoses listed in FAA records. Again, psychiatric diagnoses were all neurotic in level and type rather than psychotic.

The overall rate of medical disqualifications in the New England facilities was 3.62 per thousand man-months, a figure substantially higher than the New York rate of 2.32 per thousand man-months. The total somatic disqualification rate of non-participants was 2.78 per thousand man-months compared to 1.31 per thousand man-months for participants. Participants were disqualified at a rate of .75 per thousand man-months for psychiatric reasons whereas non-participants were disqualified at a rate of 1.31 per thousand man-months.

While the participants from the New York region were medically disqualified at a rate that was 59% of their non-participating colleagues' rate, the New England participants were medically disqualified only half as much as their non-participating colleagues. Therefore, not only was the rate of disqualification greater in the New England facilities, but the difference between participants and non-participants was greater.

The comparisons between participants and non-participants were particularly important because they indicated that the rates of health change reported in this study were probably conservative estimates of health change among the larger population of air traffic controllers.

Data regarding the disposition of medical disqualification were available for a limited number of facilities. Table 10 displays the distribution of Second Career Program and Office of Workmen's Employment Compensation (OWEC) actions granted to 95 men who were disqualified medically from New England facilities and the New York Common IFR room. A total of 78% were granted Second Career and/or OWEC options after their medical disqualification from the FAA. Five percent were pending at the time we received this data in late 1977, and 17% had a disposition with no compensation other than their possible retirement benefits. These men were both participants and non-participants, but it was not possible to separate the groups and compare their dispositions.

Finally, Table 11 displays a comparison of promotion and transfer rates between non-participants and participants in the Health Change Study from January, 1974, through October, 1977, in the par-

TABLE 9

Distribution of Medical Disqualifications
From January 1974 Through October 1977
New England Region Facilities
Journeyman ATCs Only

1. Number of ATCs who were medically disqualified = 83

A. Primary somatic diganoses

	Non-Participants	Participants
Hypertension	5	1
Hypertension and psychoneuroses	2	0
Hypertension and diabetes	1	1
Cardiovascular disease	9	0
Cardiovascular disease and other	3	1
Diabetes	3	0
Hearing loss	11	2
Hearing loss and blindness	1	0
Ulcer	3	0
Hyperplasia	0	1
Multiple Sclerosis	1	0
Vertigo and medications	1	0
Musculo-skeletal disorder	1	0
Eye disease	2	0
General medical	5	1
Dermatitis and medications	1	0
Subtotals	49	7

B. Primary psychiatric diagnoses

	Non-Participants	Participants
Psychoneurosis plus other	1	0
Psychoneurosis	2	1
Personality disorder	4	0
Disturbance of consciousness	1	0
General psychiatric	13	3
Anxiety and depression	1	0
General psychiatric plus other	1	0
Subtotals	23	4

2. Comparative rates of medical disqualifications

A. Total man-months at risk = 22944

1. Non-participant man-months at risk = 17615
2. Participant man-months at risk = 5329

B. Overall medical disqualification rate = 3.62 per 1000 man-months

1. Non-participants = 72/17615=4.09 per 1000 man-months
2. Participants = 11/5329=2.06 per 1000 man-months

TABLE 9 cont'd.

C. Total somatic disqualification rates

1. Non-participants = 49/17615=2.78 per 1000 man-months
2. Participants = 7/ 5329=1.31 per 1000 man-months

D. Total psychiatric disqualification rates

1. Non-participants =23/17615=1.31 per 1000 man-months
2. Participants = 4/5329=.75 per 1000 man-months

III. Frequencies of Medical Disqualification by type of Disqualifications and Age Group: New England Facilities

	Non-Participants		Participants	
Age as of 9/73	Medical	Psychiatric	Medical	Psychiatric
25-37	9	8	1	3
38-49	14	7	6	1
50+	26	8	0	0
	49	23	7	4

TABLE 10

Disposition Of Medical Disqualifications From Selected Facilities For 95 Men[1]

	Number of Men	% of Total
Second Career Program only	45	47%
OWEC only	3	3%
Both Second Career and OWEC	15	16%
Second Career - pending	3	3%
OWEC - pending	2	2%
Second Career granted with OWEC pending	11	12%
Neither Second Career or OWEC	16	17%
	95	100%

Total granted Second Career and/or OWEC	=74	(78%)
Total with Second Career and/or OWEC pending	=5	(5%)
Total with no compensatory disposition	=16	(17%)

[1] Complete data for this presentation was available only for Boston ARTCC, Quonset TRACON, Bradley Tower, Providence Tower, Otis TRACON, and the New York Common IFR Room.

TABLE 11

Comparison Of Promotion And Transfer Rates
Between Non-Participant And Participant ATCs From
January 1974 Through October 1977 In Participating Facilities

	Rate per 1000 man-months	
	Non-participants ATCs	Participant ATCs
Promotions[1]	2.33	3.94
Transfers[2]	4.63	1.94

[1] Promotion data was available from the Boston ARTCC, Logan Tower, Otis TRACON, Quonset TRACON, and Bradley Tower

[2] Transfer data was available from New York ARTCC, Boston ARTCC, Logan Tower, Quonset TRACON, and Bradley Tower.

ticipating facilities. Full data was not available from all facilities, but the comparisons should still be valid since the largest part of the data were available for examination.

First, one can note that participant ATCs were promoted at a rate approximately 1.7 times that of non-participant controllers. Secondly, non-participant controllers were transferred to other facilities more than twice as often as participant controllers. Since a number of contributory factors possibly accounted for these differences, we were hesitant to interpret this data. A conservative interpretation might be that men who volunteered to participate in the study were not expecting to be transferred and furthermore may have had higher aspirations for themselves.

In summary, the most important conclusion from these comparisons of participating vs. non-participating controllers was that the controllers participating in the health change study were somewhat less likely to be medically disqualified than non-participants. Thus, the health change rates reported in the later sections may be somewhat conservative in that all journeymen controllers probably experience health changes at a rate slightly higher than those who were selected and who volunteered for this study.

B. Predictor Variables

Several criteria were applied to determine the selection of health change predictor variables. First, relevance to air traffic controllers and to the air traffic environment was required. Second, the variable had to be measurable. If a measurement technique was not available, we considered the feasibility of devising an assessment technique. On the other hand, if techniques were available, we considered them from the standpoint of practicality in usage. Third, we evaluated the reliability and validity of assessment techniques. And finally, we searched for and required evidence that a given variable would have either a direct or moderating influence on health change.

Using these criteria, ten predictor variables were chosen: Socio-demographic characteristics, history of health and health-related behavior, personality characteristics, marital, family and social supports, job-related characteristics, life changes, workload, psychological and behavioral measures at work, cardiovascular measures at work and endocrine measures at work.

A description of each predictor variable follows, along with the rationale for its use in the study, explanations of the method of measurement, and descriptive findings.

The ability of these variables to predict any of the health or job outcomes is discussed in Section V, Predictive Findings.

SUMMARY

SOCIO-DEMOGRAPHIC CHARACTERISTICS

Data were collected on the ATCs' parents; social, educational, economic, military and religious backgrounds; and adolescent experiences for analyses of their relationship to health change. Some descriptive data is reported in Section IIIA1. In addition, the data showed that:

1) The ATCs' parents were generally compatible;
2) 62% of the ATCs recalled a growth rate about the same as their adolescent peers;
3) 91% of the ATCs were equally or more athletic and fit than their high school classmates;
4) 91% had as many, or more, friends in high school as others;
5) 35% had considered quitting high school;
6) 77% socialized with girls at least weekly in high school;
7) The relationships of 86% of the ATCs with their parents were average or better than those of most others; and
8) 62% had alcoholic beverages outside their home by age 16.

1. Socio-demographic Characteristics

As we were interested in determining the extent to which certain background or current biographical data might predict future health or job outcomes, we used the information obtained from the Air Traffic Controller Biographical Questionnaire (ATC BQ) not only to describe the study participants (see Section III A 1) but also to ascertain whether or not there were any significant relationships between specific socio-demographic characteristics and health or job outcomes. Although general information descriptive of the participants has already been reported in the previous section, the rationale for attempting to relate socio-demographic characteristics with health changes, and additional information pertaining to aspects of the personal lives of the participants, is reported below.

Rationale

It is well accepted that the circumstances and life style into which a person is born and raised has an effect, often indeterminate, on his or her future well-being. Some attempts have been made to predict future outcomes by use of biographical inventories that elicit information on personal background. Early uses were in the prediction of success of salesmen (Goldsmith, 1922; Kornhauser, 1965; Kurtz, 1939), prediction of college success (Bittner, 1945) and vocational success (Ghiselli, 1955; Super, 1957), success in an officer training program (Guilford and Lacey, 1947), and detection and treatment of accident-prone drivers (Johnson, 1946).

Biographical data has been extensively used in industrial studies (Owens and Champagne, 1965), in psychological research to predict amenability to psychotherapy and severity of anxiety reaction (Lawrence, 1947) and in health research to establish a relationship between status incongruity and level of illness symptoms (Jackson, 1962), as well as between coronary disease and both status incongruity and social mobility (Cohen, 1974; Wardwell and Bahnson, 1973), and between health change and a variety of personal life history and current life variables (Jenkins, 1976). Finally, such background factors as religious affiliation (Kaplan, 1975; Wardwell and Bahnson, 1973) and church attendance (Kaplan, 1975; Comstock and Partridge, 1972) have also been related to a variety of health problems.

The study of health change in air traffic controllers presented a good opportunity to determine under carefully controlled conditions whether or not social background or current life circumstances were related to future health changes among this population.

Findings

Characteristics already summarized in Section III B 1 to describe the sample included their social setting while growing up, the economic situation and family size and membership of their parental families, education and occupations of fathers and mothers as compared with the ATCs themselves, and their religious affiliations. Other social background characteristics of interest for predictive purposes included the extent of compatability between parents in background, personality and goals for their families, and the adolescent experiences of the ATCs themselves, summarized below.

The ATCs perceived their parents as having very similar goals for the ATC, but perceived the parents' personalities as being dissimilar. More than 50% of the participants judged their parents to be very similar in social background, education and religious emphasis. To obtain an overall measure of parental compatibility on these issues, the scores of the five items inquiring about parental compatibility were summed, yielding a range from 5 (very similar on all items) to 15 (sharply different on all items). Using the total scores 5-7 as "very similar," 8-12 as "a little different," and 13-15 as "sharply different," the following results emerged: (1) very similar, 48%, (2) a little different, 47%, and (3) sharply different, 5%. The results indicated considerable compatibility between the parents of the study population.

Because of the importance of the teen-age years to later development, we included seven items in the ATC BQ to assess the ATCs' experiences at that time, as they were recalled.

When asked to compare their physical growth with that of their peers in adolescence, 62% of the study group believed they were about the same, 15% perceived their own growth rate as faster, and 23% as slower.

Asked how athletic and fit they were in comparison with their adolescent classmates, 44% recalled that they were about average, 47% above average, and 9% less fit.

For 75% of the controllers, the number of friends they had in high school seemed to be about the same as others, while 16% recalled that they had more friends than others, and 9% recalled having fewer friends than others.

Twenty-six percent of the study group reported that they had seriously considered quitting high school once or twice, 9% had considered it more frequently than once or twice, and 58% had never considered quitting.

Seventy-seven percent of the ATCs reported that during their last year in high school, they went out socially with girls one or more times a week, with 11% reporting that they went out four or more times a week. Twenty-three percent recalled going out with girls less than once a week.

The relationship between the ATCs and their parents during the ATCs' teen years was described as about average by 56% of the participants, as better than that of most others by 30%, and as worse than that of others by 14%.

The majority of the study group (62%) had their first alcoholic beverage outside of their home by the time they were 16 years of age. Another 35% had their first drink outside of their home by the age of 20; 1% never drank, and 2% were over 21 years of age before they drank outside of their home.

The extent to which the social characteristics described both in Section III A 1 and above are predictive of future health change is reported in Section V, Predictive Findings.

SUMMARY

Health History and Health-Related Behavior

Family health histories, elicited during visits to Boston University Medical Center, revealed the following:

1) 73% of the ATCs had one or more relatives who had had a heart attack or who suffered from angina;
2) 25% of the ATCs had fathers who had suffered either a heart attack or angina;
3) 50% of the controllers had some family members with hypertension, 20% of the controllers having fathers with hypertension and 20% having mothers with hypertension;
4) 40% of the men reported cancer among their relatives;
5) 43% of the men reported allergy in their families.

2. Health History and Health-Related Behavior

Rationale

Any attempt to explain health change in a population is incomplete without knowledge of the health histories of its members. Large studies of the health of industrial populations have shown that a significant, though weak, predictor of future health change is the frequency of illness in the past. In addition, a complete medical examination has traditionally included data both on family history of disease and personal disease history.

The impact upon health of a variety of daily life habits has been well established. Adequacy of sleep and regularity of meals have been associated with five-year survival rates in free-living communities as studied by Belloc and Breslow*. The use of cigarettes and other tobacco products has been associated with ill health and increased mortality in a variety of studies. One's choice of beverages has been related to future incidence of gastric ulcer. Hence, it was deemed important to obtain a family history of disease and a history of personal health and health-related behaviors at each of the five examinations of the ATCs at Boston University Medical Center.

There were two specific reasons, in addition to medical tradition, for including these factors. First, it was important to determine whether the rate of health change in air traffic controllers as a group followed from a particularly high or particularly low frequency of such health problems in the family background or earlier life history of the ATCs. The health histories and habits were necessary to test the hypothesis that a high rate of a particular disease might be observed in this sample of ATCs because some inadvertent selection factors had brought a high risk group into the study population. Secondly, to determine the independent contribution of current life stresses in illness causation, one would need to know the extent to which the illness was predictable on the basis of family or personal health history or habits alone.

Findings

At the intake examination, ATCs were asked to indicate which, if any, blood relatives had ever had any of the following conditions: heart attack or angina, stroke, high blood pressure, asthma/

* Belloc, N.B., RELATIONSHIP OF HEALTH PRACTICES AND MORTALITY. Prepublication report. Human Population Laboratory, California State Department of Public Health, Berkeley, California.

hay fever/hives or eczema, diabetes (sugar disease), nervous disorder, liver disease (jaundice), anemia or bleeding disease, kidney disease, cancer or leukemia, suicide, peptic ulcer, emphysema or other chronic lung disease. At each subsequent examination, ATCs were asked for any additional report of relatives who had developed any of these diseases and also for an update of the record if they had, in the interim, learned of the presence of any of the diseases in relatives with whom they may not have had regular contact. A medical history was elicited for the father, mother, grandparents, uncles or aunts (if blood relative), siblings and offspring of the ATC. The Medical History Questionnaire is appended in the Exhibits section of this report.

The most commonly reported familial health condition was heart attack or angina. In all, 73% of the men reported one or more of the relatives with such a history, including 25% who reported this for their fathers. High blood pressure was reported in the family by 50% of the controllers, with fathers and mothers each mentioned by about 20% of the men. The next most frequently reported health condition was cancer, mentioned by 40% of the men. In this case, grandparents and other blood relatives older than the subject were reported more than three times as frequently as either parent. The allergy category was mentioned by 43% of the men. Interestingly, the relatives most commonly affected were sons and daughters, reported by 17% of the men.

At each of the subsequent examinations, 2% to 3% of the ATCs reported an additional relative to have been affected by one or more of the more common of diseases from among a comprehensive list. Coronary disease, stroke, high blood pressure, and cancers were the most commonly reported subsequent family illnesses, as might be expected by their frequency in the general population.

The personal health history of the ATCs is reflected in the discussion on past and present illnesses at intake into the study (Section IIIC3), and health-related behavior is discussed in the context of its relationship to later health change (Section VA).

SUMMARY

Personality Characteristics

Although personality characteristics have frequently been related in research studies to illness experience, few studies have been able to separate personality characteristics resulting from, and those antecedent to, any illness outcome, nor have these studies shown the biological mechanisms transmitting the influence of personality.

Our prospective study of air traffic controllers provided an opportunity to attempt to clarify these and other issues raised regarding personality influence on health outcome.

Six different instruments or scales were selected to measure or assess these personality characteristics, each described in turn in the sections following.

3. Personality Characteristics

One of the long-standing controversies in medicine relates to the question of whether or not personality characteristics predispose either to specific diseases or to disease in general. An extensive literature (Weiner, 1977) has explored the relationships particularly to certain diseases considered to have more probable connection with personality and hence labeled "psychosomatic" in nature.

The crux of the controversy over the influence of personality on disease lies in two main uncertainties: First, most of the studies reported to date were unable to prove that the personality characteristics often found to be concomitant with certain diseases were in fact antecedent to the onset of the disease. As the studies were typically retrospective in nature, it was usually not possible to separate personality characteristics resulting from, and those antecedent to, any disease outcome. Second, studies in which personality characteristics were correlated with disease initiation or maintenance did not pursue how the personality attributes might have been linked to the biological mechanisms of the disease. As physiological and endocrine measurements were rarely made together with personality assessments, little evidence of the process of personality influence on illness could be offered.

Another provocative question was whether personality characteristics could in themselves predict certain diseases, whether a personality pattern (e.g., a pattern of dependency) was only activated under certain circumstances; or whether personality attributes played only a partial role in combination with other predisposing factors (e.g., life change, genetic predisposition, or other body factors) to initiate illness. Complicating these many questions has been the additional possibility that specific personality characteristics may relate to certain sub-types of disease and that better understanding of the influence of personality on disease initiation or maintenance may await greater clarity in specifying disease sub-types.

As our prospective study of air traffic controllers provided a good opportunity for possible clarification of some of these questions as well as an opportunity to answer, for this particular population, whether or not their illnesses could be predicted in any way by personality characteristics, we were eager to utilize measurements of personality characteristics that would allow comparison with other research findings, and would provide us with scores having generally accepted reliability and validity.

After considerable thought, we selected six different instruments or scales, each of which will be described in turn, along with the rationale for selection of that particular measure. The predictive value of personality characteristics as measured will be discussed in Section V, Predictive Findings.

SUMMARY

The California Psychological Inventory (CPI)

The CPI, a commonly used inventory of everyday personality traits arranged to measure 18 different dimensions, was administered to the ATCs at their second Boston University Medical Center examination.

Relative to 5000 U.S. males in the original standardization sample, the controllers as a group scored within a normal range, and showed group tendencies toward strong, dominant, agressive, independent, sociable, conscientious and self-confident personalities who dislike and feel uncomfortable with outside regulation and authority and who are somewhat intolerant of people different from themselves.

a. The California Psychological Inventory

Rationale

Our first requirement was for an instrument that would provide an acceptably valid and reliable general assessment of personality traits. We considered a number of multitrait personality inventories, including the Minnesota Multiphasic Personality Inventory (MMPI), the Sixteen Personality Factor Questionnaire (16PF) and the California Psychological Inventory (CPI).

As the MMPI concentrated primarily on psychopathology, which we had other means of assessing, and the 16PF was initially rejected by the air traffic controllers because of their bad experience with this measurement by other investigators at a previous time,* we decided against use of these measures.

On the other hand, the CPI seemed to be an ideal choice for several reasons, and therefore was selected to be administered during the second examination at Boston University Medical Center as an instrument for assessing general personality traits.

The CPI assesses eighteen personality characteristics (Gough, 1960). It had been used in at least one previous study of air traffic controllers (Trites, Kurek and Cobb, 1967), providing us with data for possible comparison and interpretative assistance.

The CPI focuses on more positive, everyday personality traits, such as dominance and responsibility. Its content consists of questions about typical behaviors, feelings, opinions and attitudes of a respondent about social, ethical and family matters. It is notable for its lack of psychopathologically or symptom-oriented material. Finally, the CPI has been used in literally hundreds of research studies, and therefore a large literature was available regarding its use and interpretation. These considerations were bolstered by high praise from a leading psychometrist (Anastasi, 1968) and by the detailed review of validity, reliability and psychometric research on the CPI by Megargee (1972).

Findings

Table 12 lists the names of the eighteen CPI scales and descriptive characteristics of persons who score high on each of

---

* Although the air traffic controllers initially rejected the 16PF, by the time of the fifth round of examinations at Boston University Medical Center, they had changed their view and demonstrated their confidence in us by allowing us to administer the 16PF to them.

TABLE 12

California Psychological Inventory

Scale Names and Descriptive Characteristics Scale

Scale Name	Descriptive Characteristics of High Scores
Dominance	Ambitious, dominant, forceful, aggressive, persistent
Capacity for Status	Discreet, forgiving, imaginative, independent
Sociability	Clever, confident, wide interests, outgoing, sociable
Social presence	Adventurous, pleasure-seeking, self-confident, relaxed
Self-acceptance	Confident, enterprising, independent thinking, sophisticated
Sense of Well-being	Dependable, good-natured, free from worries
Responsibility	Capable, conscientious, serious, useful, responsible
Socialization	Adaptable, efficient, honest, sincere, high social maturity
Self-control	Dependable, logical, self-controlled, freedom from impulsivity
Tolerance	Forgiving, generous, good-natured, soft-hearted
Good impression	Adaptable, changeable, considerate, kind
Communality	Cautious, conscientious, deliberate, non-randomness
Achievement-conformity	Ambitious, capable, intelligent, logical
Achievement-independence	Foresighted, independent, intelligent, versatile
Intellectual efficiency	Capable, confident, efficient
Psychological mindedness	Foresighted, independent, responsive to inner needs of others
Flexibility	Easy going, highly adaptable, optimistic, spontaneous

these scales. Figure C displays the CPI profile for the 391 controllers completing the questionnaire during their second examination at Boston University Medical Center (1 of the 392 who came to the second examination was not included because his scores indicated that he was randomly answering the questionnaire). The controllers' mean scale scores are plotted relative to 5,000 males in the original standardization study (Gough, 1960), where 50 was the average male score and 10 was the average standard deviation.

Relative to the standardization sample of middle class American males between the ages of 18 and 50, controllers scored quite high on the Social Presence, Dominance and Self-Acceptance scales, indicating dominant, independent, outgoing, adventurous and confident personalities in general.

The controllers scored quite low on the Responsibility and Socialization scale of the CPI. These two scales assess what one might call "respect for authority" in that very low scores are typical of juvenile delinquents, parolees, prisoners and other social misfits (Megargee, 1972). High scorers on these two scales tend to be in positions of leadership in their schools, communities and organizations. The low score for the controllers indicates not a lack of responsibility nor a lack of socialization as the scale names would suggest, but rather a low regard for rules and regulations and a tendency to do things as one wishes in order to get them done.

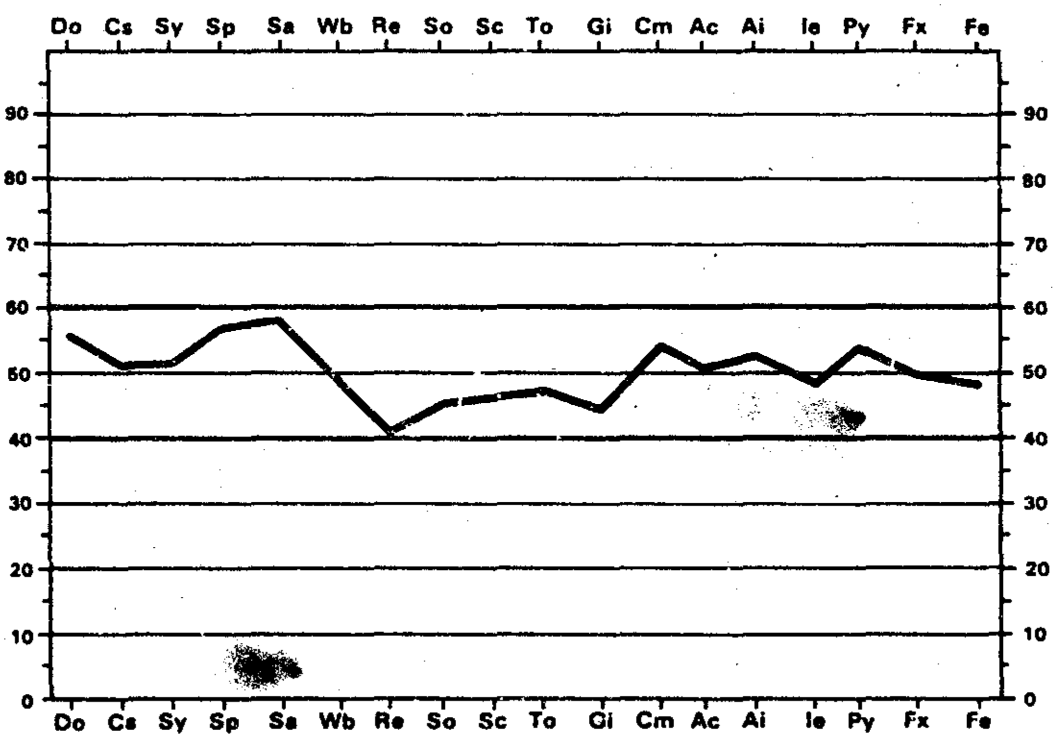
The controllers scored in the low average range on the Self-Control scale reinforcing the previous profile of dominant, forceful, aggressive and self-confident persons who dislike rules and regulations, and do not mind breaking laws or other external restrictions.

In all of the remaining scales the controllers scored in the average range, with slight elevations for being cautious and conscientious (Communality), for achieving in more independently oriented ways rather than conforming ways, and for being psychologically minded. Their Intellectual Efficiency scores were average as were their Flexibility and Femininity scores.

The CPI also provided three of the validity scales used in this study. The Sense of Well Being scale is an inverse "faking-bad" scale; that is, a person scoring extremely low (below a scaled score of 30) is likely to be describing himself in an excessively negative fashion. Studies have shown that people who are very upset will score in a more moderate range (30 to 40) than persons who are attempting to make themselves look very bad (Megargee, 1972). As a group the controllers had a slight tendency to describe themselves in less than optimum fashion, but an analysis of indivi-

FIGURE C

CALIFORNIA PSYCHOLOGICAL INVENTORY PROFILE OF 391 AIR TRAFFIC CONTROLLERS



Do = Dominance
Cs = Capacity for Status
Sy = Sociability
Sp = Social Presence
Sa = Self-Acceptance
Wb = Sense of Well-Being
Re = Responsibility
So = Socialization
Sc = Self-control
To = Tolerance
Gi = Good Impression
Cm = Communality
Ac = Achievement-Conforming
Ai = Achievement-Independent
Ie = Intellectual Efficiency
Py = Psychological Mindedness
Fx = Flexibility
Fe = Femininity

Note:
The standardization of over 5000 males scored at an average of 50 for all scales. The darker lines at 40 and 60 on the profile would include about 68% of the standardization sample scores.

dual records indicated that only a very few scored in the "faking-bad" direction. These individual records were examined and other data were used to confirm whether these men were in fact feeling as bad as their scores indicated. In all cases the additional data (medical, psychiatric and biological) confirmed the men's self description.

The second validity scale provided by the CPI was the Good Impression scale. The Good Impression scale assesses "faking-good," an extremely high score indicating an attempt to make oneself look better than is probably true. Again only a very few controllers scored very high on the Good Impression scale. For these few men, other data indicated a strong need to be seen in a good light, but on the other hand, they did not deny instances where they were not performing or feeling up to their ideal.

Finally the Communality scale indicates the degree to which a respondent's answers are consistent and systematic. Very low scores on this scale indicate random answering, concealing feelings and beliefs. One of 392 men tested scored below 30, which suggested random answering. Examination of his individual profile and the remainder of his data showed that the CPI results were confused and inconsistent but that his other data were not disordered. It may have been that this one man did not align his answer sheet with the proper questions and therefore systematically misanswered all of the remaining items. Since his score was so deviant and his answers were so inconsistent on the CPI, we discarded his results from all analyses using CPI scores.

In summary, the CPI indicated that controllers scored in essentially a normal fashion with group tendencies indicating strong, dominant, agressive, independent, sociable, conscientious and self-confident personalities who dislike and feel uncomfortable with externally imposed regulations and authority, and who are somewhat intolerant of people different from themselves. The validity scales indicated that very few men (a maximum of 30 over all three faking scales) had strong tendencies to describe themselves either very favorably or very unfavorably. In all but one case the results were considered valid because there was confirming data from other sources as to their different status.

The value of the 18 CPI personality variables in predicting health changes occurring after the second examination at Boston University are reported in Section V, Predictive Findings. Use of the "faking-good" and "faking-bad" scales to check the validity of other self-reported data is discussed in Section IV, Evidence for the Validity of Self-Reported Data.

SUMMARY

The Jenkins Activity Survey (JAS)

The JAS, measuring "Type A behavior pattern," was administered to the ATCs as a potential predictor of health change. Type A persons are hard-driving, competitive, impatient, agressive, hurried, job-involved, achievement oriented, conscientious and responsible. Type A behavior has been associated in other studies with increased risk of coronary heart disease and atherosclerosis.

Four JAS scores were provided:

1) Extent of Type A behavior pattern;
2) Speed and Impatience;
3) Job Involvement, and;
4) Hard-driving competitive.

b. The Jenkins Activity Survey

The Jenkins Activity Survey (JAS*) was designed to measure the Type A behavior pattern which has been associated in recent studies with increased risk of coronary heart disease and atherosclerosis. As this self-administered, computer-scored multiple choice questionnaire has been administered to many thousands of American and European men over 13 years of development and standardization, it was quite natural that the questionnaire was selected to be administered to the ATC population of the present study as an additional test of its value in predicting health changes. The ATCs were given this questionnaire during their second visit to Boston University Medical Center.

The JAS provides four scores:

The first score represents a clinical judgment as to the extent of Type A behavior pattern defined by the cardiologists Rosenman, Straus, Friedman et al. (1964). Scores in the positive direction indicate increasing intensity of the Type A pattern, whereas scores in the negative direction signify the converse Type B pattern. Type A persons are hard-driving, competitive, impatient, always in a hurry, aggressive, committed to their profession, achievement oriented, conscientious and responsible. In contrast, Type B individuals are more calm and relaxed, more interested in the quality of life than quantity; for the most part they are also interested in good performance and good success, but they seek their goals in a less pressured, less intense manner. Scores on the Type A scale have been found in several independent studies to be higher in patients with coronary disease than in controls, and high scores were also associated with higher risks of development of new clinical coronary disease and recurrent myocardial infarction. The Type A behavior pattern is a complex syndrome which is composed of three major independent dimensions. These dimensions are measured by the remaining three scores generated by the JAS.

Speed and Impatience. This scale, derived by factor analysis, measures the tendency to do things quickly, to be always in a hurry, to be impatient with others when they speak or work slowly, to have a strong temper and to be willing to express feelings forthrightly.

Job Involvement. This scale is composed of items depicting one's work situation as containing many pressures, important deadlines, demands for overtime work, many challenges and unexpected problems needing attention and correction. In addition, this scale

---

* Available from the Psychological Corporation.

includes items which indicate that high scorers will also receive promotions and pay increases in response to their job dedication.

Hard-driving, Competitive. High scorers on this scale describe themselves as being hard-driving, competitive, conscientious, and more serious than other people in the same line of work. They will also state that they put forth more effort, need to hurry more often, and have been told by family members and others that they are too active and need to slow down. High scorers can be characterized by the motto, "We try harder."

The relationship between these scores and health changes is discussed in Section V.

SUMMARY

Sixteen Personality Factor Questionnaire (16 PF)

Because the 16 PF had been used in previous studies of air traffic controllers, it was selected to be administered to this study population for purposes of comparisons across ATC groups. However, as previous experiences of the ATCs with this instrument in other studies caused reluctance to cooperate in administration of the questionnaire until the last examination period, the results could not be used for predictive purposes, but could be used in retrospective comparisons of men with different health change history over the three-year study.

In comparison with 1,147 males in the U.S. comprising the original normative sample used in devising the 16 PF, the present ATC group scored significantly higher on dimensions of intelligence, ego strength, dominance, elation, group conformity, boldness, self-sufficiency, compulsivity and free floating anxiety. The ATCs scored lower than the U.S. males on dimensions of emotional sensitivity, anxious insecurity, sophistication, guilt proneness, and criticalness.

The results were consistent with those found for a national sample of air traffic controllers.

c. Sixteen Personality Factor Questionnaire

In an earlier section we described how our subjects, located at Air Route Traffic Control Centers and Towers, compared with respect to age and experience with national samples of controllers at these same types of facilities. The earlier results were used to help establish the comparability of our subjects with other controllers around the country. This section describes and interprets certain comparisons between our air traffic controllers and the original normative sample used in devising the Sixteen Personality Factor Questionnaire (16PF). In addition, the rationale and procedure for use of the 16PF in this study is explained.

Rationale

The 16PF exists in four forms (A,B,C,D) and two revisions (1962, 1967). Because a number of studies were available in which the 1962 Form A version of the 16PF was used on air traffic controllers, we selected this same version for our study even though it was not the most recent version. However, use of the 16PF presented several problems. First, the controllers were very sensitive about taking the 16PF because, according to them, there had been repercussions on individuals as a consequence of its previous use by other investigators. Hence, although the 16PF would have been most useful at the intake examination, we were unable to obtain cooperation at that time for its administration. Rather, we had to wait until the fifth and last examination before the men were sufficiently comfortable and confident in their interactions with us to allow us to administer the 16PF.

Of the 388 men present for the last examination, 387 completed the questionnaire, with one man still refusing to cooperate. For descriptive purposes, the results based on 387 are adequately representative of the results we might have obtained with the original 416.

Findings

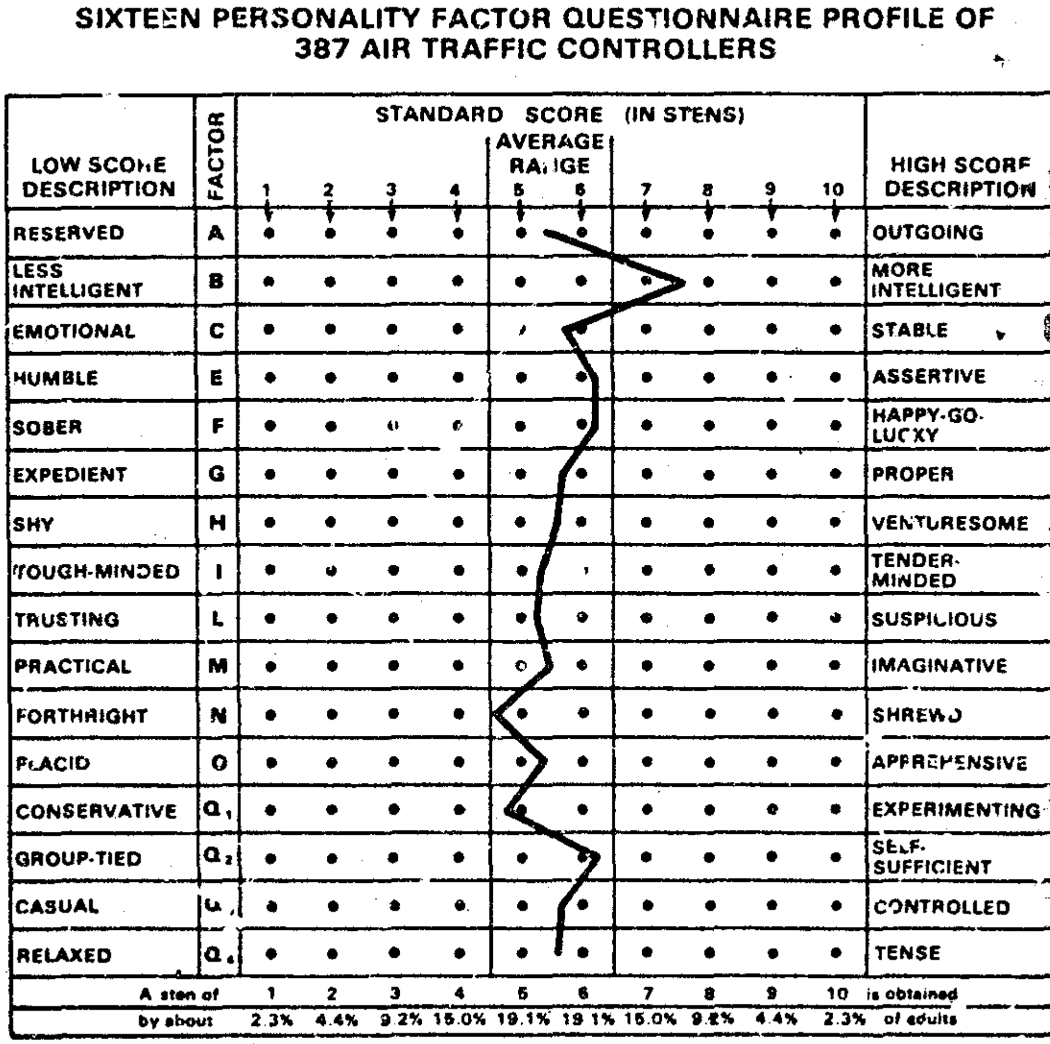
Table 13 lists the 16PF scales and provides a raw score description, the general characteristics being assessed and the high score description for each scale. Figure D plots the 16PF profile of controllers in this study with the distribution of scores among 1,147 males in the general U.S. population. The comparison indicates that controllers score significantly higher on the personality dimensions of intelligence (B), ego strength (C), dominance (E), elation (F), group conformity (G), boldness (H), self-sufficiency (Q2), compulsivity (Q3), and free floating anxiety (Q4). On the other hand, the controllers scored lower than the normative population on the personality dimensions of emotional sensitivity

TABLE 13

Sixteen Personality Factor Questionnaire
Scale Names and Descriptive Characteristics

Scale Name	Low Score Description	General Characteristic	High Score Description
A	Reserved, detached, critical, aloof, stiff	warmth	Outgoing, warmhearted, easygoing, participating
B	Dull Low intelligence	Intelligence	Bright High intelligence
C	Affected by feelings, emotionally less stable, easily upset, changeable	Ego strength	Emotionally stable, mature, faces reality, calm
E	Humble, mild, easily led, docile, accommodating	Dominance	Assertive, aggressive, competitive, stubborn
F	Sober, taciturn, serious	Elation	Happy-go-lucky, enthusiastic
G	Expedient, disregards rules	Group Conformity	Conscientious, persistent, moralistic, staid
H	Shy, timid, threat-sensitive	Boldness	Venturesome, uninhibited, socially bold
I	Tough-minded, self-reliant, realistic	Emotional Sensitivity	Tender-minded, sensitive, clinging, overprotected
L	Trusting, accepting conditions	Anxious Insecurity	Suspicious, hard to fool
M	Practical, "down-to-earth" concerns	Bohemianism	Imaginative, bohemian, absent-minded
N	Forthright, unpretentious, genuine but socially clumsy	Sophistication	Astute, polished, socially aware
O	Self-assured, placid, secure, complacent, serene	Guilt Proneness	Apprehensive, self-reproaching insecure, worrying, troubled
Q1	Conservative, respecting traditional ideas	Criticalness	Experimenting, liberal, free-thinking
Q2	Group dependent, a "joiner" and sound follower	Self-Sufficient	Self-sufficient, resourceful, prefers own decisions
Q3	Undisciplined self-conflict, lax, follows own urges, careless of social rules	Compulsivity	Controlled, exacting will power, socially precise, compulsive, following self-image
Q4	Relaxed, tranquil, torpid, unfrustrated, composed	Free-floating Anxiety	Tense, frustrated, driven, overwrought.

FIGURE D


SIXTEEN PERSONALITY FACTOR QUESTIONNAIRE PROFILE OF 387 AIR TRAFFIC CONTROLLERS
LOW SCORE DESCRIPTION
FACTOR
STANDARD SCORE (IN STENS)
AVERAGE RANGE
1 2 3 4 5 6 7 8 9 10
HIGH SCORE DESCRIPTION
RESERVED A OUTGOING
LESS INTELLIGENT B MORE INTELLIGENT
EMOTIONAL C STABLE
HUMBLE E ASSERTIVE
SOBER F HAPPY-GO-LUCKY
EXPEDIENT G PROPER
SHY H VENTURESOME
TOUGH-MINDED I TENDER-MINDED
TRUSTING L SUSPICIOUS
PRACTICAL M IMAGINATIVE
FORTHRIGHT N SHREWD
PLACID O APPREHENSIVE
CONSERVATIVE Q1 EXPERIMENTING
GROUP-TIED Q2 SELF-SUFFICIENT
CASUAL Q3 CONTROLLED
RELAXED Q4 TENSE
A sten of 1 2 3 4 5 6 7 8 9 10 is obtained
by about 2.3% 4.4% 9.2% 15.0% 19.1% 19.1% 15.0% 9.2% 4.4% 2.3% of adults

(I), anxious insecurity (L), sophistication (N), guilt proneness (O) and criticalness (Q1). These results were entirely consistent with those reported by Karson and O'Dell (1974) regarding the general personality characteristics of their national sample of air traffic controllers.

In general, the 16PF Questionnaire results indicated that the controllers were more like the general population of males than not. But, on the other hand, they had several predominant personality characteristics including intellectual orientation, compulsivity, and self-sufficiency. Along with the California Psychological Inventory results (see Section III B 3), these results indicated that the air traffic controllers were intelligent persons whose anxieties were controlled by meticulous compulsive behavior, who tended to be dominant and group conforming but who also held beliefs that authority and regulations were obnoxious. They tended to be bold and to be somewhat lacking in emotional sensitivity. Anxiety and conflict, rather than causing guilt, tended to result in conformance to group expectations. This type of conformance included verbal manifestations, but not necessarily conformance in actual behavior or physical symptoms of illness. This result has been suggested before (Karson and O'Dell, 1971; Karson and O'Dell, 1974), but in this study we had the first opportunity to validate this suggestion.

Since the 16PF was administered at the last examination, we could not use it for predictive purposes. Rather, it was included primarily for descriptive purposes and to provide retrospective comparisons of men who had different health change histories over the course of the study. Since our subject population of controllers scored so similarly to the national samples of Karson and O'Dell (1974), we could expect that our retrospective comparisons would have utility for screening purposes in later studies.

SUMMARY

PSY 130

A series of true-false scales were administered to measure the following dimensions: Hypochondriasis, Depression, Hysteria and Anxiety.

The Hypochondriasis and Hysteria measures were used as validating checks on self-reports of illness to ascertain whether or not these psychological tendencies were responsible for increased reporting of symptoms. The Depression and Anxiety scales were used as potential predictors of health change.

d. PSY 130

At the third examination a series of true-false psychological scales were administered as a single instrument which was labeled "PSY 130." This same set of 108 true-false items had been used in previous studies at Boston University School of Medicine under the name, "The Boston Health Study." The scales of this study were derived from the Minnesota Multiphasic Personality Inventory (MMPI) (see Dahlstrom, Welch, and Dahlstom, 1973, for development and validation information· also Zyzanski, et al. 1976), and consisted of the following dimensions:

Hypochondriasis. This scale is a measure of the tendency to complain about symptoms of physical ill-health, particularly under the circumstance that no organic basis can be found. Although MMPI research has found that persons diagnosed as hypochondriacs by psychiatrists score very high on this scale, it has also been established that persons with documented organic disease also score somewhat above the population average in most instances. The hypochondriasis scale was included in the ATC HCS on a single occasion as a validity check to see to what degree symptom complaints by participants might be correlated with this personality tendency rather than with demonstrable physical disease.

Depression. The standard MMPI depression scale containing 60 items was judged to be too long for most research purposes. The short form of this scale developed by Dempsey (1964), and found to have greater reliability than the original scale, was included in the PSY 130.

Hysteria. The full hysteria scale of the MMPI was included for use along with the hypochondriasis scale to determine whether hysterical hypochondriacal features were possibly making a major contribution to reports of ill health. For many kinds of health problems, the diagnosis depended far more heavily on self-report of signs and symptoms than it did on laboratory findings, and hence a check for the presence of bias and particularly for individuals who might be prone to such bias in the reporting of symptoms was a methodological precaution.

Anxiety. A short form of the Taylor Manifest Anxiety Scale developed by Bendig (1956) was the fourth and final scale included in PSY 130. The Taylor Manifest Anxiety Scale has been frequently used in research and has become established as an acceptable measure of clinical anxiety. The Bendig scale, containing the 20 strongest of Taylor's 50 items, served as an independent measure of anxiety at the third examination. Both the Bendig anxiety scale and the Dempsey depression scale were found, in a study by Zyzanski and colleagues at Boston University Medical Center (Zyzanski et al., 1976), to be associated with increased amounts of atherosclerosis as determined by coronary angiography. The PSY 130 thus contained two scales which had demonstrated association with organically determined cardiovascular illness (a general category of disease to which the ATCs were considered to be particularly vulnerable). The other two scales (HS and HY) had been shown in other studies to be associated more with complaints of disease and symptoms of psychological origin as contrasted with organically demonstrable disease. The PSY 130 was used, therefore, both as a validating check on other data and as a potential predictor of health change.

SUMMARY

Anomie

To test the hypothesis that individuals who feel alienated from society and powerless would be at greater risk for illness, five statements representing feelings of alienation (Srole role anomie scale) were embedded in the ATC Biographical Questionnaire administered at intake. The ATCs were asked to indicate their relative agreement or disagreement with the anomic indicators.

The total score distribution for the anomie scale allowed identification of three categories: 16% of the ATCs were classified as anomic, 12% as eunomic and the remaining 72% as "neither." Anomie correlated significantly (and inversely) with five measures of job satisfaction, as would be expected.

E. Anomie

Rationale

The degree to which an individual feels integrated in his interpersonal relationships and reference groups can be described on a eunomia-anomia continuum. The former refers to a person who is well integrated while the latter describes a person who feels separated from society and reference groups. The term anomie has been used generally to describe this entire continuum in which a person low in anomie would have a generalized, pervasive sense of "belongingness," whereas a person high in anomie would feel a sense of "alienation" (Srole, 1956). At the high end, anomie represents the following five components:

> the individual's sense that community leaders are detached from and indifferent to his needs;
>
> the individual's perception of the social order as essentially fickle and unpredictable such that he can accomplish little toward realizing future life goals;
>
> the individual's view that he and people like him are retrogressing from the goals they have already reached;
>
> the individual's perception that his frameworks of immediate personal relationships are no longer predictive or supportive.

Not surprisingly, anomie has been related empirically to a variety of negative factors, such as elimination of creativity, feelings of estrangement from the products of one's work, self-alienation, feelings of meaninglessness, isolation, and powerlessness (Israel, 1968; Mills, 1951, 1962; Seeman, 1959). It has also been related to mental health problems (Reinhardt, 1972; Reimanis, 1974). The study team hypothesized that ATCs who were anomic would be at a greater risk for negative health change; thus, a measure of anomie (Srole, 1956) was included.

Findings

Since anomie was seen as an enduring trait, it was measured only at intake, and embedded in the ATC BQ using the following statements:

> These days a person doesn't really know whom he can count on.

Nowadays a person has to live pretty much for today and let tomorrow take care of itself.

Most public officials (people in government offices) are not really interested in the problems of the average man.

In spite of what some people say, the condition of the average man is getting worse, not better.

It is hardly fair to bring children into the world with the way things look for the future.

Each of these statements was followed by four response categories from "strongly agree" to "strongly disagree," thus permitting a range of scores from 5 to 20 with low scores representing the more anomic responses. On the basis of the men who completed the scale at intake (n=409), it was determined from the total score distribution that three categories of individuals could be identified -- anomic, eunomic, and neither of these. Sixteen percent of the ATCs were classified as anomic, 12% as eunomic, and the remaining 72% were classified as "neither." This rather lopsided categorization occurred because of the large number of scores in the center of the total possible score distribution and the smaller frequencies of extreme scores. Since anomie was being measured by only five, strongly-worded attitudinal questions, this conservative approach to categorization seemed appropriate.

One way to determine the validity of the measure of anomie was to examine its relationship with other variables. From the empirical literature, it was expected that anomie should be related to job dissatisfaction. Since the work role is an important part of the person's life, feelings of anomie should be associated with apathy and dissatisfaction with one's job. Table 14 contains the correlations between anomie and the five job satisfaction variables. Since higher scores on the anomie variables are indicative of less anomic feelings, the relationships in the table are as expected. As can be seen from the table, anomie is significantly related to all five satisfaction variables. Furthermore, the higher values for coworker satisfaction, individual satisfaction, and group morale are consistent with our conception of the anomie variable, that is, one would expect more general feelings of job satisfaction to be more closely related to anomie than attitudes toward specific job factors. The higher value for coworker satisfaction is logical in that anomie feelings are critically dependent on one's social support system.

The relationship of anomie to health changes is discussed in Section V, Predictive Findings.

TABLE 14

Correlations Between Anomie and Job Satisfaction

faction Variable	Correlation With Anomie Score (Higher Score = less anomie)
atisfaction	.15*
ker Satisfaction	.32**
tisfaction	.16*
idual Satisfaction	.29**
Morale	.27**

.05
.01

SUMMARY

Marital, Family and Social Supports

It was hypothesized, on the basis of available information, that a strong support system would modify or reduce an ATC's risk for illness and injury. Three methods of determining support resources were:

1) 4 questions in the ATCQ on marital support;
2) 4 questions in the ATCQ on the availability of help from relatives, friends and co-workers; and,
3) several items frm the ATC BQ, including household composition, relationships with wife and children, family decision-making and agreement between ATC and wife on family goals.

Psychometric analysis showed the scales used to be acceptably reliable and to be measuring an independent variable.

Social coping resources were found to be negatively related to depression and anxiety and positively related to sociability, social presence and tolerance.

The ATCs generally perceived themselves as more dominant in family decision-making.

A large proportion of the ATCs were satisfied with their marital situation and felt in close agreement with their wives on family goals.

4. Marital, Family, and Social Supports

Rationale

One would intuitively surmise, even without reading the scientific literature, that a person's marital, family and social well-being would have a substantial impact on his physical and/or mental health. This hypothesis was reinforced by evidence from the scientific literature that work satisfaction had been related to life satisfaction in general (Kavanagh and Halpern, 1977), and to specific family characteristics such as the presence of preschool children in the household (Quinn, Staines and McCullough, 1974) and marital satisfaction (Ridley, 1973). Investigators have also found a relationship between angina pectoris and family problems (Floderus, 1974; Medalie, Snyder and Groen, 1973) as well as satisfaction with co-workers (Medalie et al., 1973); and Jenkins (1976), in a review of the literature, concluded that a number of psycho-social factors, many of them directly intertwined with an individual's social support system, affect the risk of coronary disease.

Health statistics for all developed societies, moreover, have shown that marital status is related to physical or mental well-being. It has been consistently reported that married people have lower rates of both mental and physical illness, as well as life expectancies that are different from those of single, widowed or divorced people. Recent analyses have indicated in addition that married men have an advantage over married women in lowered risk of death from those diseases that are linked to social interactions, e.g., suicide, homicide, accidental death, or cirrhosis of the liver.

Altogether, it was clear that the marital, family and social supports of the air traffic controllers in our study merited investigation and that an attempt should be made to determine whether or not these factors altered their risk for illness. The methods used to measure this aspect of the controllers' lives, and the descriptive results follow, while a discussion of the importance of these factors in the controllers' health changes is included in Section V, Predictive Findings.

a. Marital Coping Resources: The Air Traffic Controller Questionnaire

Because we hypothesized that a controller who perceived his wife to be interested and supportive would have less illness than one who perceived his wife to be uninterested and rejecting, we embedded in the Air Traffic Controller Questionnaire (ATCQ), described in detail in Section IIIB5, a measurement of the marital resources available to the ATC to cope with any felt strain or stress in his job.

The following four items were used to measure this variable:

"How often do you talk with your wife about your feelings that are a consequence of your work?

How do you feel about your wife sharing her trials and tribulations of the day when you arrive home?

How understanding is your wife of your need to unwind at the end of a day's work?

How often do you feel embarrassed by many things your wife says and does"

Based on the results of a pilot study and other psychometric evidence described in Section IIIB5, a scale was developed, using responses to these four questions as a measure of the ATC's perception of his marital coping resources. The purpose of the remainder of this section will be to discuss the psychometric properties of this scale to determine the quality of the measurement of this variable.

Measurement Qualities of the Marital Coping Resources Scale

Based on the responses of the 410 men who completed the ATCQ during the intake examination, the internal consistency reliability of the marital coping resources scale was .71, an acceptable reliability under established measurement standards. The other major form of reliability of considerable importance to effective measurement is test-retest reliability, which reveals the stability of the measurement of the variable over time. Based on the responses of 383 men who completed the ATCQ on Rounds 1 and 2, the test-retest reliability of the marital coping resources scale was .69, an acceptable value, particularly since test-retest values are usually lower than internal consistency reliabilities. It was concluded, therefore, that the scale measuring marital coping resources was adequate in internal consistency and in stability over time.

Along with determination of the scale's reliability, we needed to determine its validity. Although it was not possible to determine whether or not our four items in fact defined and measured what was intended, i.e., marital coping resources, it was possible to determine whether or not the scale at least was a measure of an independent dimension. This test was made by examining the relationships between various scales. A demonstration that no relationship exists between measurement of one variable and others, where no relationship is expected conceptually or empirically may be considered positive evidence for the independence of the measures. There was no reason to expect the marital coping resources scale to be correlated with any of the other ten scales in the ATCQ or with the industrial psychological variables described in IIIB5, and none of the correlations with these variables did, in fact, exceed .20. On the basis of these results it was concluded that the measurement of marital coping resources as defined by our questions was sufficiently independent of these other self-report measures.

As discussed in Section IIIA2, a variety of measures from the research project of Caplan et al. (1975) were included in Round 5 so that our results could be compared with the results from their ATC samples. Among the variables included were measures of social support provided by "your wife, friends, and relatives," "your immediate supervisor," and "other people at work," in the following four situations: (1) making the ATC's work life easier; (2) ease of talking to these people; (3) being able to rely on these people when things get tough at work; and (4) how willing these people were to listen to the ATC's personal problems. The correlations between responses concerning these three social support groups and those on the marital coping resource scale are presented in Table 15.

TABLE 15

Correlations Between Marital Coping Resources and Social Support Variables
(N = 343)

	Marital Coping Resources
Social Support - Supervisor	.07
Social Support - Coworker	.07
Social Support - Wife, Friends, Relatives	.65

Marital coping resources were not expected to correlate with the first two social support variables, and as expected, the correlations were statistically non-significant. However, the rather close relationship with the third social support variable involving the wife provided further validity evidence for the marital coping resources scale.

b. <u>Social Coping Resources:</u>
<u>The Air Traffic Controller Questionnaire</u>

Along with the marital coping scale, the ATCQ also contained a measure of the social resources available to the ATC to cope with any felt strain or stress on his job. Again the hypothesis was that a strong social support system consisting of friends, coworkers, or relatives would be associated with less health change than a weaker support system.

Four items were used to assess the ATC's perception of his social coping resources. The first three asked: "If you or your family were struck by a crisis or tragedy, how many persons of the following categories could you really count on to help you?" This question was followed by "relatives," "friends," and "people at work" to complete the items. The ATC was to use a scale from "none" to "8 or more" to answer these items. The last item asked: "How many people do you consider <u>close friends</u> who live within an hour's drive of your home?" The ATCs responded on the same scale as for the first three items. Based on the results of a pilot study and on psychometric evidence presented in Section IIIB5, these four items were combined into a scale to measure the ATC's perception of his social coping resources.

<u>Measurement Qualities of the Social Coping Resources Scale</u>

Using the responses of 410 men who completed the ATCQ during the intake examination, the internal consistency reliability of the social coping resources scale was determined to be .70, which is acceptable under measurement standards. Based on the responses of 383 men who completed the ATCQ on both Rounds 1 and 2, the test-retest reliability was .63, which is also acceptable. It was thus concluded that this scale had a good degree of internal consistency and test-retest stability.

This scale was not expected, conceptually or empirically, to be related to the other scales on the ATCQ or to the industrial psychological variables described in Section IIIB5. The results of the correlational analysis verified this expectation in that none of the correlations exceeded .20. It was thus concluded that this scale was measuring this variable independently of the other self-report measures.

In accordance with the work of Kasl and Cobb (1970), social coping resources were expected to be related to certain personality characteristics. Since both the CPI and certain scales from the MMPI (see Section IIIA3) were administered during the intake visit, we were able to correlate these scales with the social coping resources scale. Social coping resources were associated with low MMPI measurements of depression, anxiety, and the tendency to admit symptoms, and with a positive tendency to deny symptoms. These findings are in agreement with those of Kasl and Cobb. High social coping resources were also related to high sociability, social presence, and tolerance dimensions on the CPI. Considering the meanings of the CPI scales, these results are also supportive of the validity of our measurement of social coping resources. It should be noted that all of these correlations were in the .20's, indicating moderately low relationships.

As described above, three social support variables were measured in Round 5. Table 16 contains the correlations between the social coping resources scale and the three social support measures.

TABLE 16

Correlations Between Social Coping Resources and Social Support Variables

(N = 387)

	Social Coping Resources
Social Support - Supervisor	.20
Social Support - Coworkers	.36
Social Support - Wife, Friends, Relatives	.24

Although all three values are statistically significant, the higher value for the social support by coworkers is in agreement with the emphasis in the social coping resources scale. The other lower, but statistically significant, values are also understandable since the social coping resources scale included friends, relatives, and people at work. These results provide further evidence for the validity of the measurement of social coping resources in the ATCQ.

c. Biographical Items As Possible Supports

Household Composition

Eight-seven percent of the ATCs had children living with them (2.3 average). In addition, as many as 14% had other persons (not including spouse) living with them, such as relatives, in-laws, or friends. We combined marital status and other persons in the home to measure household composition for the sample. Not unexpectedly, the largest percentage of ATCs (66%) lived with a wife and children, and no other persons in the home. Seven percent lived alone, 4% lived with a wife only, while 2% lived in a broken home with their children. Interestingly, 17% lived in what we called a complex home environment, which meant that the ATC was living with friends or "others," regardless of marital status. Finally, 3% were living in "extended" home environments, which meant the presence of relatives or in-laws regardless of the ATC's marital status.

Relationship With Children

Seventy percent of the ATCs in this study reported they had taken their children somewhere during the week preceding administration of the ATC BQ. When asked if they had taken their children somewhere during the past month, only 6% responded that they had not taken their children to an eating place, the movies, some entertainment or recreation, or to visit a friend.

Family Decision-Making

Two important areas measured in the ATC BQ were the manner in which family decisions were made and the agreement in family goals between the ATC and his wife. Note that these variables were measured in terms of the ATC's perceptions; thus, they may not accurately assess the wife's perceptions and/or the objective situation. Since the ATC is the unit of study, his perceptions were considered to be the important ones for their possible impact on his health.

The ATC's perception of family decision-making was measured on the basis of nine major types of family decisions: the husband's job choice, family car, life insurance purchase, vacation plans, house choice, wife going to work, choice of doctor, food budget, and discipline of the children.

The ATC was asked whether the decision was made (1) by the husband always, (2) by husband more than wife, (3) by husband and wife exactly the same, (4) by wife more than husband, or (5) by wife always. Not surprisingly, there were considerable differences in the way the ATCs described the different decision situations. The husband was more dominant in making the decision on his job choice, the choice of a car, and the purchase of life insurance. Husband and wife shared, in a large percentage of households, decision-making on where to go for a vacation, house or apartment choice, and discipline of children. Finally, decisions on whether the wife should work, choice of doctor, and family food budget were made often by the wife; however, the percentage endorsement by the ATCs on these items was not nearly as strong as for the first two sets of decisions. In fact, by summing across all nine items, an overall index of family decision-making was derived. The distribution of the ATCs' scores is clearly skewed such that husband dominance in family decision-making is more prevalent, a finding in agreement with anecdotal and interview data from the study.

Family Goals

The ATCs were asked to rank on a scale from one to ten the importance they attached to ten different family goals, and also to rank the same goals a second time as they thought their wives would rank them. Measuring the differences in rankings provided an index of value consensus on family goals between the ATCs and their wives. The distribution of these scores indicated that 67% of the ATCs saw themselves and their wives in fairly close agreement on family goals, 23% perceived moderate agreement, and 10% saw disagreement on family goals.

Husband-Wife Relationships

The final two items on the ATC BQ dealing with husband-wife relationships inquired into the frequency with which the ATC took his wife out, and his overall marital satisfaction. Seventy-nine percent of the ATCs reported that they had taken their wife outside the home socially at least once or more during the week preceding their completion of the ATC BQ, and the percentage jumped to 96% when the reporting period was increased to a month. Eight-four percent reported they were currently satisfied with their marriage, 7% had mixed feelings, and 9% expressed dissatisfaction.

In general, these descriptive results indicate that a large proportion of the ATC group was satisfied with their marital situation at the time of entry into the study; however, a proportion of the men indicated on all of the indices a less than satisfactory state of marriage.

The extent to which these various coping resources among the study group relate to their health change is discussed in Section V, Predictive Findings.

5. Job-Related Characteristics

Among the possible predictors of health change that we wished to test were a number of job-related characteristics, some associated specifically with the air traffic controller occupation, and others relevant to jobs generally.

We wanted to explore whether or not health changes or specific illnesses were in any way related to differences among the ATCs in such specific variables as:

i. attitudes towards their skills, and their psychological "investment" in their work;

ii. the affective discomfort produced by their work;

iii. styles or ways of coping;

iv. attitudes toward the FAA;

v. competence on the job.

We also wanted to assess general and specific satisfactions of the ATCs with their work, and we wanted to know how they perceived the behavior of their leaders.

Our efforts to define characteristics and to find or devise methods of measuring them are related below, together with the results of the measurements taken. The extent to which differences in these characteristics among the ATCs were related to health outcomes is discussed in Section V, Predictive Findings.

SUMMARY

Specific ATC Variables

The ATC Questionnaire, administered at all 5 rounds measured psychological and attitudinal dimensions specific to the occupation, including 11 factors: Psychophysiological anxiety reaction, good controller, coping-by-drinking, social coping resources, work avoidance on bad days, subjective cost, bounceback-burnout, investment in work, tension discharge rate, coping by physical activity, marital-coping-resource. Correlations with several standard personality inventories showed the following:

1) There was little overlap between the job-specific psychological characteristics and general personality traits assessed by the CPI;
2) There were few, and very low, correlations with the coronary-prone behavior pattern;
3) High social coping resources were associated with low depression, low anxiety, high tendency to deny symptoms;
4) Psychophysiological anxiety was positively correlated with MPI hypochondriasis, depression, anxiety, and tendency to admit symptoms.

Items in the ATCQ intended to measure satisfaction with management, with training and with promotion policy revealed that the ATCs were considerably dissatisfied with all three.

Nominations for competence by peers were not associated with any unique psychological characteristics. Nominations for amicability were related negatively to dominance and self-acceptance.

a. Specific ATC Variables

During the first year of the ATC HCS, we realized that there were several psychological dimensions of being an air traffic controller that could not be measured with standard, available psychological tests. For example, although the phenomenon of "burnout" among ATCs had been reported to us many times by controllers and FAA management, there was no existing method for its measurement, and it had never been directly assessed. Similarly the capacity for "bounceback" had been described to us as a controller's ability to adapt quickly and easily between heavy and light workload. The bounceback had been described and discussed informally and clinically, but no published assessment procedure was available. Consequently, we undertook the major task of defining a number of constructs that seemed particular to the air traffic controller and his job, and then devising questionnaire scales that would assess these constructs reliably and validly. The resulting instrument was called the ATC Questionnaire.

Investigators and psychological consultants for this study spent many hours in air traffic facilities familiarizing themselves with the ATC work environment. Many group and individual conversations were held with ATCs, supervisors, and FAA officials. A few ATCs became particularly interested in this work, and their advice and suggestions were extremely valuable; particularly George Kerr and Ted Paolotti of the New York Air Route Traffic Control Center were instrumental in this preliminary work.

Development of A Priori Scales

After eight months of observations, discussion, and immersion in the ATC environment, the investigators held formal meetings to define the variables of salient interest. With the definitions and constructs in mind, the investigators then wrote more than 1,000 questions. Further refinements and the elimination of redundant items led to the preliminary selection of 206 items defining 18 different constructs.

This original list of 206 items was compiled into a questionnaire and pilot tested in the Oakland, California, Air Route Traffic Control Center with 55 air traffic controllers.

The results in the pilot test were carefully analyzed in accordance with several criteria (see Appendix 1), and 147 items were retained. These 147 items were administered to all 416 men in this study population during the entrance examination at Boston University Medical Center. Another item analysis was performed. The second analysis resulted in shortening the questionnaire to 95 items.

The 95-item questionnaire was administered at the remaining four examinations. Two additional scales were devised and administered beginning with the second examination. These scales were concerned with symptoms of anxiety caused by incidents and conflictions, and anxiety symptoms related to giving on-the-job, live-traffic training. In all, we developed scales to assess 20 ATC-specific job characteristics. The definitions of these characteristics are displayed in Table 17, along with data on the internal consistency and stability of each of these scales.

Internal consistency reliability was important as an indicator of how well the scale's items converged on a common construct. Internal consistency is high when items overlap and are somewhat redundant. On the other hand, when items are independent of one another, there is low internal consistency. Coefficient alpha (Nunnally, 1967, p. 193) was used to measure internal consistency. The results shown in Table 17 indicate that the internal consistency was reasonable for most of the longer scales, but it was very poor for a number of the short scales, particularly the coping scales. We did not consider all of these scales entirely adequate and therefore performed another series of analyses and refinements. However, in spite of some of the poor internal consistencies, a number of these *a priori* scales were considered sufficiently important from the conceptual and explanatory standpoint to be used for predictive purposes.

Another important characteristic was the test-retest reliability of the scales, referring to the stability of the measurement of the constructs and the stability of the constructs themselves. With perfect measurement, low to moderate stability would indicate that the construct was a "state variable" with wide individual variation over time. On the other hand, high stability over repeated testing would indicate that the construct was a trait of enduring substance. The test-retest reliability was derived from scores on the first and second administrations of these job-specific scales. The results are presented in Table 17.

As shown in the Table, 15 of the 20 scales had a test-retest stability of .54 or greater, indicating reasonable long-term (9-month) stability. Several of the scales with the lowest internal consistency had higher retest stability. Since internal consistency sets the upper limit for the true reliability of assessment (Nunnally, 1967, p. 210), the scales with higher test-retest stability than internal consistency were influenced by stable response modes rather than by stable measurement of the target construct.

Development of Job-Specific Psychological Factor Scales

Consequently, our dissatisfaction with the psychometric characteristics of all 20 *a prior* scales led us to conduct a number of

TABLE 17

ATC Job-Specific Scales And Definitions

Bounceback - four items. Internal consistency = .69 Stability = .49.

The four items composing this scale were devised to assess an ATC's self-perceived ability to cycle between heavy traffic and light traffic conditions at work. This construct represents a flexibility in coping and adaptation to fluctuating workload. The hypothesis was that low bounceback abilities were related to the development of psychological and physical illness, whereas a high level of bounceback was protective and adaptive.

Burnout - five items. Internal consistency = .68 Stability = .63.

This construct was conceptualized as the premature exhaustion of a man's physical, cognitive, and psychological resources leading to early retirement for medical and/or psychological reasons. It is characterized by a loss of ability to control aircraft in any but the easiest circumstances and a heightened sense of anxiety and fear when called upon to do so. Due to the sensitivity of this issue, the items were phrased projectively and impersonally rather than with first person self-evaluations.

Toughmindedness - six items. Internal consistency = .35 Stability = .60.

This construct referred to a man's emotional style while working. That is, did the man feel he let situations get to him, or could he pass them off and get on with the job of separating aircraft. We expected an inverse relationship with burnout.

Investment - eight items. Internal consistency = .68 Stability = .65.

Investment was conceptualized as a man's involvement with his job and particularly, the strength of his identification with being an air traffic controller. In a sense the items were devised to assess how important it was for the ATC to be an idealized controller of exceptional character and ability. This intensity of involvement was expected to be a moderator variable in that those men with a high degree of investment but low coping and controlling abilities would be more susceptible to health change than ATCs of similar abilities but with low investment.

Tension Discharge Rate - eight items. Internal consistency = .82 Stability = .67.

This construct referred to an ATC's felt ability to relax and unwind.

TABLE 17 cont'd.

Questions were phrased in terms of rate of tension release. It was hypothesized that those who could release tensions quickly would be less susceptible to health change than those who had difficulty discharging tensions.

6. Tension Discharge Mechanisms on Bad Days - six items. Internal Consistency = .70 Stability = .57.

The items assessing this construct were a list of possible means of releasing tension during difficult days at work. Theoretically we were considering this construct as an indication of tension release effectiveness. This conception of tension release effectiveness assumed that a person who reports that no mechanism is utilized on bad days must (a) not be affected by bad days, or (b) discharge the tensions so rapidly and effectively that the mechanism is not recalled. On the other hand a person who uses more mechanisms might (a) use less effective mechanisms, and/or (b) have poor utilization of the available mechanisms. We hypothesized that this variable would be inversely related to health change.

Before Work Anxiety - six items. Internal consistency = .88 Stability = .59.

Six psychophysiological symptoms were listed after an introduction which asked how often the ATC experienced each of the symptoms before going to work when the weather or air traffic conditions were bad. This construct was meant to assess anticipatory anxiety. It was expected to have a positive relationship with health change.

During Work Anxiety - seven items. Internal consistency = .84 Stability = .65.

Seven more psychophysiological symptoms were listed. The ATCs were asked how often they experienced each of them during difficult air traffic situations. Again, a positive relationship with health change was hypothesized.

After Work Anxiety - six items. Internal consistency = .81 Stability = .65.

Yet another six psychophysiological symptoms were listed. The controllers were asked how often they experienced the symptoms after being relieved from a long period of heavy traffic. We hypothesized that all three temporal types of anxiety would be related to each other as well as to health change. In addition, we hypothesized that individuals who did not consistently report the symptoms across the three time periods may differ in health change and/or on-the-job physiological responsivity.

Subjective Costs - eight items. Internal consistency = .74 Stability = .60.

This scale was concerned with an ATC's subjective evaluation of how much his job took away from eight areas in his life (such as family time, social life, and so forth). In other words, how much did it "cost" him

TABLE 17 cont'd.

to be an ATC? It was expected that those who felt higher levels of "cost" would be more susceptible to health change than those who felt lower levels.

Shift Change Cost - three items. Internal consistency = .47 Stability = .36.

This short scale assessed the felt "cost" due to changing shift schedules. Many men had complained that the rapid rotation of shifts interfered with many aspects of their lives, including their sense of well-being and social membership. Again, we expected a positive correlation between shift change "costs" and health change.

Coping by Drinking - six items. Internal consistency = .57 Stability = .76.

We devised several coping scales. The coping-by-drinking scale assessed how often the ATC consumed alcoholic beverages in order to unwind, relax, or get rid of problems. The questions were phrased in terms of frequency and purpose but not quantity. We hypothesized a positive relationship with health change.

Coping by Physical Activities - four items. Internal consistency = .53 Stability = .54.

This coping scale was concerned with the usage of physical activities (sports and active hobbies) for the purpose of relaxation, tension release and the like. A negative relationship with health change was expected.

Coping by Outside Activities - four items. Internal consistency = .28 Stability = .58.

Hobbies, another job, and handyman tasks were the focus of this scale. Once again, items were phrased in terms of frequency and purpose. A negative relationship with health change was expected.

Coping by Socializing - three items. Internal consistency = .38 Stability = .59.

These three items assessed the frequency of using social situations (group conversations, parties, and so on) for unwinding and relaxation. A negative relationship with health change was hypothesized.

Coping Mechanism Effectiveness - three items. Internal consistency = .15. Stability = .42.

This short scale was composed of three overall summary items inquiring about the effectiveness of the behaviors reported in the coping scales. A negative relationship with health change was expected.

17. Marital Coping Resources - four items. Internal consistency = .71 Stability = .69.

Four items were devised which specifically queried the ATC's perception of marital support. It was hypothesized that the perception of an interested, supportive spouse would be associated with less health change than the perception of an uninterested, rejecting spouse.

18. Social Coping Resources - four items. Internal consistency = .70 Stability = .63.

The number of friends, coworkers and relatives on whom the ATC could rely for help or support were three items on this scale. The last item assessed the number of close friends living within an hour's drive of the ATC's home. It was theorized that this scale would reflect the extent and depth of the ATC's social support system. The hypothesis was that a stronger social support system would be associated with less health change than a weaker social support system.

19. Anxiety Due to Incidents - six items. Internal consistency = .60 Stability = .44.

These six items listed symptoms of anxiety that might be elicited by having involuntary incidents and conflictions (violation of separation standards). It was hypothesized that men who were upset greatly and who took longer to recover from their upset would be at higher risk for health change.

Anxiety Due to Training - six items. Internal consistency = .74 Stability = .41.

It was reported that giving on-the-job, live-traffic training was particularly stressful. These six items inquired about how anxiety-provoking certain aspects of this experience were for the respondent. A positive relationship to health change was expected.

* * * * * *

factor analyses to arrive at the most reliable and valid scales from the standpoint of psychometric characteristics. The factor analyses all used a common factors approach in which the shared variance between items was factored rather than the total variance, thus excluding chance correlations between error variances (Harmon, 1960). The squared multiple correlation between a given item and all others was used as the initial communality estimate. After factoring, a rotational method was applied to clarify the factor structure. Presented in Table 18 are the results from a common factors solution with an orthogonal (Varimax) rotation since this was the most understandable methodology. Other factoring methods and other rotational solutions yielded the same results and therefore are not reported.

Description of Job-Specific Psychological Factors Content and Reliability

As seen in Table 18, eleven factors were found to subsume significantly most of the items originally devised for the twenty *a priori* scales. These eleven factors were significant by Kaiser's (1960) and Cattell's (1966) criteria. They accounted for 89.4% of the common variance among the items.

The item loadings and the item content for each of these factors is presented in Appendix I, Section III B 5. A brief discussion below of each scale includes its primary content in meaning and its composition with respect to the *a priori* scales.

The psychophysiological anxiety reaction factor was composed of 23 items with a high degree of both internal consistency and nine-month stability (see Table 18.) Basically, it was composed of the items describing psychophysiological anxiety symptoms before, during and after work. These symptoms were all phrased in the context of various ATC work situations, resulting in a factor that is completely job-specific.

The good controller factor was composed of 12 items, all of which were phrased as positive self-evaluative statements with high agreement indicating a higher level of their positive quality. Consequently this factor may have represented a social desirability factor for air traffic controllers, i.e., the extent to which they respond in a manner that wins approval. This factor was the only new combination of items resulting from the factor analysis. The *a priori* scales were not highly related to it.

The coping-by-drinking factor was very similar to the *a priori* "drinking-to-cope" scale, with which it correlated .97. The factor analysis merely shortened the scale and slightly improved the reliability.

TABLE 18

ATC Specific Job-related Characteristics

[illegible]TOR "NAME"	NUMBER OF ITEMS	INTERNAL CONSISTENCY	9-MONTH TEST-RETEST STABILITY
[illegible]hophysiological [illegible]iety Reactions	23	.92	.74
[illegible] Controller	12	.51	.46
[illegible]g by Drinking	8	.81	.80
[illegible]l Coping Resources	4	.70	.63
[illegible] Avoidance on Bad Days	5	.71	.55
[illegible]ctive Costs	6	.73	.58
[illegible]back - Burnout	4	.80	.65
[illegible]tment	6	.66	.64
[illegible]ion Discharge Rate	6	.82	.65
[illegible]g by Physical Activity	4	.50	.60
[illegible]l Coping Resources	4	.71	.69

Social coping resources emerged from the factor analysis identical to the a priori social coping scale. The same four items were involved; therefore the reliabilities remained as before. The a priori scale definition remained unchanged.

A new scale created by the factor analysis was named "work avoidance on bad days." The five items in this scale had much in common with the a priori scale called "tension discharge mechanisms on bad days." The content was slightly tightened by the factor analysis yielding good internal consistency and nine-month stability. The definition for this factor is essentially the same as it was for the a priori "tension discharge mechanisms on bad days."

Subjective cost also came out as a factor nearly identical to an a priori scale. Again the factor analysis merely tightened the item structure. The definition for the subjective cost factor is the same as for the a priori scale.

A bounceback-burnout factor was identified which was equally split between the a priori bounceback and the a priori burnout scales. The new combined factor scale was composed of two burnout items and two bounceback items which together had very good internal consistency and quite reasonable stability over time. High scores on this scale were equivalent to having a high level of bounceback and a minimal level of self-reported burnout, whereas a low score corresponded to a high level of self-reported burnout and low bounceback abilities.

An investment factor, identified by the factor analysis, was also almost identical to the a priori factor of the same name. The meaning and reliability of this factor were the same as for the a priori scale.

Tension discharge rate was a construct identified by both the a priori analysis and factor analysis. The meaning and reliability stayed essentially the same for the factor-analyzed scale.

The coping-by-physical activity factor also was very similar to the a priori scale of the same name. One item was dropped and a new one added as a result of the factor analysis. The reliability figures were improved, but the meaning was the same. Men scoring high on this factor would tend to participate in physical activities to cope with tension and anxiety.

Finally, Table 18 showed that the marital-coping-resource a priori scale also was identified by the factor analysis. The same four items were involved, and therefore the meaning and reliabilities remained the same.

In summary, the scales resulting from the factor analysis were equal or superior in psychometric characteristics to the a priori scales. In addition, the factor scales were highly interpretable since their content was so carefully described and highly related to the a priori scales. Consequently, for most analyses presented in this health change study, the factor-analyzed job-specific psychological scales were used as predictors and correlates of other variables. On occasion, however, where a particular variable was of extreme interest to the contracting agency (for example, the effects of shift changes), we have used the corresponding a priori scale. However, the remaining descriptive results are applicable only to the factor scales and by inference to their most closely related a priori scales.

Relationship of Job-Specific Psychological Factors to General Job-Related Variables

Since the 11 job-specific factor scales formed a core set of the predictor variables for health change, we considered it important to ascertain their validity to whatever extent was possible. We used the device of assessing concurrent validity, i.e., correlational relationships at a given point in time between two or more measures of related constructs. If the measures are intended to assess aspects of a common construct and if they have significant correlations with one another, there is evidence of convergent validity. If measures are intended to assess different constructs and have very low correlations with one another, there is evidence of discriminant validity. Our purpose in developing these factor scales was to assess job-specific psychological characteristics rather than general psychological traits. Consequently, we expected low magnitude orrelations with other scales that were more general in nature. he following series of results documents a number of these correlative validities. Another validity of great importance is predictive lidity, which is discussed with other predictive findings in Section V.

Table 19 presents the significant correlations between the job-ecific psychological factors and certain standard industrial psy-ology scales. A description of the standard industrial psychology les is provided in Section III B. The main point of the results sented in Table 19 is that all of the correlations between the -specific factors and the standard industrial scales were low. se low correlations indicated that the 11 job-specific psychologi- factors were relatively independent of the standard published ts of more general characteristics.

Specifically, satisfaction with work was positively related to good controller factor and the bounceback factor, indicating a controller felt more satisfied with his work if he himself he was a good controller and had good abilities to bounce back

TABLE 19

Significant Correlations Between Job-specific Factors and Standard Industrial Psychology Scales*
(N = 392)

Job Description Index

Satisfaction with Work

1.	Good Controller	0.27
2.	Subjective Costs Factor	-0.20
3.	Bounceback-burnout Factor	0.29

Satisfaction with Coworkers

1.	Psychophysiological Anxiety Factor	-0.25
2.	Subjective Costs Factor	-0.20
3.	Bounceback-burnout Factor	0.20
4.	Tension Discharge Rate Factor	0.21

Satisfaction with Pay

No Significant Correlations equal to or over 0.20

vanagh Life Attitude Profile

Individual Morale

1.	Psychophysiological Anxiety Factor	-0.42
2.	Subjective Costs Factor	-0.31
3.	Bounceback-burnout Factor	0.29
4.	Tension Discharge Rate Factor	0.31

Group Referenced Morale

1.	Psychophysiological Anxiety Factor	-0.29
2.	Subjective Costs Factor	-0.27
3.	Tension Discharge Rate Factor	0.26

r Behavior Description Questionnaire

1. Initiation of Structure

No significant correlations equal to or over 0.20

2. Consideration of Others

Psychophysiological Anxiety Factor	-0.22

3. Tolerance of Freedom

Psychophysiological Anxiety Factor	-0.26

correlations $\geq .20$, $p < .001$, are reported between high and low workload conditions. On the other hand, if he felt that he was paying a large subjective cost in many areas of his life for being an air traffic controller, he tended to be less satisfied with his work. These correlation results were evidence for convergent validity of the good controller, subjective costs, and bounceback-burnout job-specific psychological factors.

The Kavanagh Life Attitude Profile provided two scales of interest, individual (self-referenced) morale and group morale (one's morale considering one's peer group issues). Several job-specific psychological characteristics had significant relationships to these two general characteristics. Men with high individual morale tended to report fewer psychophysiological anxiety symptoms and fewer subjective costs as a consequence of being an air traffic controller. In addition, men with high individual morale also tended to report higher bounceback and higher rate of tension discharge.

Similar results were found with respect to group-referenced morale. This type of morale taps a man's sense of satisfaction with his job in more group-oriented terms such as quality of supervision. High group-referenced morale was negatively related to psychophysiological anxiety and subjective costs, and positively related to tension discharge rates. Both of these findings with respect to the Kavanagh Life Attitude Profile scales provided additional evidence regarding the meaning of the job-specific scales and their convergent validity.

The Leader Behavior Description Questionnaire required air traffic controllers to describe their immediate supervisors' behavior. Since they were describing someone else, one would expect low correlations with reports about themselves. In fact, Table 19 shows that there were only two very low correlations between the job-specific psychological scales and the Leader Behavior Description Questionnaire scales. The only significant correlations were between anxiety and a supervisor's consideration of other people and a supervisor's tolerance of freedom. In both of these cases, high anxiety was associated with descriptions of supervisors as having little consideration of other people and having low tolerance for workers' freedom. These results from the Leader Behavior Description Questionnaire provided evidence of discriminant validity as previously discussed.

<u>Relationship of Job-Specific Psychological Factors to General Personality Traits</u>

Originally, we believed that there were ATC psychological characteristics which were not appropriately assessed by widely-used personality tests. As we demonstrated, this was certainly true for job-related attitudes such as work satisfaction, co-worker satisfaction, and so forth. The next series of results presents data on the relationship between the California Psychological Inventory (CPI) scales, several Minnesota Multiphasic Personality Inventory (MMPI) scales, and the job-specific psychological factors.

The CPI (see Section III B 3) assesses 18 personality traits which are listed and described in Table 12. As this inventory was administered at the second examination at Boston University Medical Center, the following results are based on the testing of 391 of the 392 controllers who answered the CPI validly.

Table 20 displays correlations between the CPI personality traits and the job-specific psychological factors. Only these correlations of .20 or greater, $p < .001$, are reported. The two additional anxiety scales (anxiety caused by training or by incidents) are included since they were not part of the factor analysis. As can be seen in Table 20, most of the correlations were in the expected direction.

The CPI traits are scaled such that more of a "desirable" trait is reflected by a higher score (Megargee, 1972). The job-specific psychological factors are scaled such that a higher score represents more of the ATC characteristic. Some ATC psychological characteristics were considered to be negative or undesirable from a theoretical standpoint inasmuch as they would be risk factors for illness (e.g., the anxiety factor). Other ATC psychological factors represented characteristics that we felt were protective from illness (e.g., bounceback). Thus, for example, the positive correlations of the CPI Well-Being scale with marital coping resources, bounceback and tension discharge rate were in the expected direction (all low, but positive). Similarly, the negative correlations of the CPI Well-Being scale with the anxiety, coping by work avoidance, subjective costs, anxiety caused by incidents, and anxiety caused by training also were in the expected direction (negative and low).

Although they were in the expected direction, the correlations between the job-specific psychological factors and CPI personality traits were all very moderate in magnitude. Our decision to devise ATC job-specific psychological scales was supported by the results shown in Table 20 which indicated very little overlap between the job-specific psychological characteristics and the general personality traits assessed by the CPI.

Table 20 shows that subjects in this health change study were neither attempting to make themselves look worse nor better than they really were when answering the job-specific psychological items. The low correlations between the CPI Well-Being scale and the 11 factors and 2 additional anxiety scales, and the very low correlation between the CPI Good Impression scale and the job-specific factors and anxiety scales provide some evidence that they were not "faking." Finally, there was no evidence whatsoever of any tendencies for random answer-

TABLE 20

Job Specific Factor Scale
Correlations with CPI
Personality Traits *
(N=391)

CPI ...ility Trait	Marital Coping Resources (N=350)	Social Coping Resources	Psychophysiological Anxiety	Good Controller	Coping by Drinking	Coping by Work Avoidance	Coping by Physical Activities	Subjective Costs	Bounceback	Investment	Tension Discharge Rate	Anxiety due to Training	Anxiety due to Incidents
...ce	20												
...ty for status													
...ility		23	-23	25				-20	24				
...Presence		20	-26					-24	26				-20
...ceptance													
...ing	20		-38			-20		-26	31		27	-20	-22
...bility													
...zation	22				-24								
...ntrol			-23		-28				21		20		
...ce	23	20	-22					-23	23				-20
...pression	22		-25		-22			-21	23		20		
...ility													
...ent-...ity	21		-26		-25			-24	28		21		
...ent-...ence									22				
...tual ...cy								-26	27				
...gical-...ss			-22										-22
...lity													
...ty													-22

*...orrelations ≥.20, $p < .001$, are reported; decimal points omitted.

ing on the job-specific factors and anxiety scales. Thus we could be reasonably sure that the scores on the job-specific psychological factors were not unduly distorted by possible "faking good," "faking bad," or random answering.

The subjects in this study also completed the Jenkins Activity Survey (JAS) on their second visit to BUMC (see Section III B 3) and several short MMPI scales on their third visit (see Section III B 3). The results from the JAS and the MMPI scales are presented in Table 21. There were few and very low significant correlations between the JAS personality dimensions and the ATC job-specific psychological characteristics. The results indicated that the job-specific psychological characteristics were not particularly associated with the four main dimensions of the coronary-prone behavior pattern (Jenkins, 1972). However, there were a number of correlations between the job-specific psychological characteristics and those assessed by the short MMPI scales.

High social coping resources were associated with low depression, low anxiety, low tendency to admit symptoms, and high tendency to deny symptoms. These findings were reasonable and expected since they confirmed the data and findings of Kasl and Cobb et al. (1970).

The psychophysiological anxiety factor was positively correlated with the MMPI hypochondriasis, depression, anxiety, and tendency to admit symptom scales, but it was negatively associated with the denial of symptoms index. Thus although the psychophysiological anxiety symptoms were specifically related to ATC work situations, these data suggested that these symptoms also were related to more generic and psychological traits.

Other job-specific psychological scales with a number of moderate MMPI correlates were subjective costs, bounceback, tension discharge rate, and anxiety due to incidents. High scores on the subjective cost and anxiety caused by incidents scales seem to be more related to general psychological traits since they had moderate, positive correlations with the psychopathological characteristics in the MMPI scales. On the other hand, high scores on bounceback and tension discharge rates were more indicative of healthy psychological functioning, since they had negative, though moderate, correlations with the MMPI scales.

Attitudes Toward the FAA

After we had completed the intake examinations, we received considerable feedback from the study participants that there were gaps in the coverage of ATC job attitudes by the Job Description Index (JDI) scales (see Section III B 5). This is a fairly common phenomenon in attitude research and measurement in organizations since general scales

TABLE 21

Job Specific Factor Scale
Correlations with MMPI(N=351) and
Jenkins Activity Survey(JAS-N=390)
Traits*

	Marital Coping Resources**	Social Coping Resources	Psychophysiological Anxiety	Good Controller	Coping by Drinking	Coping by Work Avoidance	Coping by Physical Activities	Subjective Costs	Bounceback	Investment	Tension Discharge Rate	Anxiety due to Training	Anxiety due to Incidents
A Person- (JAS)											-24		
and In- nce (JAS)			31					20	-21		-32		
nvolve- (JAS)<													
driving				25									
ondriasis			36					25	-35		-22		
sion(MMPI)		-26	41	-26				26	-37		-28		26
a (MMPI)													21
(MMPI)		-20	43					30	-30		-37	20	26
ymptoms		-20	38			21		29	-34		-24		27
mptoms		24	-40			-22		-30	30		30	-21	-27

orrelations ≥ .20, p<.001, are reported; decimal points omitted
al coping resources - N=310

from previous research studies can never capture all of the specific elements of the job of the respondents in any given study. Specific areas indicated as missing were attitudes toward management, training, and promotions. Based on interviews and informal discussions with the ATCs, the ATC HCS team developed items to measure satisfaction with management (7 items), satisfaction with promotion policy (5 items) and satisfaction with training (3 items). These scales were administered at the second through fifth examinations. Internal consistency reliability and test-retest reliability of these three scales proved to be adequate and are displayed in Table 22.

Descriptive Results

Since norm groups for comparison samples do not exist for these attitude variables, a threshold criterion of 20%-25% dissatisfaction was used to identify a problem area. This "rule-of-thumb" is well accepted in the industrial psychological literature; the complete rationale is discussed in Section III B 5. The threshold criterion of 20%-25% dissatisfaction can be used to interpret the three job-specific attitude dimensions: Satisfaction with Management, Satisfaction with Training, and Satisfaction with Promotion Policy. However, since these scales were not based on previous literature, examination of item response distributions is appropriate. Tables 23-25 contain the percentage of agreement with the items which represent dissatisfaction in the three scales.

The consistent, negative results on the items concerned with management and management policy in Table 23 would indicate that this is a significant area of dissatisfaction for controllers. This finding is consistent with the results found on the Group Morale scale in Section III B 5 below. This dissatisfaction is perhaps revealed most succinctly in responses to the following single item: "The greatest problems I face in doing my job well are due to FAA policies and not the work itself." Responses to this item indicated that 61% of the ATCs agreed.

As the final two scales, Satisfaction with Training and Satisfaction with Promotion Policy, also represented aspects of management policy, the negative attitudes indicated by the large percentages in Tables 24 and 25 were consistent. It was most interesting to note that the one item showing a relatively lower negative percentage (28%), "One loses the respect of peers when promoted," did not involve management policy, and, in fact, was closer to what has been defined below in Section III 5 as Individual Satisfaction.

However, it was clear from the results that a substantial number of ATCs were quite dissatisfied with management, training and promotion policies. These results were criticized by the FAA after an

TABLE 22

Internal Consistency And Test-Retest Reliabilities of Satisfaction With FAA Policies Scales

Scale	Internal Consistency Reliability Coefficient*	Test-Retest Reliability Coefficient**
Satisfaction with Management	.65	.70
Satisfaction with Training	.53	.64
Satisfaction with Promotion	.72	.67

*Based on Round 2 data, N = 392

**Based on Round 2 - Round 3 data, N = 352, with an average test-retest period of nine months.

TABLE 23

Responses to Satisfaction with Management Scale

7 Items, Round 2
(N = 332)

ITEM	CUMULATIVE FREQUENCY OF AGREE (COMPLETELY, MODERATELY OR SOMEWHAT)
FAA does not reward consistently good performance.	85%
Facility reluctance to use OT results in dangerous understaffing.	68%
Biggest problems are FAA policies, not the job itself.	61%
FAA more concerned with hardware than helping people do their job.	60%
Controllers' welfare is too low on the facility chief's priorities.	48%
Washington, DC policies hinder facility chief's ability to help ATCs.	44%
Facility reluctance to use OT seriously disrupts personal life.	29%

TABLE 24

Responses to Satisfaction with Training Policy Scale

3 Items, Round 2
(N = 392)

	CUMULATIVE FREQUENCY OF AGREE (COMPLETELY, MODERATELY OR SOMEWHAT
too easy now to become fully qualified.	75%
Quality of trainees is too low.	74%
ining should not be done with live traffic.	56%

TABLE 25

Responses to Satisfaction with Promotion Policy Scale

5 Items, Round 2
(N = 392)

	CUMULATIVE FREQUENCY OF AGREE (COMPLETELY, MODERATELY OR SOMEWHAT)
formance counts very little for promotions.	91%
ward for controlling well is recognition om other ATCs.	73%
ent only responds to the quality of work en something goes wrong.	70%
on to supervisor means becoming a "yes-man".	52%
es the respect of peers when promoted.	28%

interim report because of potential response bias. Most of the items, it was argued, were phrased negatively, inducing a high negative response.

As a result of this criticism, it was decided to check this potential bias during Round 5 data collection. All items from Round 2 were retained verbatim within the same questionnaire. A new questionnaire was created primarily to include measures identified in the Caplan et al. (1975) study for comparison with their results (see a discussion of this questionnaire in Section III A). Our previous questions on attitudes toward management, training and promotion were reworded so that agreement would indicate satisfaction with management policies, and the reworded questions were embedded in this questionnaire. For example, the original item, "The FAA is more concerned with improving hardware than with helping people do their job," was reworded, "The FAA is more concerned with helping people do their job than improving hardware." Finally, in order to minimize memory effects, the two questionnaires were administered at separate sittings, one in the morning and one in the afternoon, and the order of presentation was counterbalanced.

The results from Round 5, in terms of percent dissatisfied for each item by attitude area, are presented in Tables 26-28. These results did not support the contention that negatively-stated items biased the responses toward a high degree of dissatisfaction. There were no significant decreases in dissatisfaction with management (Table 26) when the items were stated positively. In fact, as indicated in the table, two items showed significant increases in dissatisfaction when the meaning was reversed.

The situation is the same when Satisfaction with Training results are examined (Table 27). Although two of the three items showed slight decreases, the percent of dissatisfaction was still so high as to make the small changes negligible for practical purposes.

The Satisfaction with Promotion Policy results were somewhat different in that four of the five showed changes in the direction of greater satisfaction. However, the levels of dissatisfaction on the items reversed in wording were still seriously high, and should be of concern to management. The original and reversed wording for the three items showing significant changes were:

(1) Job performance counts very little for promotion (original).

I believe excellent performance as a controller is an important consideration when promotions are made (reversal).

(2) One of the problems of being promoted to supervisor is that you become a "yes-man" for management (original).

TABLE 26

Responses to Satisfaction with Management Scales
(Round 5, N = 387)

	PERCENT DISSATISFIED	
	ORIGINAL WORDING	REVERSE WORDING
ore concerned with hardware than helping people do their job.	49%	62%*
st problems are FAA policies, not the job itself	58%	59%
lity reluctance to use OT seriously disrupts personal life	24%	59%
lity reluctance to use OT results in dangerous understaffing.	69%	65%
does not reward consistently good performance.	81%	83%
ington, DC policies hinder facility chief's ability to help ATCs	36%	62%*
rollers' welfare is too low on the facility chief's priorities.	43%	48%

ignificant increase in dissatisfaction.

TABLE 27

Responses to Satisfaction with Training Policy Scale

(Round 5, N = 387)

	PERCENT DISSATISFIED	
	ORIGINAL WORDING	REVERSE WORDING
should not be done with live traffic.	52%	57%
lity of trainees is too low.	86%	77%
o easy now to become fully qualified.	71%	65%

TABLE 28

Responses to Satisfaction with Promotion Scales

(Round 5, N = 387)

	Percent Dissatisfied	
	Original Wording	Reverse Wording
performance counts very little promotion	79%	40%[1,2]
otion to supervisor means becoming es-man"	56%	38%[2]
loses the respect of peers promoted	39%	34%
reward for controlling well is nition from other ATCs	76%	62%[2]
ement only responds to the quality rk when something goes wrong	67%	71%

content inadvertently changed - see discussion in text.

icant change

An ATC's attitude toward controllers becomes much worse when he is promoted to supervisor (reversal).

(3) About the only reward for controlling well is the recognition other controllers give you (original).

Exceptional controller performance results in recognition from supervisors that helps in getting promoted (reversal).

Examination of these items revealed that the reversed items were not exactly the opposite of the original items. In accordance with the belief that respondents do not care to answer questions that are very similarly worded and that appear to be repetitive, we attempted to change the wording so the questions would not be recognized from the previous sitting. Although the other job-specific items were easy to change and reverse, it was particularly difficult to change and reverse these three items. Thus, increase in satisfaction may well have been more due to semantic subtleties in the items than to the reversal itself. However, even though the 79% of the ATCs who believed that "job performance counts very little for promotion" decreased to 40% who disagreed with the statement that "excellent performance as a controller is an important consideration when promotions are made," that proportion of dissatisfied ATCs was considerably greater than the 20%-25% criterion, indicating a problem in the area of promotions policy.

A comparison of Round 2 data with Round 5 data in terms of percent dissatisfaction revealed a high degree of stability over time. Of course, different men may have been dissatisfied at Round 2 than at Round 5, but of the 15 items, only 4 showed shifts greater than 10%. These results may indicate that the ATC's attitudes in these job dimensions are not easily changed. It might also indicate that many ATCs accept these negative feelings about management, training and promotion much as other workers accept high noise levels or poor illumination as "part of the job." If this is true, one would expect these attitude responses to have little ability to predict health change. Results of analyses of their predictability are noted in Section V.

In conclusion, reversing the item wording had little impact on the extent of satisfaction revealed. The only significant changes appeared to be more a consequence of changed item meanings than response bias. The responses of our study group at Round 5 indicated that ATCs were considerably dissatisfied with management policy, training and promotion policy, the same conclusion reached from Round 2 data.

Peer Nominations for Competence

In our discussions with the FAA administrators and air traffic controllers, we often were presented with statements that competent controllers might be affected less by the pressures of their job. Other less competent men, it was suggested, struggled more and were strained more by their job. These suggestions were based on many years of involvement with the air traffic system and therefore were considered with great care.

Our visits to air traffic facilities and continued observation suggested that these ideas had considerable merit. It was not difficult to find anecdotal evidence that supported the idea that competence could moderate the effects of workload and possibly influence the effects of air traffic control work on health. Other observational evidence also suggested that competence had a more direct effect on health. Thus, we considered it relevant to find the best method of assessing the competence of the individual ATCs.

Alternative Techniques for Assessing ATC Competence

A number of techniques were available for assessing competence. These included (1) supervisor ratings from annual performance evaluations, (2) semi-annual task performance profile ratings, (3) expert-observer ratings of all ATCs, (4) peer ratings of all ATCs, (5) self-ratings, (6) peer group nominations of the most competent. All of the techniques were feasible, but each had drawbacks as well.

Supervisor ratings are a means of evaluating job performance and merit in the FAA. Although this established use of supervisor ratings would have given them a special relevance, their use also presented problems: neither the ATCs nor the supervisors felt the annual performance evaluations were an appropriate measure of competence. Their comments indicated that these evaluations had changed from a performance to a compensation orientation as is common in most companies. Consequently they felt that these evaluations would yield skewed distributions with nearly everyone being rated at the competent end in order to insure annual step-in-grade raises. Thus, supervisor ratings were eliminated as useful measures of competence because of inadequate discrimination between individuals.

Semi-annual task performance profile ratings also are conducted within the air traffic control system. These ratings are made by supervisors using a rating and task definition system that is quite comprehensive (System Development Corporation, 1972). However, the developers of this system noted that it was not applicable for studies across different air traffic facility environments. Further, this system is an "official" system, and consequently, there were potentially serious political problems associated with its use for research. Therefore, we eliminated this method of competence rating from consideration.

Global expert-observer ratings could have been used. We could have trained ex-ATCs or supervisors, for example, to rate the ATCs in the study using scales such as those of Milne and Colmen (1972). However, a major problem with this method was that a standard series of work situations was not available and such a series would be necessary since the depth of observation would necessarily be limited for our particular study. For that reason, this method was not used.

Global peer ratings would have avoided the problem of lengthy, in-depth observations because men would know each other's performance based on years of observation in many different situations. On the other hand, we were informed that we could expect virtually no negative evaluations and a preponderance of very favorable evaluations if we used global peer ratings. Many controllers felt this kind of procedure would be a psychological popularity study. The method was dropped from further consideration.

Self-ratings were ruled out by the study team for a multitude of reasons. One overriding reason for exclusion was the lack of accuracy and validity typical of self-ratings (Lawler, 1967). A probable reason for this inaccuracy of self-ratings is that they are influenced by self-esteem. Other reasons included the lack of comparable standards between individuals, the extremely skewed distribution that could be expected, and the general lack of acceptability of such ratings by either the sponsors or participants in the research.

Peer group nominations (sociometric nominations), the remaining alternative, are a form of peer group rating. Individuals nominate or rank their peers in terms of varying characteristics, such as competence. A given person's "score" in this type of system is the average rank order or the number of nominations received from one's peers. Our subjects informed us that they thought that such a system would be fair, practical, and valid as long as every individual did not have to be rated. That is, our subjects thought it would be all right to say who was "good" but not who was "bad." They did not want to rank-order all relevant individuals (e.g., end up placing someone at the bottom of a list) nor did they want to assign a rating to all individuals (because then someone would end up with the lowest average).

To alleviate these feasibility problems, we arrived at a compromise solution. The Sociometric Peer Rating Questionnaire (SPRQ) was devised so that each subject nominated only three other subjects for each of three questions. A given person could be nominated for more than one question but only once for each question. A subject was provided with a list of all of the subjects with whom he worked. For men from small facilities, all volunteers from the facility would be on each person's list. For men from larger facilities, each person's list included only men from his own work team and "sister" teams (sister teams work the same shifts). Every man had more than three other persons' names on his list, and therefore, no one was put in the position of ranking or rating everyone. This procedure was acceptable to our subjects.

Since the SPRQ seemed to be the most feasible instrument and method for assessing ATC competence, we investigated previous research on sociometric ratings. Evidence in the scientific literature for the general validity, reliability, and relevance to health of peer ratings is discussed in Appendix 1, under Section III B 5: Background. Available literature, in summary, confirmed that peer nominations could be expected to have a strong relationship to actual job performance.

Application of the Peer Rating Method

A total of 416 men were processed through the first series of examinations for the ATC Health Change Study. Prior to the first man's visit in February 1974, lists of potential participants in the study were drawn up for use with the sociometric questionnaire. Originally these lists contained the 547 names of those eligible for the study on the basis of a stratified random sampling. By January, 1974, 458 men had either volunteered to participate or were undecided.

The lists used for the sociometric nominations were designed to include only active journeyman controllers who had indicated that they were going to participate. Of the 458 men, 11 were undecided, and hence were excluded from the lists (these 11 eventually did volunteer). Another three were on specialized non-controlling assignments, and since our lists included only men working on functioning controller teams, these three were excluded from the lists. Consequently, the list for the first series of examinations included 444 men.

By the end of the first series of examinations in September, 1974, 435 men had signed informed consent forms certifying their decision to participate in the study. Twenty-three of the 458 had changed their decision to participate and/or finally decided not to participate. In addition, 19 men became ineligible to participate because they had been promoted, transferred or medically disqualified within the FAA. Consequently we saw only 416 men in the first series. Of the 416 men, two refused to complete the sociometric questionnaire. Hence, 414 men made the nominations on the first round. During subsequent administrations of the SPRQ, there were fewer subjects because additional individuals became ineligible for the study.

On the SPRQ questionnaire a man was presented with three questions. First, if all assignments were changed to correspond to his preferences, what three ATCs would he most like to work with (ideal team nominations). Second, without considering technical ability, with what three ATCs would he find it easiest to work (amicability nominations)? Third, without considering how easy it was for him to work with someone, who did he believe were the best controllers from a technical standpoint (competence nominations)?

We provided each respondent with a list of names of men in the study. Each man's list included his own team members and those of the two sister crews with which his team regularly worked. In this way we limited the choices to persons with whom they would be familiar and who were in the study.

Each question was presented separately by the computer along with the appropriate list of names. The ATC typed in the identification numbers of the three persons he chose to meet the criteria of the questions, and further, the subject indicated if each of his selections was on his own team. After receiving the answers to one question, the computer displayed the next question. Checking routines were built into questionnaire programs to preclude invalid answers (such as trying to list the same person three times for one question).

The nomination lists were updated at the beginning of each series of examinations (rounds) but were not modified once a series started. This procedure was followed to ensure that all names were presented to the same (and appropriate) people for the same amount of time. If a subject changed teams during a series, the list on which his name appeared was not modified. Rather, that subject would be given a new list of names from which to make his choices, but his own name would stay on the old list until the next series of examinations. The names of subjects who had dropped from the study between rounds were retained on the list until the next round, so that they could be nominated after leaving the study, but could not participate in nominating others.

Interrelationship of Choices

Before accepting the Sociometric Peer Rating Scale and the procedures we had devised as reliable and valid indicators of the variable we wished to measure, i.e., competence of the individual ATC, we carried out a number of tests to determine whether or not the probability of being chosen randomly might affect the frequency of any man being chosen and whether or not the values or choices that were measured by the questionnaire were stable over time. Details of these tests and their results are included in Appendix I, Section IIIB5: Background. The tests that we made persuaded us that no adjustment of the raw scores was necessary. Further, we were reassured by the test-retest reliability coefficients generated over two successive intervals that the peer ratings were reasonably stable. We were primarily interested in obtaining a measure of competence. We placed the question of competence after two questions of a more innocuous nature (who would you put on your ideal team, and who is easy to work with?). The intent of this procedure was to reduce popularity or halo effects on the nominations for competence.

The interrelationships between the three types of choices are presented in Table 29. All of the correlations in the table are highly significant ($p < .001$). A most important finding, however, was that amicability was distinguished from competence (sharing 36% of its variance) and was most highly related to whom an ATC would put on his ideal team. Thus the competence ratings were not simply "nice guys" ratings; however, clearly a number of the ATCs nominated as competent were also "nice guys."

Relationship to Other Psychological Characteristics

Pearson product moment correlations between the three main scores of the SPRQ and other psychological scales were generally very small in magnitude (the highest was .15). This result was not unusual for sociometric ratings (Korman, 1968), but it could have been due to several factors such as non-normal distributions or curvilinear relationships. Consequently we conducted one-way analyses of variance using ordinal categories of SPRQ scores as the independent variables and other psychological scales as the dependent variables.

Table 30 shows the significant analysis of variance results for ideal team nomination categories; Table 31 shows the significant results across the amicability nomination categories; and Table 32 gives the same results for the competence nomination categories. Although more than 40 psychological characteristics were examined for each type of nomination, few unique characteristics were associated with how often one was chosen by one's peers.

The results shown in Table 30 indicate that those men who reported fewer anxiety symptoms and tension discharge mechanisms were nominated to the ideal team significantly more often than

TABLE 29

Correlations Between
Types of Sociometric Nominations
(N = 401)

ination	Ideal	Amicability	Competence
al Team	1.00		
cability	.86	1.00	
etence	.80	.62	1.00

TABLE 30

Analysis of Variance Results
for Ideal Team Nominations

ychological riable		Number of Times Nominated for the Ideal Team					
		0	1	2-3	4-6	7+	
nxiety before ork	Mean	50.5	51.3	48.7	51.9	46.9	$F = 2.70$
	S.D.	9.9	10.9	9.5	11.1	6.5	$p < .03$
	N.	64	72	127	88	42	
uring work nxiety	Mean	51.5	51.8	48.4	51.0	47.6	$F = 2.64$
	S.D.	10.8	10.3	9.4	10.1	8.4	$p < .04$
	N	64	72	127	88	42	
ter work xiety	Mean	53.0	50.9	48.9	49.5	48.3	$F = 2.28$
	S.D.	10.8	10.0	10.0	9.5	9.4	$p < .06$
	N	64	72	127	88	42	
nsion Discharge chanisms on Bad ys	Mean	53.2	51.3	49.4	48.9	48.4	$F = 2.60$
	S.D.	10.9	10.9	10.4	9.0	7.2	$p < .04$
	N.	64	72	127	88	42	

TABLE 31

Analysis of Variance Results for
Amicability (Ease of Working With) Nominations

Psychological Variable		Number of Times Nominated as Easy to Work With: 0	1	2-3	4-6	7+	
Subjective Costs (Visit 1)	Mean	52.1	52.6	49.7	47.9	51.0	$F = 2.94$
	S.D.	10.5	8.6	10.4	10.6	7.9	$p < .02$
	N	56	77	132	87	41	
Social Coping Resources (Visit 1)	Mean	47.7	49.7	49.3	52.5	49.3	$F = 2.22$
	S.D.	10.0	10.9	9.7	9.6	9.4	$p \le .06$
	N.	56	77	132	87	41	
CPI Dominance (Visit 2)	Mean	58.5	56.0	57.6	55.9	53.6	$F = 2.39$
	S.D.	12.2	12.3	11.1	12.0	12.4	$p < .05$
	N	80	79	95	92	44	
CPI Self-Acceptance (Visit 2)	Mean	61.6	58.8	58.7	57.2	57.0	$F = 2.86$
	S.D.	9.2	9.7	9.7	8.8	9.0	$p < .03$
	N	80	79	95	92	44	

TABLE 32

Analysis of Variance Results for Competence Nominations

Psychological Variable		Number of Times Nominated as Most Competent: 0	1	2-3	4-6	7+	
During Work Anxiety	Mean	51.8	50.8	49.7	50.1	46.6	$F = 2.47$
	S.D.	10.9	9.5	10.5	9.8	7.4	$p < .04$
	N	84	75	100	80	54	
Tension Discharge Mechanisms on Bad Days	Mean	53.7	50.0	48.4	50.3	47.9	$F = 4.11$
	S.D.	10.5	10.0	10.0	10.1	8.0	$p < .003$
	N	84	75	100	80	54	
Outside Activities to Cope	Mean	52.4	49.2	50.5	49.6	47.3	$F = 2.46$
	S.D.	10.7	10.3	8.5	10.1	9.9	$p < .05$
	N	84	75	100	80	54	
Toughmindedness	Mean	47.6	50.2	51.9	48.8	51.8	$F = 2.85$
	S.D.	10.3	11.2	10.1	9.7	7.6	$p < .02$
Investment	Mean	47.0	49.2	51.6	50.6	51.0	$F = 2.84$
	S.D.	9.7	10.3	9.9	9.6	10.4	$p < .02$
	N	84	75	100	80	54	

those men who reported a higher number of anxiety symptoms. Although there was a high correlation (.86) between nominations for an ideal team and for being easy to work with (amicability), a different set of self-reported psychological characteristics differentiated the men chosen as most amicable (see Table 31). As shown in Table 31, those who were nominated as "easy to work with" experienced fewer subjective costs and tended to have more social resources than those who were never nominated for this attribute.

Perhaps the most interesting result was the finding that the amicability nominations were related to the traits of dominance and self-acceptance as measured by the California Psychological Inventory. The people who were never nominated as being easy to work with scored the highest on dominance and self-acceptance while those who were nominated most often for this characteristic scored significantly lower on these two traits. It appears that more dominant, self-accepting controllers are not particularly easy to work with.

Finally, the results for competence nominations are presented in Table 32. Those men who were nominated most often as competent reported themselves to be significantly less anxious at work, using fewer tension discharge mechanisms, engaging in fewer community and union activities, and having greater toughmindedness and more job investment than those who were nominated only once or never.

SUMMARY

General Job-Related Variables

Three sets of scales were used to assess the ATCs' general attitudes toward their jobs and work: the Job Description Inventory (JDI), the Kavanagh Life Attitude Profile (KLAP) and the Leader Behavior Description Questionnaire - Form XII (LBDQ).

Of five scales of the JDI, the three used were: Satisfaction with Pay, Satisfaction with Co-workers and Satisfaction with Work itself. The KLPA's two scales, Group Morale and Individual Satisfaction, were used as measures of overall satisfaction. Certain supervisory behaviors, as seen by the ATCs, were reported and measured in terms of frequency on the LBDQ. The supervisory dimensions measured were: Initiation of Structure (IS), Tolerance of Freedom (TF) and Consideration (C). Psychometric properties of internal consistency, test-retest reliability and independence of the scales were tested and found to be adequate.

Attitudes of the ATCs toward the work itself were found to be average when compared to a normative sample of other workers stratified by socio-economic characteristics; ATC attitudes toward co-workers were low and toward pay, above average when compared with the same normative sample. In comparison with workers at comparable salaries, ATC attitudes towards work, co-workers and pay were quite low.

The three supervisory behavior qualities were noted less frequently among ATC supervisors than among comparison supervisors.

The KLAP revealed that 42% of the ATCs felt "so-so" in general about their jobs. Differences in responses between the Group Morale and Individual Satisfaction scales supported a conclusion that the major sources of dissatisfaction were note in the ATCs' work itself, but rather in the work environment.

b. General Job-Related Variables

In addition to our decision to explore psychological and attitudinal dimensions particular to the air traffic controlling occupation because of their potential impact on health outcomes of the men, a reading of the industrial psychology literature persuaded us that we should also investigate more general work and job satisfactions as possible modifiers, if not generators, of health change.

A number of studies suggested that satisfaction or dissatisfaction with work were related to health (Antonousky, 1968; Caplan, 1971, Kasl and Coff, 1970; French and Caplan, 1970; Friedman, Rosenman and Carroll, 1957; Groen and Drory, 1967; House, 1972; Jenkins, 1971; Kornhauser, 1965; Maaora and Taustein, 1969; Quinn, Staines and McCullough, 1974; Russek, 1959; Sales, 1969; Sales and House 1971; Wolf, 1971; Weintraub, 1974; Palmore, 1969; Benet, 1972). However, as none of the previous studies were longitudinal in nature, it was not possible to determine what associative, modifying or possibly causal role was played by the work attitude factor, nor were the findings consistent in relating work satisfaction to types of illness.

The current study provided an opportunity to determine whether or not general job satisfactions were related to changes in health, and whether deterioration in attitudes toward work anteceded or were concurrent with health changes.

While it was necessary to devise an instrument to assess dimensions particular to the air traffic control occupation (preceding section), the more general attitudes toward work and jobs (regardless of the specific nature of the work) could be assessed by existing industrial psychological scales which provided, at the same time, normative scores from other populations so that we could compare job attitudes of ATCs with those of males in a diverse group of other industries.

We selected three sets of scales to assess general job attitudes or perceptions. The Job Description Inventory (Smith, Kendall and Hulin, 1969), the Kavanagh Life Attitude Profile (Kavanagh, MacKinney and Wolins, 1971), and the Leader Behavior Description Questionnaire - Form XII (Stogdill, 1963).

The Job Description Inventory (JDI)

Because of its extensive use, and availability of norm data, the JDI was a useful instrument for comparisons of job satisfactions between the ATC group and other populations. In addition, it had strong psychometric properties to recommend it (Smith et al., 1969).

Of the five scales contained in the JDI - satisfaction with supervision, with promotions, with pay, with co-workers and with work itself - we used only the last three named. The satisfaction with supervisors scale was omitted because another more comprehensive measure of supervisory behavior was used. The satisfaction with promotions scale was omitted because this dimension was included in the ATC Questionnaire and was directed specifically to FAA promotion policy. The scale relating to pay was shortened from nine to five items to reduce overlap with another scale and to eliminate items inappropriate to the ATCs.

The Kavanagh Life Attitude Profile (KLAP)

In addition to the measurement of satisfaction with specific aspects of the job, such as pay and co-workers, we wanted to have a measure of the ATCs' overall satisfaction with the job. Such a global measure of satisfaction, it was thought, might well relate differently to health change than would the measures of satisfaction with specific aspects of the job.

The KLAP scale (Kavanagh, et al., 1971) was chosen as the global measure of job satisfaction. The instrument contains two overall satisfaction scales -- group morale and individual satisfaction with 10 and 8 items, respectively. These scales differ from other job satisfaction scales in that the items, individually, measure various aspects of the job, but are summed to derive indices of overall individual satisfaction and group morale.

The KLAP scales differ from other job satisfaction scales also with respect to reference. Most scale development is based on factor analysis of responses to job attitude items, with the resulting scales, or factors, comprised of items similar in their reference to a given aspect of the job. For example, the JDI Work scale contains items whose common reference is the work the person is doing. The two KLAP scales were based on a different developmental analysis, one that clustered agreement among people (not items) to job satisfaction items.

Items showing the highest agreement between employees of the same work group were identified to comprise one scale, while those items showing the lowest (from a larger set of items) agreement were combined into another scale. Examination of the reference of these two sets of items revealed that those with high agreement referred to job factors that employees held in common, while those for which there was low agreement were distinctly individualistic in reference. The latter was named Individual Satisfaction, while the former was named Group Morale.

One advantage of these scales versus single items is that the respondent is not asked to sum the various job aspects "in his head," and respond with a summary answer. In addition, this scale distinguishes between two aspects of overall satisfaction, individual feelings and feelings shared with other employees (group morale). Group morale is the overall attitude a person shares with others in his work group or organization, and it typically involves factors external to the person such as the company administration, or the supervision he receives. Group morale is expected to be fairly stable over time. On the other hand, individual satisfaction as an overall measure is more personal and typically deals with factors internal to the person, such as whether or not he finds his work boring. Individual satisfaction is expected to be more variable over time.

Strong psychometric properties reported by Kavanagh, et al. (1971) added to this instrument's utility as a measure of overall job satisfaction.

Leader Behavior Description Questionnaire (LBDQ)

A measure of the behavior of the supervisor as seen by the ...s included in the ATC HCS for several reasons. The effect ...cularly of the immediate supervisor's style or behavior on ...loyee's performance, job satisfaction and role conflicts ...wn to be important (e.g., see Campbell, Dunette, Lawler ...ick, 1970; Korman, 1966; Stogdill, 1974; Kahn et al., 1964; ...and Filley, 1968; Vroom, 1964). In addition to the import... supervisory behavior generally, it has been shown to ...mportance specifically in the ATC occupation (Singer and ...anz, 1971; Smith, 1973). An unpublished report by CAMI ...ttitudes of ATCs, for example, states that "the dominant ... issue for all three ATC specialties (ARTCC, TRACON, and ... that of management. In this they share attitudes in ...ith European controllers and with employees in most other ...ons." Finally, we hypothesized that there would be a ...hip between supervisory behavior and job stress, and ...ended to investigate possible relationships between ...ors and health changes, we determined to include a ...hat would allow us to assess supervisory influence in ...environment of the ATC.

...r than measure attitudes (satisfaction) toward super... decided to measure supervisory behavior because this ... would be more useful. It could tell us what super...viors were in need of improvement, so that recom... for intervention might be made as appropriate; and ... is more easily changed than attitudes, it would prode a more efficient focus for intervention.

Although a variety of scales were available to measure ervisory behavior (see Stogdill, 1974), three major ones manded consideration, the Supervisory Behavior Description stionnaire (SBDQ), (Fleishman, 1972), The Leader Opinion stionnaire (LOQ) (Fleishman, 1972) and the Leader Behavior cription Questionnaire (LBDQ) (Stogdill, 1963).

The LBDQ was chosen because the other scales had limited in the ATC HCS. The LOQ is a self-description by the rvisor as to how he <u>should</u> behave. This seemed inappro- te for this study as we were interested in the ATCs' eptions of their <u>existing</u> work environment. The SBDQ the LBDQ both elicit employees' descriptions of their rvisors' behavior. However, the SBDQ measures only two nsions of leader behavior, is considerably longer than LBDQ, and the construct validity of one of its scales has under sharp criticism (Schriesheim and Stogdill, 1964; iesheim, House, and Kerr, 1976; Szilagyi and Keller, 1976; enberg and Kavanagh, 1972). On the other hand, the ometric evidence for the development and use of the is reasonably good (Stogdill, 1963).

The LBDQ was developed from earlier work describing ehavioral dimensions of the supervisor's role (Stogdill oons, 1957). However, these early efforts resulted in dentification of only two dimensions of supervisory ior. It seemed unreasonable to assume that two dimen- would be sufficient to describe completely the com- y of supervisory behavior. Thus, the LBDQ-Form XII red to as the LBDQ in this report) was developed to additional dimensions of leader behavior, and the form contains 100 items to measure 12 dimensions. In to keep the questionnaire to a manageable size for dents, the ATC HCS team limited the number of dimensions the study. Based on observations of the ATCs in job ction with their supervisors as well as on interviews rious ATCs, the following supervisor behavior dimen- ere chosen: 1) Initiation of Structure (<u>IS</u> - clearly own role, and lets followers know what is expected; erance of Freedom (TC) - allows followers scope for ive, decision, and action; and 3) Consideration (C) ds the comfort, well-being, status, and contribution owers.

is important to recognize that the LBDQ measures <u>cy</u> of supervisory behaviors, not <u>satisfaction</u> with sory behavior. The instructions for the LBDQ items clearly indicate that the ATCs were to report their perceptions of the frequency of their supervisors' behavior on the items.

Technical Considerations

Certain measurement aspects of the scales described in this section should be set forth, including specific modifications made to improve the quality of the scales, and psychometric evidence supporting the use of the various measures. As the psychometric evidence regarding the development of the scales has been referenced, this section deals with the following topics: 1) number of scale points, 2) reduction in the length of the scales, 3) reliability of the scales, and 4) independence of the variables.

umber of Scale Points

All scales were converted to 7 point response formats. This as done for several reasons. First, numerous studies have found at reliability of scales is a monotonically increasing function the number of scale points (Guilford, 1954). As Nunnally 967) has pointed out, the increase in reliability, starting th two scale points, is very rapid at first, and then tends to vel off at about seven. Secondly, from a practical standpoint, ven-point scales were used to allow the ATCs more latitude in sponding to individual items. Finally, seven-point scales, contrasted with shorter scales, could lead to increased item scale variance, thus potentially increasing the value of a en variable for predicting health change.

The original JDI scales were formulated as three-point scales, LBDQ as five-point scales, and the KLAP as 99-point scales. as important to determine if the seven-point format for the and LBDQ were equivalent to the original response format, since scales have norms which were based on their particuler res- format. Since the KLAP had no norms, this was not neces- Item response transformations were developed so that the nd the LBDQ could be scored using both the seven-point system he original response format. The re-scored scales were lated with the standard seven-point scales, with the follow- esults: Work Satisfaction: .77; Coworker Satisfaction: .74; tisfaction: .87; Initiation of Structure: .90; Toler- f Freedom: .93; and Consideration: .90. For practical es, the magnitude of these values indicates that the scores may be considered virtually equivalent. Much of fference between the observed correlation and the theore- ideal may be a function of the reduced range of the orig- asurement rather than a failure of the new format to the same meaning as the original.

Scale Reduction

Since the number of items contained in the first data collection phase was large, we thought that reducing the number of items within each scale would be desirable. However, the basic meaning of each scale had to be maintained. In order to accomplish this, item-total scale score correlations were calculated to determine which items were most highly related to the total scale score. The items in each scale were also regressed on the scale score using a step-wise procedure. This procedure adds items one by one, in descending order of correlation with total scale score, until the total variance accounted for in the scale score by the present equation (larger by one predictor item) is not significantly different from the variance accounted for by the previous equation of predictor items. This procedure maintains the meaning of the scale while decreasing the number of items to be used in later data collection phases.

The regression analysis was computed on the first 173 participants' responses in Round 1. For the JDI - Work Satisfaction, the regression solution indicated 10 items (original number of items was 18) could be used ($F = 212.9$, $p < .001$). These ten items accounted for 93% of total scale score variance. For the JDI - Coworker Satisfaction, the stepwise solution indicated 10 items (original number was 18) could be used ($F = 275.6$, $p < .001$). These items accounted for 94.4% of the total score variance. The Pay scale of the JDI was not analyzed because it only contained four items. Likewise, the two subscales of the KLAP were not analyzed because they contained eight and ten items respectively. In addition, the original development of these scales was empirically different from the typical scale development procedure, making item-total regression analysis inappropriate for reducing scale length. The original scale development (see Kavanagh et al., 1971) involved correlating people, not items, to identify group morale and individual satisfaction. Item-item or item-total score correlations to reduce scale length would thus violate the conceptual foundation of these scales. Therefore, these scales were retained in full for subsequent data collection rounds.

Regression analysis of the LBDQ revealed that the subscales could be reduced in length. The Initiation of Structure scale, originally 10 items, could be reduced to 5 items, which accounted for 91% of the total score variance ($F = 354.9$, $p < .001$). The Tolerance of Freedom scale, originally 10 items, could be reduced to 5 items that accounted for 96% of the total scale score variance ($F = 247.3$, $p < .001$). Finally, the regression solution indicated that the Consideration scale, originally 10 items, could be reduced to 5 items that accounted for 95% of the total scale score variance ($F = 596.77$, $p < .001$).

However, the exclusive use of items from the regression analyses could have lowered the internal consistency of the scales since the regression technique tends to choose items relatively independent of each other so as to maximize the total explained variance. To insure that the internal consistency would not be seriously affected, the highly correlated items identified in the first analysis but not included by the regression were included in the revised, shortened scales.

The result of this dual procedure was to reduce the total number of items in the combined scales from 89 to 54. The table below shows the item reduction by scale.

SCALE	ORIGINAL NUMBER OF ITEMS	REVISED NUMBER OF ITEMS
JDI		
Work Satisfaction	19	8
Coworker Satisfaction	18	7
Pay Satisfaction	4	4
KLAP		
Individual Satisfaction	8	8
Group Morale	10	10
LBDQ		
Initiation of Structure	10	6
Consideration	10	6
Tolerance of Freedom	10	5

These are the scales that were used in subsequent data collection phases. The modified form of these scales is included in the Exhibits appended.

Reliability of the Scales

Internal Consistency. Internal consistency reliabilities were calculated using coeffcient alpha (Nunnally, 1967, p. 193-196) for the eight shortened scales. As can be seen from Table 33, all scales met acceptable standards of reliability (Nunnally, 1967, p. 226). The reliabilities also compared favorably with previous research. Smith et al. (1969) found internal consistency reliabilities of .73, .78, and .67 for the Work, Coworker and Pay scales, using a sample of 80 males from two electronic plants. Stogdill (1963) reported the following ranges of reliabilities for the LBDQ scales: 1) Initiation of Structure - .72 to .80; 2) Tolerance of Freedom - .58 to .86; and 3) Consideration

.76 to .87. As can be seen in Table 33, the reliabilities for the ATC HCS compared quite favorably with these values and supported the use of the scales. No previous internal consistency reliabilities were reported for the Individual Satisfaction and Group Morale scales since this was not appropriate in their development (Kavanagh et al., 1971). However, the values in Table 33 were quite acceptable. The lower value for the individual satisfaction scale as contrasted with the group morale scale was consistent with their conceptual basis. One would expect items with reference shared by team (or group) members to be more internally consistent than satisfaction items individualistic in nature.

Test-Retest Reliability. Not only was it important for scales to be internally consistent, but it was also important that the scores for the ATC study group were relatively consistent over time as the variables measured by the scales were thought to be relatively enduring attitudes. Table 34 presents the test-retest reliabilities for the scales for three rounds. The test-retest reliabilities met generally accepted standards for reliability of measurement over time (Nunnally, 1967). The relatively lower values for the LBDQ were most likely a result of changes in supervisors between data collection points, rather than weakness of the measurement. The rather remarkable result in Table 34 was that the reliability over eighteen months did not show substantial deterioration from the nine-months test-retest reliabilities.

Independence of Variables

The independence of the industrial psychological variables was tested by inter-scale correlations. If two variables were highly correlated, it would make little sense to include both in a study as a score on one would allow prediction of the score on the other. Table 35 presents the results of the inter-scale correlation for the Round 1 data. Most of the values in Table 35 were low enough to support the conclusion that the variables were indeed independent measures. However, the correlations above .60 indicated some overlap between the measures. No values were high enough to warrant dropping variables from the study.

Inter-scale correlations for Round 2 between the job-specific attitude scales discussed in the previous section and the general job-related scales discussed in this section are presented in Table 36. Again, these values did not warrant the dropping of any scales. The significant correlations between the five general satisfaction scales and the job-specific scales provided evidence that the latter scales were indeed measuring the ATCs' affective (attitudinal) responses to their

TABLE 33

Internal Consistency Reliabilities
For Industrial Psychological Scales

NSTRUMENT	SCALE CONSTRUCT	RELIABILITY COEFFICENTS
DI*	Work Satisfaction	.81
	Coworker Satisfaction	.73
	Pay Satisfaction	.69
LA?*	Individual Satisfaction	.62
	Group Morale	.82
DDQ*	Initiation of Structure	.72
	Consideration	.75
	Tolerance of Freedom	.89

Based on Round 1, N = 410

TABLE 34

Test-Retest Reliabilities For Industrial Psychological Scales

SCALE CONSTRUCT	ROUND 1-ROUND 2 RELIABILITY* (N = 386)	ROUND 2-ROUND 3 RELIABILITY* (N = 352)	ROUND 1-ROUND 3 RELIABILITY** (N = 347)
JDI			
Work Satisfaction	.71	.67	.60
Coworker Satisfaction	.62	.67	.51
Pay Satisfaction	.65	.69	.63
KLAP			
Individual Satisfaction	.63	.66	.63
Group Morale	.66	.65	.59
LBDQ			
Initiation of Structure	.58	.54	.43
Consideration	.58	.58	.48
Tolerance of Freedom	.56	.50	.42

* An average test-retest period of nine months.

**An average test-retest period of eighteen months.

TABLE 35

Correlation Matrix For The Industrial Psychology Scales - Round 1

Work Satisfaction	1.00						
Coworker Satisfaction	0.34	1.00					
Pay Satisfaction	0.01	0.19	1.00				
Individual Satisfaction	0.19	0.33	0.66	1.00			
Group Morale	0.16	0.31	0.30	0.53	1.00		
Initiation of Structure	0.11	0.13	0.03	0.20	0.52	1.00	
Consideration of Others	0.11	0.18	0.09	0.30	0.55	0.63	1.00
Tolerance of Freedom	0.11	0.18	0.06	0.22	0.39	0.32	0.61

TE: These correlations are based on an N of 375, since only ATCs with scores on all variables were included in this analysis.

TABLE 36

Correlations Between The Job-Specific Attitude Scales And The General Job-Related Scales*

	Job-Specific Attitudes		
eral Job-Related Scales	Satisfaction with Management	Satisfaction with Training	Satisfaction with Promotion
rk Satisfaction	.11	-.13	.16
orker Satisfaction	.25	-.30	.22
y Satisfaction	.38	-.24	.21
dividual Satisfaction	.57	-.35	.43
oup Morale	.57	-.28	.47
itiation of Structure	.22	-.05	.18
nsideration	.34	-.06	.28
lerance of Freedom	.21	-.01	.15

r Round 2, N = 392

jobs.

The inter-scale correlations for the JDI scales compared favorably with the original work done by Smith et al. (1969). They reported the following values based on 981 males pooled across 21 plants: (1) Work-coworker .36; (2) Work-Pay, .40; and (3) Coworker-Pay, .28. These values for the ATC HCS were (1) .34; (2) .01; (3) .19. Although the ATC results appear better in terms of being more independent measures in two of the three cases, it must be remembered that the ATC results were based on a single occupational group, whereas the Smith et al. (1969) results were based on many different occupations in different organizations. No comparison interscale correlations were available for the other industrial psychological variables; however, the values were generally low with a few exceptions. Interpretation of other results must take these correlations into consideration.

Normative Comparisons

JDI. We compared the ATC scores to norms in order to determine how satisfied the ATCs were with aspects of their job compared with other workers. Table 37 presents the results of the normative comparisons for the ATC scores on the JDI scales. In order to understand these results, it is necessary to describe the normative sample for the JDI (Smith et al., 1969). The male norms were based on 2,019 males pooled across a total of 21 plants, representing 19 different companies and 16 different Standard Metropolitan Statistical Areas. According to Smith et al. (1969, p. 88), "the basic sample was drawn from individual units (with 50 or more employees) of business and industrial firms in the continental United States (excluding Alaska) as of 1956," and they claimed that "the 21 plants and 16 communities are reasonably representative of the range of conditions found in American industry and business." The sampling of workers in terms of job levels was as follows: 22% unskilled, 30% semiskilled, 18% skilled, 13% white collar, and 17% supervisory (almost entirely at the first-line level). Selected company and community characteristics are displayed in Tables 38 and 39.

The importance of understanding the normative sample characteristics can be seen by examining the percentile rankings in Table 37. Compared against the total array of normative samples, the ATC means converted to percentiles for the work, coworker and pay satisfaction scales were 46 percentile, 24 percentile and 60 percentile, respectively. The attitudes of the ATCs sample toward the work itself were therefore about average com-

TABLE 37

Norm Comparisons for the Job Description Inventory

SCALE	ATCS	MALES IN 21 INDUSTRIES	ATC %ILE OVERALL	ATC %ILE BY INCOME GROUP	ATC %ILE BY EDUCATION GROUP	ATC %ILE BY JOB TENURE GROUP	ATC %ILE BY COMMUNITY PROSPERITY	ATC %ILE BY COMMUNITY DECREPITUDE
Satisfaction with Work	36.20	36.57	46	8	30	46	43	50
	6.56	10.59						
	408	1977						
Satisfaction with Coworkers	39.64	43.49	24	15	31	27	32	34
	6.90	10.02						
	410	1928						
Satisfaction with Pay	33.29	29.90	60	17	52	53	68	64
	12.93	14.53						
	400	1966						
				Highest	13-14 yrs educ	10-15 yrs exp.	37-50 yrs	19-25 yrs

All cells read: Mean
Standard deviation
Number of subjects

TABLE 38

Characteristics Of Companies In Normative Sample, 1960

NT	LOCATION	PRODUCT OR SERVICE	NUMBER OF EMPLOYEES	AVERAGE ANNUAL EARNING	
				MALE	FEMALE
	Mich.	Brass fittings	475	5220	3810
	Md.	Research	75	6380	3220
	Mass.	Office files	360	3140	1770
	Mass.	Electronics	3100	5310	3250
	Mass.	Aero weapons	2500	5670	3050
	Conn.	Pumps	745	5890	3800
	Wis.	Heavy machinery	630	4830	3500
	N.Y.	Iron foundry	105	6400	--
	Ind.	Instruments	620	6330	4040
	N.Y.	Appliances	4200	6010	3790
	Wis.	Iron castings	115	4950	--
	Ill.	Retail	2000	5230	2190
	Ill.	Phone equipment	900	4560	3130
	N.C.	Textiles	505	3080	2230
	Mass.	Valves	415	5380	--
	Pa.	Glass	490	4730	3310
	Wis.	Small machinery	1000	5340	3610
	Minn.	Banking	1550	5670	3320
	Ill.	Power	720	6300	4000
	Tenn.	Chemical production	3000	6680	4850
	Tenn.	Chemical research	4000	7305	5270

Characteristics Of Communities In Normative Sample, 1960

PLANT	POPULATION (in 1000s)	APPROXIMATE MEDIAN FAMILY INCOME	MEDIAN MONTHLY RENT	% OWNER OCCUPIED HOMES	MEDIAN HOME VALUE (in $1000s)	MEDIAN AGE
1	3800	6800	79	71	13	29
2	2000	7600	88	49	17	29
3	140	5200	56	44	12	34
4	2600	6700	82	52	16	32
5	2600	6700	82	52	16	32
6	100	8000	91	68	23	32
7	1200	7000	88	58	16	30
8	1300	6500	75	61	15	31
9	110	6500	71	78	10	28
10	560	6400	79	67	13	30
11	1200	7000	88	58	16	30
12	6200	7300	90	51	19	31
13	50	6000	87	64	14	26
14	130	4700	51	61	7	26
15	2600	6700	82	52	16	32
16	40	4800	55	74	7	28
17	1200	7000	88	58	16	30
18	1500	6800	78	68	15	28
19	290	6300	71	70	12	30
20	60	5900	63	69	7	26
21	60	5900	63	69	7	26

pared to Smith's representative sample, low (24th percentile) in comparisons of attitudes toward coworkers, and above average (60th percentile) in comparisons of attitudes concerning pay.

However, Smith et al. (1969) stratified their general norms on income, education, job level, community prosperity, and community decrepitude. This stratification of norms allowed comparisons which took into account the specific characteristics of the sample. The most striking change occurred when the norms for the ATC means were adjusted for income. All three attitude areas dropped sharply, indicating that, compared to persons in the normative sample earning comparable salaries, the ATCs' attitudes toward work (8th percentile), coworkers (15th percentile) and pay (17th percentile) were quite poor. The ATC group was, however, compared with persons in the Smith et al. (1969) sample who were earning a comparable salary of "$14,000 and over" in 1960. As contrasted with the journeymen ATCs, the Smith sample most likely would be most entirely from the supervisory and professional levels. Thus, this comparison may not have been altogether valid, and the pattern in Table 37 may present a better understanding of the ATC scores.

It seems apparent from the table that ATCs were most satisfied with their pay, and slightly above average compared to other workers. They were least satisfied with their coworkers and clearly below average in comparison with other workers. It also seems apparent from the comparison data given in Table 37 that ATCs, in comparison with other workers, were about average or slightly below average in satisfaction with work. There was no strongly favorable attitude area for ATCs when compared with other workers, although their satisfaction with pay was above average.

LBDQ. Normative comparisons for these scales were not possible, as norms were not developed. However, it was still possible to compare the ATCs' scores with other groups because Stogdill (1963) reported means and variances from a wide range of different samples. Table 40 presents the ATC results against the results reported by Stogdill. In the table, the means represent subordinates' descriptions of how frequently their supervisor/leader engaged in a given behavior. Thus, the headings of the columns refer to the supervisors of the group members who provided the behavioral ratings.

Several interesting facts emerged from Table 40. First, ATC supervisors were significantly lower than other supervisors for all three supervisory behaviors. With respect to Initiation of Structure, the ATCs described their supervisors as not

Leader Behavior Description Questionnaire Results

SCALE	ATC SUPER-VISORS	ARMY OFFICERS & NCO's	AIRCRAFT EXECU-TIVES	COMMANDING POLICE OFFI-CERS	MINISTERS	COMMUN-ITY LEADERS	LABOR UNION PRESI-DENTS	CORPORATION PRESI-DENTS
Initiation of Structure (IS)	31.78	38.60	36.60	39.70	38.70	37.20	38.30	38.50
	6.92	5.70	5.40	4.50	4.90	5.70	5.60	5.00
	409	235	165	185	103	57	44	55
Consideration (C)	33.08	37.10	37.10	36.90	42.50	41.10	42.30	41.50
	7.12	5.60	5.80	6.50	5.80	4.70	5.50	4.00
	408	235	165	185	103	57	44	55
Tolerance of Freedom (TF)	35.13	35.90	38.00	36.30	37.50	36.40	38.00	38.90
	4.77	6.50	5.90	5.30	6.00	5.00	4.00	4.90
	408	235	165	185	103	57	44	55

All cells read: Mean
Standard Deviation
Number of Subjects

clearly defining their own roles and not letting the ATC know what was expected. This finding was consistent with the observations of the ATC HCS team when visiting the facilities as well as in interviews with the ATCs. This may stem partly from unique features of ATC work that emphasize a great deal of individual responsibility. ATCs, unlike many other employees, are highly trained when they begin on the job; thus, the amount of directive behaviors by their supervisors to ensure adequate job performance may be low.

The ATCs reported less frequent consideration by their supervisors than that reported for other supervisors. This result probably indicated a low level of interaction between supervisors and ATCs.

However, the last comparison from Table 40 was inconsistent with this low interaction interpretation; that is, ATCs, in comparison with other groups, reported that their supervisors had about the same, or a slightly lower frequency of Tolerance of Freedom behavior. Thus, the ATC supervisor's behavior, in terms of allowing his employees scope for initiative, decision and action, was perceived by the ATCs as slightly lower than other groups. If there were a low level of interaction between the ATCs and their supervisors as previously argued, one would expect the mean for Tolerance of Freedom to be higher. Examination of the pattern of means for the other groups in the table illustrates the rationale for this argument. The means for Tolerance of Freedom, in general, were lower than those for Consideration and Initiation of Structure. Typically, a supervisor with high frequency of the Initiation of Structure and Consideration behavior was conversely reported to have less Tolerance of Freedom. Our ATCs showed a distinctly different pattern of means.

One explanation for this apparent inconsistency may be that the supervisors did not interact very much with the ATCs in terms of actual work performance; however, the supervisors may have been restrictive in terms of non-work performance activities, e.g., daily assignments, timing of work breaks, etc. Some anecdotal evidence from observations by the ATC HCS team of the ATCs in their work place provided some support for this observation.

KLAP. Since there were no norm groups or comparison samples for these variables, a different perspective was needed to interpret the ATC results on individual satisfaction and group morale. The primary purpose of attitude surveys in industry is to diagnose problem areas in organizations. Most companies prefer to develop locally meaningful surveys than to use standard surveys which have complete norms but are generic and non-specific in nature relative to the unique characteristics of that company. Thus, there has emerged a generally-accepted guideline or threshold as to when a certain level of dissatisfaction indicates a problem exists and management should take some action. This generally-accepted threshold value varies from 20% to 25% dissatisfied response.

This threshold value is based on several considerations. First, national surveys taken over the past 20 years indicate the level of dissatisfied workers varied from 8% to 19%, with a mean of 14% (Quinn et al., 1974). Second, from a practical viewpoint, many managers believe that having 1 in 5 or 1 in 4 of their workers dissatisfied is a serious problem. Finally, dissatisfaction of workers is costly in terms of turnover, absenteeism, and potentially lowered productivity. A recent empirical study (Mirvis and Lawler, 1977) found that "expected direct-cost savings of $17,664 in absenteeism, turnover, and performance" could result from a .5 standard deviation increase in job satisfaction.

Using this standard for dissatisfaction, an examination of Table 41. reveals several noteworthy findings. First, the consistent percentage of scores in the neutral category for both individual and group morale indicated that about 42% of the ATCs felt "so-so" in general about their jobs. The distribution for Individual Satisfaction was not severe, although about 21% of ATCs were dissatisfied with their overall personal satisfaction in their jobs, while 36% were personally happy with their jobs. The story changed markedly for Group Morale. The percentages were reversed -- 39% of ATCs were dissatisfied with the general morale in their jobs, while only 19% were satisfied.

Examination of the items in these scales (Table 42) provided one possible explanation for these results. The Group Morale scale was mainly concerned with management, supervision, and working conditions, all factors extrinsic to the ATC's actual work, while the Individual Satisfaction items were more personalistic and dealt with the controller's job. This would lead one to conclude that the major sources of dissatisfaction in the ATC's job were not in the work itself, but rather in the work environment surrounding him.

TABLE 41

Scale Score Distributions For The Kavanagh Life Attitude Profile

VARIABLE	PERCENT DISSATISFIED	PERCENT NEUTRAL	PERCENT SATISFIED
Individual Satisfaction	20.9	42.9	36.2
Group Morale	38.9	42.0	19.1

TABLE 42

KLAP Scales

INDIVIDUAL SATISFACTION

I am extremely underpaid for the work that I do.
I can learn a great deal on my present job.
There is too much pressure on my present job.
I am satisfied with my salary in relation to my job responsibilities.
I am satisfied with the extent to which higher management keeps me informed by written communication (about my job).
I often feel tired and worn out on my job.
My supervisor has always been fair in his dealings with me.
For my kind of job, they expect too much work from me around here.

GROUP MORALE

Management here sees to it there is cooperation between departments.
This facility operates efficiently and smoothly.
My supervisor has the job well organized.
Management really knows its job.
My supervisor knows very little about his job.
Poor working conditions keep me from doing my best work.
My supervisor lives up to his promises.
Management here does everything it can to see that employees get a fair break on the job.
For my kind of job, the working conditions are okay.
My supervisor gets employees to work as a team.

SUMMARY

Impact of Life Changes

To test whether or not life change events were important in the development or exacerbation of disease, an inventory of life events, called Review of Life Experiences (ROLE) was developed and consisted of appropriate items from the two major existing inventories, the Holmes and Rahe Schedule of Recent Experiences (SRE) and the Paykel, Uhlenhuth and Prusoff (PUP) questionnaire, and an additional 20 items representing important life events not included in the other two. A total of 103 items appeared in the ROLE.

The ATCs were asked to make two ratings for each event that had occurred to them, a rating from 1 to 99 for the adjustment the event required of them, and another rating from 1 to 99 of the distress the event caused them. Four different life change assessment scores were derived:

1) a score based on the Holmes and Rahe items, using their scoring method, which was based on weights placed on the different items in accordance with the amount of adjustment that a normative population group had indicated that item required;
2) a score based on the PUP items, using the Paykel, Uhlenhuth and Prusoff scoring method, which weighted the items by the amount of distress a normative population group had assigned to each item;
3) a score on the Holmes and Rahe items based on the adjustment weight each ATC himself assigned to that item; and
4) a score on the PUP items based on the distress weighting that the individual ATC assigned to that item. In addition all 103 items could be summed for a total distress score.

It was found that the ATCs assigned greater weight to the adjustment required by their life events than the Holmes and Rahe normative population did, but they were less distressed by their life events than the PUP normative population.

Because further analysis suggested that the normative weights reflected more the number of life events than the distress caused by them, it was decided to use the ATCs' own ratings of distress caused by the life changes in further predictive studies.

6. Impact of Life Changes

We knew from our previous reading of the psychology literature that life events were considered by many researchers to play an important role in the development or exacerbation of disease (Hurst, Jenkins and Rose, 1976, 1978). We were therefore interested from the outset of the study in ascertaining whether or not significant events occurring in the lives of the ATCs were related in time and in emotional effect to the onset of either physical or psychological diseases, or both. However, the method of quantifying life change events and attributing the derived measurements to appropriate conclusions presented certain problems.

Existing Methods of Measurement

Two major approaches had been developed for the systematic recording of life events and their effects on individuals. The Holmes and Rahe approach (Holmes, Rahe, Hawkins, Davies and others, 1957, 1967a, 1967b, 1967c, 1969, 1970, 1976) involved the development and testing of an instrument to assess the frequency of occurrence of 43 life events. This instrument, the Schedule of Recent Experience (SRE) was later revised, with an addition of 12 items, and renamed the Recent Life Change Questionnaire. Another instrument, the Social Readjustment Rating Questionnaire (SRRQ) was devised to be used with the SRE. By administering the SRRQ to research populations, the developers were able to obtain population ratings of the amount of readjustment required after the occurrence of specific SRE life events. These population estimates of readjustment, called life change units or LCU, were applied to the events checked by an individual on the SRE and summed to give an individual a total life change score.

The second approach, that of Paykel, Prusoff and Uhlenhuth (1971, 1972), involved use of a 61-item questionnaire to elicit the life changes that had occurred, and also to attach to these life events varying weights (LCW) that represented population ratings of the amount of distress caused by each specific life change. The Total Life Change score in this approach was the sum of the life changes that had occurred weighted by the distress a general research population had attached to each such life event.

The Holmes and Rahe and the Paykel, Uhlenhuth and Prusoff questionnaires overlapped somewhat in item content, and were similar to the extent that they both applied population estimates of the amount of distress or adjustment to the events reported by an individual to avoid the circularity that might occur in retrospective studies; that is, it was important in retrospective studies that the individual who became ill did not on that account ascribe belated importance to the life event occurring previously.

Although both the approaches were based on the number of significant life changes occurring to an individual, the Holmes and Rahe weighted the readjustment required while the Paykel, Prusoff and Uhlenhuth weighted the distress produced by the event.

Our research into the value of using one or the other method of assessing life change raised a number of questions that led eventually to our use of the items from the two available instruments with other items that we added together with a different weighting perspective, in a new questionnaire that we called the Review of Life Experiences (ROLE).

These questions involved the applicability of the population ratings established by the original research to other groups and cultures, the possible differences in ratings of impact when the events were actually experienced as opposed to estimates of impact when they were not actually experienced, the effect on the ratings of the recency of the event, the age and the sex of the rater.

Although the developers of both existing instruments declared that the normative weighting units used for each were applicable to many different groups, and additional research supported this contention in terms of the rank ordering of life changes, little was known about the way in which cultural group differences affected the total life change score, the variable we intended to use for our predictive studies. In addition, results of several studies of differences in ratings for recent events and for the same events when they were more remote, and results of studies of age and sex differences in rating, all caused us to decide that the state of the art of measuring life change events and interpreting the measurements was neither sufficiently advanced nor expedient to merit the use of the existing methods without modification.

One modification that appeared to us to be very important and that was suggested in the literature by Rahe (1976) and others (Goldberg and Comstock, 1976; Paykel, Prusoff and Uhlenhuth, 1971) after some years of research with the original approach was the use of individuals' idiosyncratic ratings for weights instead of the normative ratings. The results of several studies suggested that the individual's own ratings may have better power to predict disease than normative ratings (Lundberg, Theorell and Lind, 1975; Rubin, Gunderson and Arthur, 1969, 1971).

Additional details of the characteristics of the existing instruments, the results of research involving their use, and the questions that have been raised can be found in three papers prepared for publication under auspices of this study which are included in Appendix II (Hurst, Jenkins and Rose, 1975, 1978, and Hurst, forthcoming, 1979).

Review of Life Experience (ROLE)

After studying the questions briefly summarized above and concluding that neither one instrument nor the other sufficiently satisfied our purposes, we devised a composite questionnaire, called the Review of Life Experience (ROLE). The ROLE included 39 of the 43 items of the Holmes and Rahe SRE and 52 of the 61 items of the Paykel et al. questionnaire. We added 20 new items that represented important life crises not included in the two existing scales. In addition some of the standard items in both lists were broken down into their positive and negative aspects. The ROLE as it was constituted allowed us to create Holmes and Rahe equivalent total LCU's, and Paykel, Prusoff and Uhlenhuth equivalent total LCW's. After indicating that any listed event had occurred, the ATC was asked to make an adjustment rating and a distress rating from 1 to 99 for that event.

The instructions for adjustment and distress were carefully written to distinguish between the two, and a linear scale was presented for each to illustrate how to assign a value. The scale had markers every 10 points. Beneath the 1-10 range on the adjustment scale was written "vacation." "Marriage" was written beneath the range of 44-55. "Death of spouse" was written beneath the range of 90-99. The distress ratings, to be made after the adjustment ratings, were accompanied by a scale with "child married with your approval" written under the 1-10 range, "additional person in house" written under the 45-55 range, and "child died" written under the 90-99 range. A copy of the ROLE with instructions is included in the Exhibits appended.

Although we had excluded certain items of each of the schedules because they were not applicable to our group (e.g., questions concerning pregnancy, retirement, etc.), the total scores of our ATC respondents on the Holmes and Rahe items would be the same as if we had used the original schedule because the items excluded would not have occurred to them anyway; and the ATCs' scores on the Paykel et al. items would also correspond to their scores on that original schedule.* Therefore, we were able to compare our ATCs with the normative groups on each set of items, and as the adjustment and distress ratings were made for all 103 ROLE items, we were also able to ascertain whether or not there were significant differences between the amount of adjustment and the amount of distress caused by the specific life change events.

---

* The original Paykel et al. inventory used a range of 1-20 rather than the 1-infinity scale, later reduced to 1-100, used in the Holmes and Rahe schedule. Therefore, for comparison with the Paykel norms, we simply divided the ATCs' scores by 5 to obtain equivalence.

The ROLE questionnaire provided, therefore, four primary life change assessment scores: 1) the Holmes and Rahe normative total LCU equivalent, 2) the Paykel et al. normative total LCW equivalent, 3) an individual's total life change adjustment rating for the Holmes and Rahe items, and 4) an individual's total life change distress rating for the Paykel et al. items. In addition it was possible to sum across all 103 items for a total distress score.

Procedure

At the intake visit to Boston University Medical Center, the ATCs completed the ROLE at video terminals, responding interactively to the computer displays. Previous instructions on using the terminals had been provided and checking and verification programs were built in to preclude invalid responses.

The entire list of 103 items was presented first, one at a time. A subject indicated if the event never occurred in the prior 2 years, if it occurred in the past 6 months, if it occurred in the previous period of 7-24 months, or if it occurred both in the past 6 months and in the preceding 7-24 months. The computer was programmed to accept only one of the four alternatives. At the end of the presentation and response to all 103 items, the subject had an opportunity to change answers. The computer re-displayed the item, and the new answer could be substituted.

When satisfied with his response concerning occurrence or non-occurrence of events, a subject was presented with the instructions for adjustment ratings. The anchoring scale was continuously displayed in a protected field of the video display. The computer retrieved items that the subject indicated had occurred. It displayed the item and requested an adjustment rating in accordance with the instructions. After receiving adjustment ratings for all of the events that had occurred to the subject, the computer displayed the instructions for distress ratings along with the distress scale. Again the items were displayed and a distress rating was requested. Hence ratings were obtained only for those events that had occurred to a subject.

Descriptive Findings

Table 43 displays a comparison of total life change scores using the individual's ratings with the scores produced when the normative ratings were used. The table shows that the individuals' own ratings of the adjustment required by their life change events produced a significantly higher life change score than that produced by the same life changes weighted by the normative LCU of the Holmes-Rahe items ($t = ?.99$, $P < .01$). On the other hand, the

TABLE 43

Comparison Of Total Life Change Scores Using Individuals' Ratings versus Normative Ratings

	N	Mean Total	Standard Deviation	t for correlated means	df
SRE total LCU	356	111.04	75.85	2.99*	355
Total individual adjustment ratings for SRE items	356	124.84	118.38		
PUP total LCW	356	30.68	23.88		
Total individual distress ratings for PUP items	356	15.97	17.32	-16.91*	355

*$p < 0.01$

ATCs' own ratings of distress resulted in a significantly lower mean total score than that produced by applying the normative LCW to the Paykel et al. items.

The intercorrelations of the life change total scores produced by the four different assessment methods are displayed in Table 44. The highest correlation, .29, was between the SRE total LCU and the PUP total LCW, attributable partially to the overlap of 25 items between the two schedules. The total individual adjustment rating scores correlated .68 with the normative scores and the total individual distress rating scores correlated .73 with the normative weighted scores.

In summary, the life change events of the ATCs appeared to require more adjustment on their part than other groups have indicated, and appeared to cause less distress than other groups have experienced.

The frequency distributions of the number of life events that occurred for each man in the most recent 6-month period are shown for the SRE and the PUP schedules in Table 45. The distributions for both inventories were highly skewed, with that for the SRE events less skewed than the PUP inventory (significance of the difference, Kolmogorov-Smiroff D = .23, P <.01). The median number of Holmes and Rahe events per individuals was 4, or 10.2% of those possible; for the PUP events it was 3, or 5.8% of those possible. Overall, for all 103 items in the ROLE, the average proportion of events per individual was 11.7%. The ROLE event frequency distribution was less skewed than that of the two contributing inventories.

As further analysis suggested that the normative weights may reflect more the number of life change events than the stress generated by them (see Hurst, Jenkins and Rose, 1978), we saw additional justification for proceeding with our predictive studies using the individually rated life change scores.

The extent to which this variable was related to the development of physical or psychiatric illness is discussed in Section V, Predictive Findings.

TABLE 44

Intercorrelation Matrix Of Life Change Total Scores For Four Assessment Methods (N = 356)*

	SRE Total LCU	PUP Total LCW	Total Adjustment	Total Distress
SRE total LCU	1.00	0.79	0.68	0.59
PUP total LCW	0.79	1.00	0.56	0.73
Total adjustment score for SRE items	0.68	0.56	1.00	0.71
Total distress score for PUP items	0.59	0.73	0.71	1.00

* All correlations significant at $P < 0.0001$, 2-tailed

Number of events	HR SRE Absolute frequency	HR SRE Relative frequency	HR SRE Cumulative relative frequency	PUP Inventory Absolute frequency	PUP Inventory Relative frequency	PUP Inventory Cumulative relative frequency
0	12	3.0	3.0	21	5.2	5.2
1	27	6.7	9.7	61	15.1	20.3
2	40	9.9	19.6	78	19.3	39.6
3	61	15.1	34.7	73	18.0	57.6
4	76	18.8	53.5	43	10.6	68.2
5	51	12.6	66.1	47	11.6	79.8
6	30	7.4	73.5	31	7.7	87.5
7	33	8.1	81.6	16	4.0	91.5
8	29	7.2	88.8	13	3.2	94.7
9	21	5.2	94.0	6	1.5	96.2
10	10	2.5	96.5	5	1.2	97.4
11	6	1.5	98.0	9	2.2	99.6
12	4	1.0	99.0	2	0.5	100.1
13	2	0.5	99.5	0	0	100.1
14	2	0.5	100.0	0	0	100.1
15	0	0	100.0	0	0	100.1
16	0	0	100.0	0	0	100.1
17	1	0.2	100.2	0	0	100.1
Totals	405	100.2[a]	100.2[a]	405	100.1[a]	100.1[a]
Mean (SD)	4.9(2.8)			3.6(2.6)		
Median	4			3		
Mode	4			3		

[a] These totals do not add up to 100% due to rounding error.

SUMMARY

Measurement of Air Traffic Controller Workload

In order to determine whether or not the air traffic controllers' differential psychologica, behavioral or physiological responsivity on the job was related to their health changes, their differential workload needed to be evaluated. Components of workload assessed were:

1) Peak aircraft, the raw maximum number of planes coming under control within the 2 minutes before physiological measurements were made;
2) Transitioning aircraft, a count of the number of planes climbing or descending in the few seconds before the physiological measurements;
3) Proposals, the planes expected to come under a controller's legal authority;
4) Duty position, one of seven positions: data or clearance delivery, hand-off or ground control, radar or local control, check-ride monitor, training, radar coordinator, or supervisor; or on break;
5) Sector Types, one of five types: arrival/departure, enroute low, transition low, enroute high, transition high;
6) Percentage of air space not available because of military operations or weather conditions.

Other workload measures were derived as follows:

1) Normalized peak traffic, a measure reflecting the difference between peak traffic at any place/time/position and the average peak traffic that could be expected for the facility/sector/position at that time, allowing comparisons of men's responses to change in workload regardless of facility, sector or position worked;
2) Maximum peak traffic, the highest raw peak traffic count recorded for a man during a field study day;
3) Maximum normalized peak traffic, the highest normalized peak traffic for a man in a day,
4) Range in peak traffic, representing the fluctuation in a man's workload for a day;
5) Range in normalized peak traffic, representing the fluctuation in normalized work load;
6) Maximum consecutive time on position, the maximum number of consecutive observation periods during a day that a man worked a journeyman position without a break;

7) Total time on a working position, the number of times, of a possible 15 in a study day, that a man worked on a journeyman position;

8) Training time, the amount of time a man spent during the 5-hour study day in training developmental controllers;

9) Training work, the peak traffic counts of the trainee while being trained by a subject controller.

These measures were transformed into the following measures, each representing some unique dimensions of workload for the day:

1) Paceload, the sum of raw peak traffic recorded for a man during the day, maximum peak traffic and range in peak traffic over the day;
2) Normalized workload, the sum of normalized peak traffic, maximum normalized peak traffic and range in peak traffic for a day;
3) Timeload, the sum of maximum consecutive time on position and total time on position;
4) Area workload, a measure of airspace load estimated by number of transitioning planes plus a weighted index applied to the five types of sectors;
5) Time training, the measure defined as a component of workload;
6) Work training, the measure defined as a component of workload.

Because the workload of individuals varied from day to day, an average paceload, average normalized workload, average timeload, etc. was calculated to characterize each man's work across 10 field studies. These average measures were used in some analyses, while for other analyses, the specific value of a variable at a particular field visit was used.

The following findings emerged:

1) Paceload increased, on average, by 50% over the course of the study;
2) Paceload differed significantly among the facilities;
3) The oldest group of controllers tended to have a lower mean paceload, but there was no significant difference between three groups clustered by length of experience;
4) Facilities did not differ in average normalized workload, as expected by design of the variable;
5) Neither age nor experience grouping provided systematic differences in normalized workload;

6) Timeload differed at the different facilities;
7) Age and experience differences in timeload were not statistically conclusive;
8) Facilities differed significantly in amount of time spent training;
9) The oldest and most experienced ATCs spent the least amount of time in training others;
10) Trainees at the NYARTCC handled more aircraft than trainees at other facilities;
11) There was a tendency for the trainees of older and more experienced controllers to handle fewer aircraft.

7. Measurement of Air Traffic Controller Workload

A major emphasis of this study was to examine air traffic controllers on the job to determine if their differential psychological, behavioral or physiological responsivity on the job was related to the development of health changes. Consequently, in May of 1974 we began studying men for five-hour periods during randomly selected work days, measuring blood pressure, collecting blood samples for hormone assay, observing behavior and noting workload. At the conclusion of our field studies in July, 1977, we had completed 1,470 blood pressure studies and 1,156 hormone studies yielding a total of 2,626 man-days of studies at work. As some of these studies represented augmentations of the routine ten visits to the field and included a small number of highly cooperative men, our basic data consisted of 2,570 man-days of observation at work, resulting from the five studies of blood pressure and five studies of endocrine variables. Since each man-day consisted of five hours of observation and three observations were made for each hour, a total of 38,550 concurrent observations of physiology and work formed the basis for our assessment of workload and responsivity on the job. For some studies of workload alone, 2626 man-days of data were used.

Table 46 displays the number of field studies completed for the participants. As can be seen, almost all controllers were subject to at least one field study. Those few who were not had dropped out of the study or were ineligible for field studies by the time that they were approached. On the average, 6.2 days of observation were completed for the original 416 air traffic controllers who volunteered for this study. More than half of the controllers had six or more field studies completed by July, 1977.

Workload was observed and measured concurrently with measurement of blood pressure and collection of blood samples for endocrine studies. Table 47 displays the number of participants for whom blood pressure and endocrine field studies were completed. Blood pressure studies were completed for more men than endocrine studies, a result which we attribute to the less invasive blood pressure measurement equipment.

Each field study included an assessment of eight possible moods at the beginning of the day (see Section IIIB8). After the ATC completed the mood questionnaire, there were either 15 measurements made of blood pressure or 15 collections of blood serum at 20-minute intervals. Each of these physiological events took place two minutes after the evaluation of eight different work variables. At the end of the study day, a controller completed a 17-item "subjective difficulty" questionnaire that inquired about his opinion of the day's work dif-

TABLE 46

Number of Field Studies
Completed by Participants

Number of Studies	Number of Men Men Completing	% Of Original 416 ATC Volunteers
0	27	6.5
1	389	93.5
2	367	88.2
3	347	83.4
4	325	78.1
5	289	69.5
6	263	63.2
7	229	55.0
8	180	43.3
9	123	29.6
10	58	13.9

Total Field Studies = 2570

Average Field Studies per Participant = 6.2

TABLE 47

Participants For Whom
Blood Pressure and Endocrine
Field Studies Were Completed

of s	Number of Men Completing Endocrine Studies	% of Original 416 ATC Volunteers	Number of Men Completing Blood Pressure Studies	% of Original 416 ATC Volunteers
	369	88.7	382	91.8
	316	76.0	340	81.7
	243	58.4	288	69.2
	158	38.0	249	59.9
	63	15.1	162	38.9
Studies	1149		1421	

Field Studies = 2570 Man-days
Hours of Observation = 12,850 Man-hours
observations = 38,550 Observations

ficulties. A description follows of the workload variables that were assessed by direct observation and those that were derived, and our method of determining the most complete and appropriate measures of workload.

Components of Work That Were Assessed

Peak Aircraft. A number of previous studies, including two preliminary studies of this research team, indicated that the raw maximum number of planes coming under control within a unit of time was the best objective index of workload for controllers. These previous studies showed that peak traffic was significantly related to "pace" (ratings of behavioral arousal and activity while controlling aircraft), heart rate and observer-rated stress (Hurst and Rose, 1978a, 1978b; Laurig, Becker-Biskaborn and Reiche, 1971; Philip, Reiche, and Kirchner, 1971).

In these previous studies the unit of time during which the maximum peak traffic was observed varied from 1 minute to 10 minutes. The present research team conducted a pilot study varying the observation time from 2 minutes to 10 minutes at 1-minute intervals. We found that the peak traffic during a 2-minute period of time was highly correlated (above .80) with the peak traffic that would be observed in 10 minutes. Consequently, for the purpose of this study, peak traffic was assessed in ten two minutes prior to a physiological measurement or collection in order to allow the field technicians the opportunity to study more than one man per day.

Transitioning Aircraft. Two published studies also suggested that the number of transitioning planes, i.e., those ascending or descending, might be a relevant objective index of workload (Hurst and Rose, 1978a, 1978b). These studies investigated the theoretical variables of Arad (1963, 1964a, 1964b), called control load variables. Although Arad's work was primarily intended for the optimization of sector design characteristics, we used his formulations in a pilot study. The pilot study included data from the sectors in the New York and New England Air Route Traffic Control Centers and from all of the major IFR airport towers in the continental United States. The results of the pilot study showed that the number of transitioning aircraft was significantly related to Arad's airspace control load variable and was in fact an important component of workload. Consequently, our field studies also included a count of the number of planes climbing or descending (transitioning) in the few seconds before the physiological measurements or collections.

Proposals. "Proposals" were considered to be a third possible component of workload. These are the computerized flight plan reports received and posted at an air traffic controller duty station, and they represent those planes that can be expected to come under the controller's legal authority. Although no previous research has established the relevance of proposals, our observations in the facilities had indicated that at least some men experienced anticipatory reactions when their proposal board filled up. Consequently, our field technicians also noted the number of proposals on the board in the seconds before the physiological measurement or collection.

Duty Position. The position at which a controller was working also was considered a component of workload. Controllers in air traffic control centers could be working any one of seven positions: data, hand-off, radar, check-ride monitor, training, radar coordinator or supervisor. In towers, controllers could be working data, clearance delivery, ground control, local control, check-ride monitor, training or supervisor positions. Observations in the field and preliminary analyses indicated that the data and clearance delivery positions were essentially equivalent; that the hand-off and ground control positions were essentially equivalent; and that the radar and local control positions were essentially equivalent. Consequently, we combined these positions so that seven position codes were available to be assigned each time a man was observed. Only four of these positions were technically working positions for journeyman controllers with direct legal responsibility for aircraft separation. When a man was on break (either in or out of the control room) or when a man was not assigned a specific job task, a special code was assigned. The positions were as follows:

On break (in or out of the control room) or not assigned. This position code was established to allow separate evaluation of our continuous physiological measures when a man was on a working position and when he was not.

Data or clearance delivery. The data position required a controller to remove proposal strips from the computer and to update these as needed. The strips were placed in plastic holders and then put on the proposal board. The clearance delivery position in a tower required the transmittal of instructions and air field data for takeoff and landing as directed by the local or ground controller. In essence both the data and clearance delivery positions were clerical. These were not journeyman positions since there was no legal responsibility for aircraft.

Hand-off or ground control. The hand-off controller provided the majority of radio communications between sectors when an aircraft was leaving the legal control of one sector and entering the legal control of another sector. This task relieved the radar controller of these communications. Similarly the ground controller at a tower provided taxiing instructions before and after landing and relieved the local controller of communications with planes that were not yet in a legally defined air space. Hence the hand-off and ground controller positions were primarily communications positions without responsibility for the separation of aircraft in legally defined air space. They were, however, journeyman positions.

Radar or local control. The radar position was the most skilled position of a journeyman air traffic controller. He had the legal responsibility for separating aircraft within his legally defined air space (called a sector) in accordance with FAA standards. In a tower the local controller had this responsibility even though his controlling was primarily visual in nature once an aircraft was handed off from the arrival/departure radar sector of an air route traffic control center or the IFR room in a tower. These were journeyman positions.

Check-ride monitor. The check-ride monitor was a highly responsible position assumed primarily by experienced and highly-rated journeyman controllers. The controller in this position was responsible for a sector but another controller actually provided air traffic separation within that particular sector. The check-ride monitor's responsibility was not only to insure the separation of aircraft within the sector, but also to evaluate the performance of the controller actually providing separation services. A particular kind of stress accompanied this position since the check-ride monitor could not provide instructions or direct traffic without automatically failing the controller attempting to prove his mastery of the sector. This was a journeyman position.

Training. In the training position, a journeyman controller was plugged into a sector along with a non-journeyman controller. The journeyman controller provided on-the-job training for developmental controllers, i.e., those who had yet to prove mastery of all of the sectors within their areas and who had to pass the journeyman air traffic controller examination. The trainer was legally responsible for separation of traffic within the sector at all times. In order to provide appropriate training, however, it was necessary that the trainer allow the trainee some errors in judgment or planning, while making sure that this allowance did not become too generous. This was a journeyman position.

Radar coordinator. This position, primarily found in the Common IFR Room at Kennedy Airport, entailed coordination of radar controllers at several sectors. The radar coordinator an active separation responsibility and communicated both h working controllers and aircraft. This was not considd a journeyman position since it was supervisory, and a ar coordinator did not have legal control of aircraft.

Supervisor. Occasionally men in the study were assigned temporary supervision of an area. Although responsible for of the work within an area, the supervisor typically had direct responsibility for active separ tion of aircraft; ther a man assigned to the temporary su, isory position primarily responsible for maintaining ate staffing, ations on break, meal schedules, and the coordination of personnel activities. This was not considered a journeyposition since it was supervisory, and a supervisor did have legal control of aircraft.

Sector Types. Our preliminary work also indicated that type of sector worked was a relevant component of workload urst and Rose, 1978a; Hurst and Rose, 1978b). Five types of ctors were identified.

Arrival/departure sectors were defined as those in which rcraft were accepted from a local controller at a tower or r traffic was handed off to a tower or IFR room. The airspeed aircraft in such sectors was relatively low. The number of rcraft could become quite high. The area of airspace covered relatively small, and these sectors were low altitude.

A second type of sector was an enroute low sector, defined a sector in which air space extended up to 17,000 ft., and thin which less than half the aircraft were transitioning. e enroute low sectors had relatively greater numbers of airaft flying at relatively s ow speeds in areas slightly larger an those of arrival/departure sectors. Very often the majoty of traffic in these enroute low sectors was general busis aviation rather than commercial airliners since the latter y much higher when they are enroute to their destination.

A third type of sector was the transition low sector. The ansition low sector was characterized by altitude responsibity below 17,000 ft., but within which the majority of airaft were climbing or descending. Typically this was the secr to which an arrival/departure sector would hand off.

A fourth type of sector was the enroute high, defined as aring responsibility for aircraft above 17,000 ft., and withwhich less than half the traffic was climbing or descending. pically these sectors covered thousands of square miles and were concerned mainly with commercial aircraft on designated airways enroute to their destination at a defined altitude.

In contrast to the enroute high sector, a transition high sector was one in which a majority of planes were climbing or descending in the altitude strata above 17,000 ft. Transition high sectors also primarily handled commercial airline traffic and were responsible for thousands of square miles.

Percentage of Air Space Not Available. Finally, as suggested by Arad's work, our field technicians assessed the percentage of sector air space not available for controlling because of either military or weather conditions. Unavailable air space was an important consideration for workload assessment, since the amount of airspace within which a controller could maneuver aircraft affected his feeling of pressure and activity. Since most pilots do not like to fly through bad weather, airspace may become quite crowded when bad weather blocks normal air routes and controllers must direct pilots to other air space. In addition, the military occasionally block off certain sections of airspace in order to conduct military maneuvers. This reduces the amount of airspace within which a controller can direct air traffic and increases the density of operations.

These six variables were the main components of workload which we decided to observe directly. A number of other components were derived from our field observations.

Components of Workload That Were Derived

Normalized Peak Traffic. A considerable amount of time in the first year of the study was spent in the facilities. Two important observations were made. First, the air traffic conditions varied widely among the facilities. Second, the controllers adapted to and reacted to their environments in similar ways despite the different air traffic conditions. Consequently, we concluded that we needed a measure of air traffic that would be comparable across facilities and would reflect the observation that men's responses to air traffic conditions were relative to what conditions they usually encountered. The conditions to which they were accustomed included not only their facilities *per se*, but also the position that they were working and the type of sector that they worked. Table 48 displays a comparison of the peak traffic among the different facilities, positions, and sector types, confirming our observations regarding their differing traffic conditions.

	NYARTCC	BOSARTCC	NYCTFR	Tower	[illegible]	[illegible]	[illegible]	[illegible]
Number of Observations	13000	5335	2840	1320	508	377	328	63
Mean Peak ***	4.38	2.63	3.11	2.45	1.62	1.78	2.07	1.18
S.D.	2.65	1.84	1.76	1.76	1.23	1.38	1.49	1.17

$F = 541.23$ $df = 7,23763$ $p < .0001$

	POSITIONS			
	Hand Off/ Ground Control	Radar/ Local Control	Check-Ride Monitor	Training
Number of Observations	4622	15657	81	3411
Mean Peak	3.51	3.52	3.70	4.01
S.D.	2.50	2.43	2.09	2.51

$F = 38.89$ $df = 3,23767$ $p < .0001$

	SECTOR TYPES**				
	Arrival/ Departure	EnRoute Low	Transition Low	EnRoute High	Transition High
Number of Observations	9291	1283	5765	3445	3969
Mean Peak	3.20	4.45	3.27	4.28	4.07
S.D.	2.15	2.72	2.26	3.09	2.45

$F = 235.62$ $df = 4,23748$ $p < .0001$

* Of the 38,550 observations, 23,771 occurred on a working journeyman position (breaks, data coordinator and supervisor positions were excluded).

** 23,753 observations; 18 observations occurred on non-classifiable sectors and were excluded.

*** Mean peak represents the mean of all observations of the maximum number of planes within an ATC's control during a 2-minute period.

The mean peak traffic was significantly different among the different facilities, positions, and sector types, ranging from 1.18 at the Providence Tower to 4.38 at the New York Air Route Traffic Control Center. The mean peak traffic varied from a low of 3.51 at the hand-off (or ground control) position to a high of 4.01 at a training position, and from a low of 3.20 on arrival/departure sectors to a high of 4.45 on en-route low sectors.

To control for these systematic differences, we devised a measure that we called "normalized peak traffic." To accomplish this we computed the mean and standard deviation of peak traffic for every combination of facility, sector type and journeyman position in the entire series of field study observations. All calculations were based on a minimum of ten observations for each combination of the three variables. In cases where fewer than ten observations were available for a given combination, we took the next higher order mean and standard deviation. The order of importance for this special step was facility, sector, and position from highest to lowest. All of the statistics for the mean and standard deviation of peak traffic by every combination of facility, sector, and position are presented in Appendix I, Tables 46-52.

Using these statistics, we then calculated a normalized peak traffic for each sample observation in a day. This normalized peak traffic was equal to the raw peak traffic of the particular observation minus the mean peak traffic, divided by the standard deviation of peak traffic, where the mean and standard deviation were those of the facility/sector/ position being worked when the peak traffic observation was made, as in the following formula:

$$\text{NPT} = \frac{\text{PT of observation} - \text{Mean PT of Facility/Sect./Pos.}}{\text{S.D. P.T. of Facility, Sector, Position}}$$

If no peak traffic observation was made because the man was not working one of the journeyman positions, then a missing value was assigned. This method of calculating a normalized peak traffic was the equivalent of calculating a standardized z-score with a mean of 0 and a standard deviation of one for each of the 15 observations in a study day.

Therefore, the normalized peak traffic variable reflected the difference between the peak traffic at a particular point in time and the average peak traffic that could be expected for the facility/sector/position at that time. Since these normalized peak traffic measures were standardized in this possible to compare men's response to change in ardless of the facility, sector, or position they

other workload component variables were derived from observations within a work day.

Peak Traffic. The maximum peak traffic a man ing the five hours of observation provided a measure mum intensity of workload experienced. It was de- e highest raw peak traffic count recorded for a man eld study day.

m Normalized Peak Traffic. This variable represented normalized peak traffic experienced by a man during It was defined as the highest normalized peak traffic 5 but not including missing values. If a man never journeyman position as previously defined, then the rmalized peak traffic was set to a missing value.

in Peak Traffic. A man could have a day characterized total number of planes and a high maximum peak traffic, little variability. In order to assess variability, we range in peak traffic workload measure as the highest fic count minus the minimum peak traffic count for a dy day, where the minimum was not a missing value. This e us an indication of fluctuation in a man's workload.

e in Normalized Peak Traffic. Similarly, we defined in normalized peak traffic as the maximum normalized fic for the day minus the minimum normalized peak for the day, excluding missing values. Thus the range lized peak traffic represented the fluctuation in stand- workload.

imum Consecutive Time on Position. The amount of time s on a journeyman working position seemed an important t of workload. One measure of time on position was de- the maximum number of consecutive observation periods tal of 15) during a study day that a man worked a journey- ition without a break. Thus this variable could range ro, when a man never worked a journeyman position, to , when a man worked five hours straight without a break.

tal Time on a Working Position. A second method of evalu- time as a component of workload was the number of times, ossible total of 15 during a study day, that a man worked journeyman position. This variable could also range from zero fteen.

Training Time. One of the questions to be answered in accordance with our contract was whether or not training imposed a particular stress upon controllers. Consequently, we devised two measures of training load. The first measure was the amount of time that a man spent doing live traffic training of developing controllers during the five-hour study day. This number could range from zero to fifteen, representing zero to three hundred minutes.

Training Work. We also assessed the amount of work accomplished at the training position by using the peak traffic counts of the trainee. This was done by summing the peak traffic reading only for those sample observations for which the subject controller's position was that of training.

For both the training time and training work variables a missing value was assigned if the individual never worked a journeyman position. If a man worked at any time on a journeyman position but did not have any sample observation in the training position, then the training time and training work variables were set to equal zero rather than a missing value.

Summary Workload Measures

The assessed and derived components of workload required evaluation and organization into one or more measures representing workload for the day. Our measures of workload needed to account for absolute level, duration and variability if we were to demonstrate any relationship between workload and the level, duration and variability of physiological responses to workload. As a consequence of these considerations, six major summary measures were devised.

A large number of preliminary analyses were conducted in order to reduce and combine the workload components into the six summary measures presented and discussed below.

Paceload. Paceload was defined as the sum of 1) raw peak traffic recorded during the 15 2-minute observation periods of a man's study day, 2) maximum peak traffic and 3) range in peak traffic over the study day. This measure included an assessment of the absolute magnitude of workload, a measure of the highest absolute magnitude, and a measure of variability.

Table 49 displays a comparison of paceload among the facilities in the study in two random split-half samples. For these analyses the man-day was the unit of observation and the entire set of 2600+ man-days was randomly split into two. This method was used to insure that any significant results were replicable.

TABLE 49

Comparison of Paceload Among Facilities*

RANDOM SPLIT-HALF SAMPLE #1 (N=1302)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES
[illegible]of [illegible]tion	636	340	78	164	84
[illegible]ceload***	58.19	28.73	30.14	37.26	18.58
	21.99	13.73	15.12	14.39	8.23

F=214.99 df=4,1297 $p < .0001$

RANDOM SPLIT-HALF SAMPLE #2 (N=1304)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES
[illegible]of [illegible]tion	635	348	81	164	76
[illegible]ceload***	59.65	29.97	28.24	35.52	22.58
	24.58	14.03	11.34	14.19	11.06

F=187.29 df=4,1299 $p < .0001$

[illegible]vailable man-days were used, including sixth visits, N=2606, for which [illegible]oad could be calculated.

[illegible]onset, Otis, Providence, Bradley.

[illegible]ceload = the sum of raw peak traffic recorded while a man was on a journeyman [illegible]rking position, maximum peak traffic and range in peak traffic over the [illegible]udy day.

There was a significant difference in the paceload among the different facilities, the New York Air Route Traffic Control Center (NYARTCC) being certainly the busiest facility in the study in both sets of analyses. The New York Common IFR (NYCIFR) was the second busiest facility. The Boston Tower (Logan) and the Boston Air Route Traffic Control Center (BOSARTCC) were approximately equal in their paceload and finally the smaller facilities were substantially below the other facilities in their average paceload.

Although expressed in units of planes, the paceload, such as 58 planes at the NYARTCC, represents an index of how much work and how variable the work was in each of the facilities rather than an actual number of aircraft controlled.

Table 50 displays comparisons between controllers of three age ~~groups~~ and ~~three experience~~ groups in two random split-half analyses. In both analyses there was a significant tendency for the oldest group of controllers to have a lower mean paceload. On the other hand, the difference in the mean paceload between three experience groups was not significant. The amount of work accomplished during study days appeared, therefore, to be more an age-related then experience-related phenomenon.

Figure E displays the change in paceload over the course of ten chronological field studies. The vertical axis represents paceload; the horizontal axis represents the visit numbers in chronological order. The numbers at data point on the graph represent number of men included at that point. As can be seen in this figure, paceload decreased over the first three field studies and slowly rose to more than 60 aircraft, representing greater than 50% increase in average paceload over the course of the study.

It could be argued that the change in paceload over time reflected differences in the between-visit intervals for the men. However, the graphing of days after intake into the study against field visits (Figure F) showed a near-linear relationship between visits and days elapsed, indicating approximately equal intervals between visits. Therefore the displayed change in paceload over the course of the study represents a real change in paceload over the ten visits, not an artifact of widely different times between field visits.

2) <u>Normalized Workload</u>. The primary reason for devising a normalized workload measure was to enable us to compare workload across facilities in spite of their different environments. Normalized workload was defined as the sum of normalized peak traffic, maximum normalized peak traffic and range in peak traffic.

TABLE 50

Comparisons of Pace oad
Among Age and Experience Groups*

Random Split-Half Sample #1 (N=1302)

	Age (Years)			Experience (Years)		
f ion	25-34	35-38	39-49	3-7	10-14	15-20
	441	392	469	430	493	379
eload**	45.70	45.23	40.33	44.73	44.32	41.46
	23.30	23.72	22.89	24.16	22.92	23.04
	F=7.37 df=2,1299 p<.001			F=2.31, df=2,1299 p< .10		

Random Split-Half Sample #2 (N=1304)

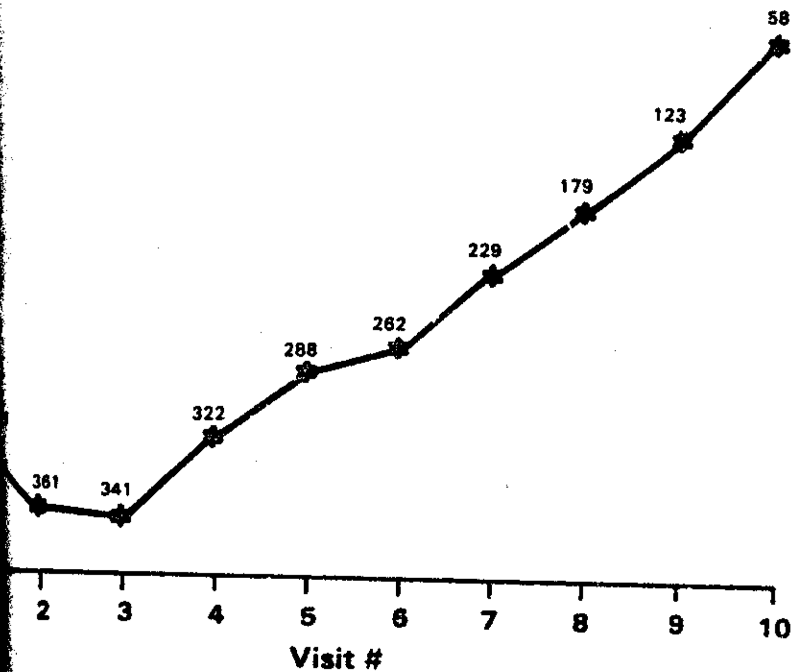
	Age (Years)			Experience (Years)		
of tion	25-34	35-38	39-49	3-7	10-14	15-20
	451	404	449	460	461	383
ceload**	46.57	45.00	42.67	45.92	44.59	42.97
	25.06	23.02	25.41	24.92	23.47	25.77
	F=2.82 df=2,1301 p<.06			F=1.49 df=2,1301 p>.20		

vailable man-days were used, including sixth visits, N=2606, hich paceload could be calculated.

oad = the sum of raw peak traffic recorded while a man was on a eyman working position, maximum peak traffic and range in peak ic over the study day.

FIGURE E

CHANGE IN PACELOAD OVER TEN FIELD VISITS



Note:
The number above each data point is equal to the number of man-days used to compute the average pace load at each visit.

The normalized workload variable was calculated first in a standard score from (mean = 0, standard deviation = 1) and then was converted to a scale where 500 represented the mean normalized workload across all men and all visits, with a standard deviation of 50.

Table 51 displays a comparison of normalized workload among facilities in two random split-half samples of over 2000 man-days of observation. In both samples, as expected, there was no significant difference between facilities; an average normalized workload in one facility was equivalent to the average normalized workload in another facility.

Table 52 displays a comparison of the normalized workload among age and experience groups. Again the analysis was done as two random split-half samples. In the first random sample there was neither a significant difference in the normalized workload between three age groups nor between the three experience groups. On the other hand, the second sample indicated a significant difference in both age and experience groups, where in both cases the middle of the age range and the middle experience range showed a significantly lower normalized workload. However, since these results were not replicated in the two random samples, they may be regarded as chance variation. We concluded that there was no systematic evidence of either older men exceeding younger men or the reverse in normalized workload over their entire period of time in the study.

The normalized workload across 10 field visits in chronolo- ical order is graphed in Figure G. As can be seen, the normali- ed workload decreased over the first three visits and increased ver the remaining seven visits. Although this graph is of nor- lized workload, it is very similar to the paceload graph de- icted in Figure E. The normalization procedure did not alter he changing trend in workload over ten consecutive field visits.

3) Timeload. The variable, "timeload," was deduced from number of preliminary analyses indicating that the two time omponents of workload, maximum consecutive time on position and otal time on position, were related more to one another than the other components of workload. Consequently, timeload was mputed as the sum of maximum consecutive time on position and tal time on position. The resulting numbers are in units of -minute periods, such that a timeload of 10 represents ten -minute periods or a grand total of 200 minutes.

Table 53 displays a comparison of the timeloads among the ilities in the study. Both split-half samples produced a nificant difference in the timeload measured at the facilities.

TABLE 51

Comparison of Normalized Workload
Among Facilities*

Random Split-Half Samples #1(N=1302)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
f ion	636	340	78	164	84
zed d***	497.93	498.14	503.22	501.58	490.72
	53.21	44.25	50.26	46.61	38.87

$F=.95$ $df=4,1297$ $p > .50$

Random Split-Half Sample #2(N=1301)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES
of tion	633	348	81	163	76
ized d***	503.53	499.69	496.91	498.76	507.82
	56.90	45.87	(34.84)	(43.60)	46.18)

$F=.94$ $df=4,1296$ $p > .50$

available man-days were used, including sixth visits, N=2603 , for
h normalized workload could be calculated.

nset, Otis, Providence, Bradley

rmalized workload = the sum of normalized peak traffic, maximum
rmalized peak traffic and range in peak traffic.

TABLE 52

Comparison of Normalized Workload Among Age and Experience Groups*

Random Split-Half Sample #1 (N=1302)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
on Days	441	392	469	430	493	379
d **	500.32	495.85	498.35	496.95	498.56	499.38
	49.07	48.48	49.83	49.48	48.28	50.03
	F=.86 df=2,1299 p>.50			F=.26 df=2,1299 p>.50		

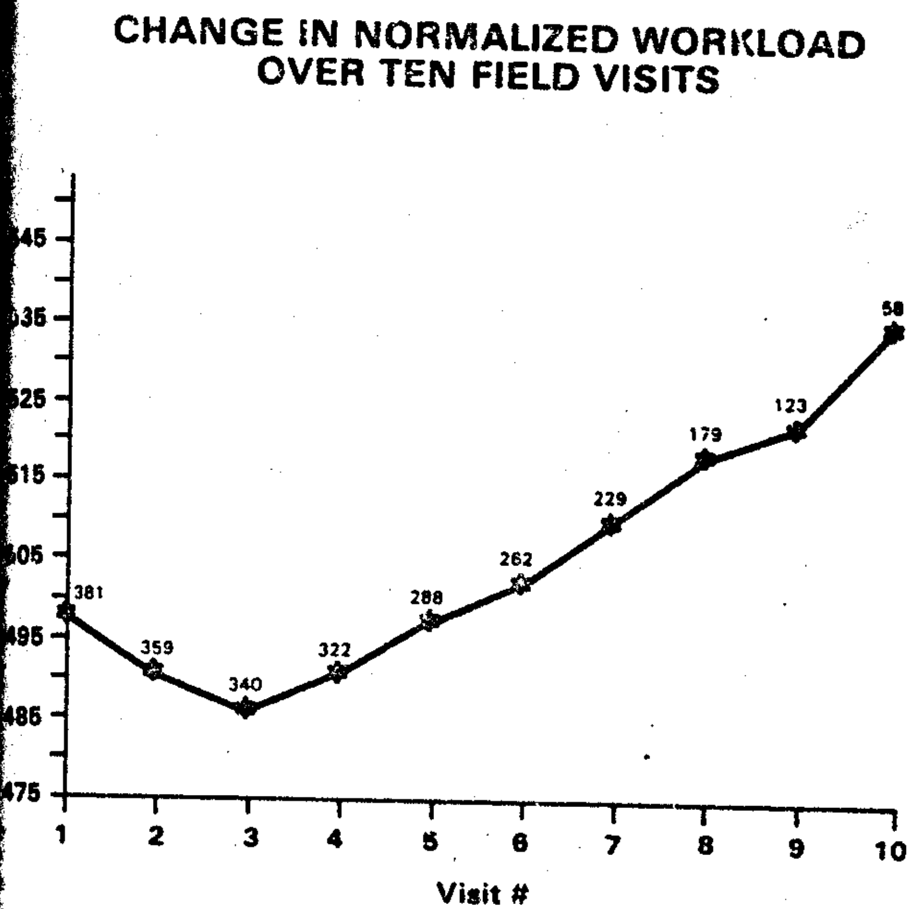
Random Split-Half Sample #2 (N=1301)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
	449	403	449	458	460	383
	506.47	498.41	500.01	506.66	498.02	500.34
	50.08	50.34	51.59	50.11	49.38	52.82
	F=3.09 df=2,1298 p<.05			F=3.55 df=2,1298 p<.03		

le man-days were used, including sixth visits, N=2603, for
lized workload could be calculated.

workload = the sum of normalized peak traffic, maximum
peak traffic and range in peak traffic.

FIGURE G



Note:
The number above each data point is equal to the number of man-days used to compute the average normalized workload at each visit.

TABLE 53

Comparison of Timeload
Among Facilities*

Random Split-Half Sample #1(N=1302)

	NYARTCC	BOSARTCC	JOSTON TOWER	NYCIFR	SMALLER FACILITIES**
of ation Days	636	340	78	164	84
imeload***	15.66	11.87	13.21	13.96	13.79
	3.16	3.71	3.72	3.49	3.75

$F=70.54$ $df=4,1297$ $p<.0001$

Random Split-Half Sample #2(N=1304)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
r of vation Days	635	348	81	164	76
Timeload***	15.50	12.28	13.53	13.48	13.13
	3.46	3.75	3.69	3.66	3.42

$F=49.97$ $df=4,1299$ $p<.0001$

available man-days were used, including sixth visits, N=2606, for which
load could be calculated.
nset, Otis, Providence, Bradley
eload = sum of maximum consecutive time on position and the total time
position.

The New York Air Route Traffic Control Center had the highest mean timeload. The Boston Tower, the New York Common IFR Room and the smaller facilities had approximately equal timeloads, and the Boston ARTCC had the lowest mean timeload.

Table 54 displays a comparison of the timeload among the several age and experience groups. The results for the first random split-half sample indicated that there was a significant difference between both age and experience groups. The oldest and the most experienced groups had significantly lower timeload. However, the second random split-half sample indicated that this result may have been spurious since there was no significant difference between the age or experience groups on the timeload variable. The same trend was, however, apparent in the second sample: the oldest group and the most experienced group tended to have lower timeload.

The mean timeload over 10 chronological field visits is graphed in Figure H. As with the paceload and normalized workload variables, timeload decreased from the first to the second visits, but while workload continued to decrease, timeload began to pickup between the second and third visits. The remainder of the figure shows that timeload generally tended to continue increasing over the remaining field studies.

Comparison of Figure H with E and G shows that the total amount of time men spent on position and consecutively on position increased very rapidly in advance of changes in either of the traffic workload variables. Timeload was characterized by rapid rises followed by a leveling off, which would seem to suggest an adaptation. It would seem that the men spent more time on position to cope with increased traffic, but after a while, their time on position leveled off, perhaps indicating greater efficiency. After a while at the same timeload, the traffic workload increased sufficiently to cause the men to increase their time on position again, after which it leveled off.

4) Arad Workload. Analyses reported in two previous studies caused us (Hurst and Rose, 1978a, Hurst and Rose, 1978b) to decide to calculate a summary measure of workload based on the formulations of Dr. Arad, a physicist who worked with the FAA to develop a theoretical framework for assessing what was called airspace control load (Arad, 1963, 1964a, 1964b). The previous two studies used data from all of the sectors involved in the present health change study and in addition, from all of the radar rooms at major airports in the U.S. The results of those studies indicated that Arad airspace load was best estimated by the number of transitioning aircraft plus a weighted index applied to the

TABLE 54

Comparison of Timeload
Among Age and Experience Groups*

Random Split-Half Sample #1(N=1302)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
on Days	441	392	469	430	493	379
load **	14.50	14.59	13.56	4.48	14.35	13.65
	3.55	3.79	3.88	3.56	3.88	3.82

F=10.34 df2,1299 $p < .0001$ F=5.56 df=2,1299 $p < .005$

Random Split-Half Sample #2(N=1304)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
n Days	451	404	449	460	461	383
oad**	14.34	14.23	13.82	14.27	14.22	13.84
	3.86	3.88	3.76	3.88	3.68	3.96

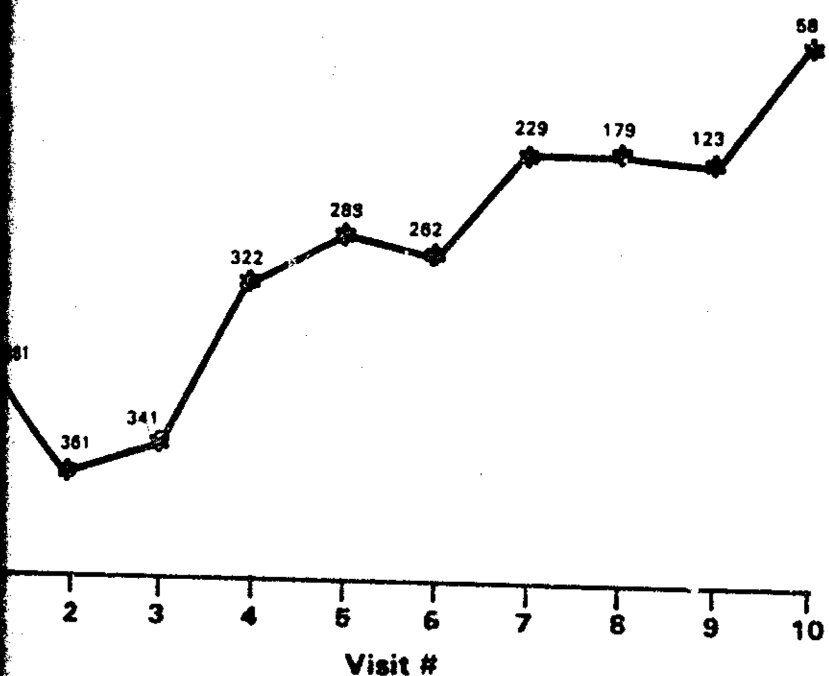
F=2.27 df=2,1301 $p > .10$ F=1.51 df=2,1301 $p > .20$

ble man-days were used, including sixth visits, N=2606, for which could be calculated.

sum of maximum consecutive time on position and the total time n.

FIGURE H

CHANGE IN TIMELOAD OVER TEN FIELD VISITS



Note:
The number above each data point is equal to the number of man-days used to compute the average timeload at each visit.

five types of sectors. The results showed a significant relationship between our simplified Arad airspace load and his airspace load calculated from complete data. Consequently we attempted to use the simplified Arad workload measure in the health change study.

The weights applied to different types of sectors, however, included a -12.84 for arrival/departure sectors because they tended to have very simple traffic patterns with many aircraft, whereas other types of sectors generally had fewer aircraft and more complex air traffic patterns. Because of this fact, the regression analysis gave a negative weight to arrival/departure sectors. The results of this negative weight on the overall calculations can be seen in Table 55.

Table 55 displays a comparison of the Arad workload across facilities in two random split-half samples of all man-days of observation. Both analyses indicated a significant difference in the Arad workload in the various facilities. However the difference seems primarily related to those facilities with many arrival/departure sectors compared to those without many arrival/departure sectors. The Boston Tower, the New York Common IFR Room, and the smaller facilities all have a predominant number of arrival/departure sectors and their Arad workload measurements were high negative numbers, whereas the centers had high positive workload measures. These results indicated that the simplified Arad workload computation probably would not be useful in the present study, since it was unacceptably biased by the weight for arrival/departure sectors.

Table 56 displays a comparison of Arad workload across age and experience groups in two random split-half samples. The results show no systematic significant differences between age groups or experience groups. In fact the pattern of results is not even repeated from one sample to the other. Again, these results suggested that the Arad workload measure was not particularly useful because it varied so much from one sample to another and in addition, was highly influenced by the proportion of arrival/departure sectors in facility. Therefore, we decided not to use the Arad measure in continuing analyses.

5) <u>Time Training</u>. Another one of our summary measures was a derived component discussed earlier. The amount of time spent in training during a day was considered both a component of workload, and a summary assessment of a particular type of work. It was used in both ways. Table 57 displays a comparison of time training across facilities in two random split-half samples of all man-days of observation. The analyses replicated one another and showed that there were significant differences in the amount

00

TABLE 55

Comparison of Arad Workload Across Facilities*

Random Split-Half Sample #1(N=1302)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
of tion Days	636	340	78	164	94
ad d ***	217.96	434.08	-803.49	-1035.16	-807.62
	1032.42	374.04	514.37	3546.78	533.28

F=43.52 df=4,1297 p < .0001

Random Split-Half Sample #2(N=1304)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
of tion Days	635	348	81	164	76
rad ad ***	239.89	454.22	-753.07	-1050.37	-759.66
	1072.68	394.59	414.10	3420.21	513.41

F=45.39 df=4,1311 p < .0001

vailable man-days were used, including sixth visits, N=2606, for which workload could be calculated.

set, Otis, Providence, Bradley

d workload - number of transitioning aircraft plus a weighted index lied to the five types of sectors.

TABLE 56

Comparison of Arad Workload
Across Age and Experience Groups*

Random Split-Half Sample #1 (N=1302)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
n Days	441	392	469	430	493	379
	-25.60	37.51	-33.36	61.49	-113.95	47.67
	2372.10	948.91	876.90	928.48	2257.68	886.91
	F=.26 df=2,1299 p>.50			F=1.81 df=2,1299 p<.20		

Random Split-Half Sample #2 (N=1304)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
Days	451	404	449	460	461	383
	39.48	23.47	-18.99	41.36	-123.97	148.83
	901.09	996.38	2229.86	888.90	2280.03	959.37
	F=.17 df=2,1301 p>.50			F=3.40 df=2,1301 p<.04		

vailable man-days were used including sixth visits, N=2606.
hich Arad workload could be calculated.
orkload = number of transitioning aircraft plus a weighted
applied to the five types of sectors.

TABLE 57

Comparison of Time Training
Across Facilities*

Random Split-Half Sample #1(N=1302)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
er of rvation Days	636	340	78	164	84
Time ning ***	1.91	.62	1.12	1.01	.68
	3.35	1.61	1.83	2.11	1.63

F=15.54 df=4,1297 p<.0001

Random Split-Half Sample #2(N=1304)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
er of rvation Days	635	348	81	164	76
Time ning ***	1.84	.47	1.42	1.14	.58
	3.30	1.31	2.00	2.46	1.70

F=16.98 df=4,1299 p<.0001

l available man-days were used, including sixth visits, N=2606, for which time
aining could be calculated.
uonset, Otis, Providence, Bradley
me training = time spent in training developmental controllers, with the units
nging from 0 to 15, representing 0 minutes to 300 minutes (all 5 hours of
servation on a study day).

of time spent training in the various facilities in the study.

Men in the Boston Air Route Traffic Control Center and in the smaller facilities spent a significantly lesser amount of time training than men in the other facilities. Men in the New York Air Route Traffic Control Center spent the most time training.

The results in Table 57 indicate that the men spent anywhere between 20 minutes and 2 hours training controllers on the job each day. Clearly, training developmental controllers on the job was an important and common task for air traffic controllers, particularly in the busier and larger facilities.

Table 58 displays a comparison of time training across age and experience groups. The oldest and the most experienced air traffic controllers spent the least amount of time in training developmental controllers.

These results were consistent and replicable and were somewhat surprising since our expectation was that older and more experienced controllers would be the ones providing training for the new developmentals. However, our results suggest that this was not true. A number of reasons could be posited for this finding. Results presented elsewhere in this study suggest that training is a difficult, arduous and disliked task for many controllers, and it may greatly tax the controller's abilities and flexibility. Further, senior controllers may see this as a task for younger controllers due to these factors. Hence older controllers may avoid training. However, we do not really know the reason for these findings.

6) Work Training. The amount of work training also was both a derived workload component and a summary measure. We calculated the amount of work while training as the simple sum of peak traffic controlled by the trainee while the observed controller was providing the training.

Table 59 displays a comparison of the amount of work while training at the various facilities in the study. Again the analyses were accomplished in two random split-half samples in order to test for replicable effects. The results show that trainees at the New York Air Route Traffic Control Center handled a significantly larger number of aircraft than trainees at the other facilities in the study. The Boston Tower and the New York Common IFR Room handled somewhat less traffic while trainees at the Boston Air Route Traffic Control Center and at smaller facilities handled the fewest number of aircraft.

Table 60 shows the comparison of the amount of work while training across the age and experience groups of the observed controllers giving the training. There was a significant difference

TABLE 58

Comparison of Time Training
Across Age and Experience Groups*

Random Split-Half Sample #1(N=1302)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
umber of bservation Days	441	392	469	430	393	379
ean Time raining **	1.51	1.62	.92	1.63	1.37	.94
.D.	2.79	3.11	2.22	2.87	2.84	2.33
	F=8.52	df=2,1299	p < .0004	F=6.55	df=2,1299	p < .002

Random Split-Half Sample #2(N=1304)

	Age (Years)			Experience (Years)		
	25-34	35-38	39-49	3-7	10-14	15-20
umber of bservation Days	451	404	449	460	461	383
ean Time raining **	1.45	1.32	1.09	1.50	1.23	1.10
.D.	2.82	2.86	2.41	2.83	2.75	2.46
	F=2.00	df=2,1301	p < .15	F=2.44	df=2,1301	p < .10

*All available man-days were used, including sixth visits, N= 2606 , for which time training could be calculated.

**Time training = time spent in training developmental controllers, with the units ranging from 0 to 15, representing 0 minutes to 300 minutes (all 5 hours of observation on a study day).

TABLE 59

Comparison of Work Training
Across Facilities*

Random Split-Half Sample #1(N=1302)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
Number of Observation Days	636	340	78	164	84
Mean Work Training***	8.50	1.78	2.50	3.35	.86
S.D.	16.52	5.07	4.02	7.65	2.39

$F=23.09$ $df=4,1297$ $p<.0001$

Random Split-Half Sample #2(N=1304)

	NYARTCC	BOSARTCC	BOSTON TOWER	NYCIFR	SMALLER FACILITIES**
Number of Observation Days	635	348	81	164	76
Mean Work Training***	8.33	1.29	3.19	3.61	1.18
S.D.	16.13	4.03	5.60	8.28	3.93

$F=23.83$ $df=4,1299$ $p<.0001$

*All available man-days were used, including sixth visits, N=2606, for which work training could be calculated.
**Quonset, Otis, Providence, Bradley.
***Work training = sum of peak traffic controlled by the trainee while the observed controller was providing the training.

TABLE 60

Comparison of Work Training
Across Age and Exerience Groups*

Random Split-Half Sample #1(N=1302)

	Age (Years) 25-34	Age (Years) 35-38	Age (Years) 39-49	Experience (Years) 3-7	Experience (Years) 10-14	Experience (Years) 15-20
on Days	441	392	469	430	493	379
	5.88	6.67	3.42	6.58	5.30	3.61
	12.57	14.91	10.07	13.93	12.88	10.29

F=8.11 df=2, 1299 p< .0006   F=5.68 df = 2, 1299 p< .004

Random Split-Half Sample #2(N=1304)

	Age (Years) 25-34	Age (Years) 35-38	Age (Years) 39-49	Experience (Years) 3-7	Experience (Years) 10-14	Experience (Years) 15-20
Days	451	404	449	460	461	383
	5.90	5.06	4.40	5.88	4.99	4.37
	12.94	12.60	11.48	12.62	13.04	11.10

F=1.67 df=2, 1301 p<.20   F =1.61 df =2, 1301 p<.20

ailable man-days were used, including sixth visits, N=2606,
hich work training could be calculated.

raining = sum of peak traffic controlled by the trainee while
served controller was providing the training.

between the age and experience groups in how much traffic their trainees controlled. One of the analyses shows that the difference was highly significant in that older and more experienced controllers did not have their trainees handle as many aircraft. The second split-half analysis showed the same general trend although it was not statistically significant. Consequently, we can say only that there was a tendency for the trainees of older and more experienced controllers to handle fewer aircraft.

The difference in the number of aircraft handled by the trainees of controllers in different age and experience groups can be explained particularly by the fact that the older and more experienced controllers did not spend much time giving training. The number of aircraft handled by the trainees may have been a function of time exposure rather than of restraint by the training controller in allowing the trainee to handle planes.

Reliability and Stability of Summary Workload Measures

The summary workload measures were all based on assessments of particular components of workload as previously discussed. Considerable effort was expended in learning how to assess peak traffic, transitioning planes, weather, and so on. Before we began the field studies, we held a large number of training sessions with the medical technicians who were to make the observations in the field. We conducted a final series of 99 observations with 4 medical technicians and the senior investigators of this study. The intraclass reliability coefficients ranged from .93 to .99 for the assessments of peak traffic, transitioning traffic, proposals, percent of weather or military blockage, sector classifications and position classifications. Consequently we can be reasonably sure that the components from which the summary measures were derived were reliably assessed.

As medical technicians joined or dropped from the study team, new technicians were trained to the same standards. No one was allowed to make formal field assessments until he had a demonstrated reliability of over .9 with the continuing technicians and in particular with the chief medical technician. Typically this training took approximately a month since a new technician had to become familiar with the terminology and the reading of the radar equipment to make the assessments correctly.

The summary workload measures themselves were based not only on the directly assessed components of workload such as the observed peak traffic but also on derived components such as maximum peak traffic, calculated by computer and therefore entirely accurate.

Another question of some concern was the stability of workload. We were interested in whether or not the amount of work performed was consistent over time for individuals.

Table 61 displays the correlations between paceload, normalized workload, timeload, time training and work training over the first three visits. Paceload and timeload were significantly related from visit to visit for individuals. Time training and work training were only marginally stable between the first two visits. Finally, normalized workload was not significantly correlated from visit to visit.

Considering the differences presented earlier between facilities in paceload and timeload, and the inconsistency of normalized workload from visit to visit, we concluded that the relative constancy of the paceload and timeload measures from visit to visit was primarily the result of the constancy of traffic conditions and time required to handle the traffic in the various facilities. The correlational results shown in Table 61 reflected the constancy of environmental conditons rather than the constancy of workload for individuals. Once we took into account the systematic differences between facilities in peak traffic and the amount of time worked, there was no association of workload from visit to visit (e.g., using normalized workload). However, some individuals were nonetheless relatively consistent.

To obtain the best and most stable measure of workload performed by individuals, we devised variables which represented the average workload across all of the field studies of a given individual. These variables gave the best estimate of the average amount of work performed by an individual even though the work performed on given days varied considerably. In later sections of this report, the reader will occasionally note variables called <u>average paceload</u>, <u>average normalized workload</u>, <u>average timeload</u>, and so forth. <u>These average measures of summary workload refer to the average across all field studies for given individuals</u>. In other sections of the report, the notation might be normalized workload visit 1, visit 2, visit 3 and so forth. In these cases the workload measures specifically refer to a given observation day and not to the entire series of field studies.

<u>Relationship Between Workload Variables</u>

Table 62 displays the correlations between the summary workload measures at the first and second field studies for blood pressure. The results were essentially the same for both visits and hence the relationships were highly reliable.

Normalized workload was significantly related to paceload and work training. It was not significantly related to timeload or time training. Hence normalized workload was most clearly related to the

TABLE 61

Stability of Summary Workload Across The First Three Field Studies

	Correlations		
Summary Workload Measure	First Study With Second	Second Study With Third	First Study With Third
Paceload [1]	.49**	.50***	.49***
Normalized Workload [2]	.10	.10	.08
Timeload [3]	.27***	.26***	.24***
Time Training [4]	.12*	.09	.08
Work Training [5]	.21**	.11	.08

* $p < .05$, 2-tailed
** $p < .01$, 2-tailed
*** $p < .001$, 2-tailed

1 N = 278 - 329

2 N = 280 - 331

3 N = 280 -331

4 N = 238 -340

5 N = 288-340

TABLE 62
Correlations between
Summary Workload Measures

First Visit, N=373

	(1)	(2)	(3)	(4)	(5)
(1) Normalized Workload	1.00				
(2) Paceload	.57***	1.00			
(3) Timeload	-.04	.52***	1.00		
(4) Time Training	.00	.20**	.24**	1.00	
(5) Work Training	.12*	.32***	.20**	.92***	1.00

* p< .05
** p< .01
*** p< .001

Second Visit, N=338

	(1)	(2)	(3)	(4)	(5)
(1) Normalized Workload	1.00				
(2) Paceload	.60**	1.00			
(3) Timeload	.05	.59***	1.00		
(4) Time Training	.05	.21**	.22**	1.00	
(5) Work Training	.26**	.44**	.20**	.84***	1.00

* p <.05
** p< .01
*** p< .001

amount of work performed and not to the amount of time a man had been sitting at a position.

On the other hand, paceload, which was based on raw peak traffic measures, was significantly related to every other measure of workload. The strongest correlation was with normalized workload, but it was as strongly correlated with timeload. Hence, paceload had two major dimensions, the amount of work as well as the amount of time being worked. Paceload also was related to time spent training and work training, both of which could be components of paceload on days that a man, in fact, did training.

Timeload was significantly related to paceload, time training and work training, but not to normalized workload. It was most highly related to paceload and had a low magnitude correlation with time training and work training.

Work training and time training were very highly correlated with one another but they were only slightly related to other measures of workload.

These correlational results indicated that the workload measures were related to one another, but they were not so highly related that any one could be used as the sole measure of workload. Rather, each of these summary measures represented some unique dimension of overall workload. Consequently, throughout the remainder of this report, these summary workload measures will be evaluated as separate indices of the amount of work a man performed.

As noted earlier, for some purposes in this study, we have used the average normalized workload, the average paceload and the average timeload for men over as many as 10 or as few as 3 field visits. These average summary workload measures provided relatively stable indices of how much work a man did on the average. Other analyses used the normalized workload, paceload and timeload for a given field visit with defined characteristics. In these instances, the actual value for the particular field visit was used.

SUMMARY

Psychological and Behavioral Measures at Work

To ascertain whether or not moods of the ATCs at the beginning of a work day were related to, or modified, work performance, behavioral or physiological response, a highly reliable instrument for mood assessment, the Profile of Mood States (POMS), was administered. Comparisons of ATC scores with male college student norms revealed that the ATCs started their work days with about as much positive mood state as the students, and with far less negative mood state. No differences in mood levels between facilities, age or experience groups were noted, except that more experienced men felt more friendliness than less experienced men. No significant linear correlations were obtained between any of eight moods and five workload measures.

To ascertain possible relationships between behavioral arousal while working, workload, and mood states, behavior was assessed at work by use of a Behavioral Rating Scale devised for the purpose. Most of the ATC behavior on the job was found to be in the low moderate range of activity and arousal. Total behavioral arousal was highly related to timeload and paceload, and very weakly related to normalized workload, time training and work training. Range of behavior was not related to work performed. Behavior during the work day was very little affected by mood states at the beginning of the day.

To ascertain the influence of perceptions of difficulty or ease of a work day on other responses, the ATCs were asked to fill out a 17-item Subjective Difficulty Questionnaire after work on the days that their behavioral and physiological responses were studied and their workload was assessed. Only two facets of the work day were considered above average in difficulty: giving on-the-job training and participating in the field study itself.

Based on the first three field studies, 70%-75% of the ATCs considered the day they were studied to be average, and 8%-11% above average, in difficulty. Perception of difficulty was moderately related to objective measures of work accomplishment.

An additional predictor variable was created: psychological response to work, which was the difference between perceived difficulty and expected difficulties based on workload assessment.

A mood of friendliness before work was related to a perception of less difficulty over the day. No over moods were found to be related to perceptions of difficulty of the work.

Total behavioral arousal at work was related to perceived overall difficulty at work. The predictive utility of these measures is reported in Section V.

8. Psychological and Behavioral Measures at Work

In addition to the objective measures of work, we were interested in certain psychological and behavioral measures. We wanted to investigate whether an ATC's mood at the beginning of the day, his behavior during the day, or subjective response to the day was related to, or modified, his propensity for physiological response, either at work or in terms of illness over the course of the study. Consequently, we included in our field studies self-reports of moods and job difficulty as well as assessments of behavior on the job by trained observers.

Moods At Work

A large number of methods and instruments for mood assessment were available in the psychological literature. One of the best validated and most reliable instruments for this purpose was the Profile of Mood States (POMS). The POMS was originally devised to assess six mood states: tension/anxiety; anger/hostility; fatigue; depression; vigor and confusion (McNair, Lorr, and Droppleman, 1971). These six mood scales were devised on the basis of many studies of patients, college students and other adults in natural as well as in experimental settings. In all of these settings, and with these various subjects, it was found that these six mood scales were consistently and significantly related to other measures of the same moods as well as to mood states caused by experimental procedures using psychotropic medication, stressful films and stressful events. In addition, these scales were repeatedly created in factor analyses of the items comprising the POMS.

At the beginning of the field studies we were advised by the senior author of the POMS (Dr. McNair) that two new scales were being used with it. These new scales assessed moods of friendliness and elation. They were experimental scales, but the author thought that they might prove to be valuable. Therefore the POMS as used in this study included eight scales, the original six and the two new experimental ones, comprising a total of 78 items. A subject indicated how he felt with respect to each item at the time of administration, according to a scale running from 0 (not at all) to 4 (extremely). The items were descriptive adjectives such as friendly, tense, happy, angry, worn-out, and so on.

In the field studies, the controller to be studied for a day was randomly selected by a medical technician. After agreeing to be studied, the controller immediately completed the POMS. In this way we sought to control for possible modifying effects of the eight moods on a man's physiological responses, workload or subjective responses to the day.

Table 63 lists the names of the eight mood scales and the reliability figures reported by the original authors as well as the test-retest reliability over a nine-month period of time for controllers in this study. The table shows that the internal consistency of the scales is quite high, indicating accuracy of assessment. In addition the table shows that the moods were variable over time, as one would expect for state variables, with the stability coefficients ranging from .38 to .61 over 9 months. This is exactly the range one expects for state measures. Thus, we were confident that the POMs was appropriately assessing mood states.

The POMS scales could have been scored according to norms established for psychiatric inpatients and outpatients or college students. We selected the norms of male college students (N=340) because they would be more similar to our non-patient group of air traffic controllers than would samples of psychiatric patients even though the college students would be approximately sixteen years younger on the average.

To score the POMS we used T scores with a mean of 50 and a standard deviation of 10. Thus a score of 50 would correspond to the average of the male college students on a given scale. A score of less than 40 would be obtained by less than 17% of the college males, and a score of greater than 60 would be obtained by less than 17% of the college males. The average range for T-scores is 40 to 60. Thus the descriptive results that follow are in comparison to college student males using T-scores.

Table 64 displays the descriptive results for the friendliness and tension/anxiety scales of the POMs on the first three field studies. The results for friendliness indicated a strong similarity to the proportion of college males who would score low, average, and very high. This similarity was also reflected in the overall mean which was quite close to the expected value of 50. These results were replicated across the first three field studies.

The same table shows that more than 80% of the ATCs experienced very low tension/anxiety at the beginning of the day. Approximately 15% to 18% experienced tension/anxiety, in the average range. Very few (less than 1%) experienced tension/anxiety in the very high range. The overall mean on each of the field study days was substantially below the normative average of 50.

Table 65 displays the POMS results for the elation and anger/hostility scales. Approximately the same proportion of controllers experienced an elated mood as of normative college students. On the other hand the controllers experienced less anger and hostility than college students. Very few were very angry at the beginning of any of the first three field study days. Again the results were very consistent across the three studies.

TABLE 63

Profile of Mood States (POMS)
Mood Names, Internal Consistency, and Stability

	Internal Consistency[1]	6-week Stability[1]	9 month Stability[2]
iness	N/A*	N/A	.61
/Anxiety	.92	.51	.41
	N/A	N/A	.56
ostility	.92	.53	.40
	.94	.45	.38
ion	.95	.49	.47
	.89	.43	.57
ion	.87	.52	.45

not available. These were new experimental scales.

McNair, Lorr, and Droppleman (1971)

64 air traffic controllers

TABLE 64

Descriptive Statistics
and
Distribution of Men Experiencing
Friendliness at the First Three Field Studies
according to the Profile of Mood States

Score	First Field Study	Second Field Study	Third Field Study
y low (< 40)	23.1%	23.4%	21.8%
rage (40-60)	65.7%	67.8	65.1%
y high (> 60)	11.2%	8.8%	13.1%
er of Subjects	385	364	344
all Mean	47.54	46.64	47.28
all S.D.	11.01	10.97	10.81

Descriptive Statistics and Distribution
of Men Experiencing Tension/Anxiety
at the First Three Field Studies
according to the Profile of Mood States

core	First Field Study	Second Field Study	Third Field Study
ow (< 40)	81.6%	81.4%	83.9%
e (40-60)	18.4%	18.3%	15.8%
igh (> 60)	.5%	.3%	.3%
of Subjects	385	366	347
Mean	33.88	34.30	33.69
S.D.	6.77	6.59	6.01

TABLE 65

Descriptive Statistics
and
Distribution of Men Experiencing
Elation at the First Three Field Studies
according to the Profile of Mood States

ore	First Field Study	Second Field Study	Third Field Study
(< 40)	18.2%	19.4%	15.9%
(40-60)	68.8%	70.8%	69.7%
h (> 60)	13.0%	9.8%	14.4%
f Subjects	384	366	347
ean	48.49	48.03	49.36
.D.	10.64	10.60	10.78

Descriptive Statistics and Distribution
of Men Experiencing Anger
at the First Three Field Studies
according to the Profile of Mood States

	First Field Study	Second Field Study	Third Field Study
(< 40)	55.6%	57.4%	54.4%
0-60)	41.8%	49.9%	42.7%
(> 60)	2.6%	2.7%	2.9%
ubjects	385	364	344
n	41.18	41.45	41.51
.	6.85	7.34	7.02

Table 66 displays the results for the POMS fatigue and depression scales. More than half of the men experienced very little fatigue at the beginning of the field study day. Approximately 40% scored in the average range, and less than 4% felt extremely fatigued at the beginning of the study days. Similarly 65% or more of the controllers experienced very little depressive mood at the beginning of the study day. Approximately 30% scored in the average range, and less than 3% had a depressed mood in the very high range. Again these results were consistent across the first three field studies.

Finally, Table 67 displays the results for the POMS vigor and confusion scales The controllers matched the college student norms rather closely with respect to vigorous mood, but scored very differently from the student norm on the conclusion scale. Ninety-five percent or more of the controllers scored in the very low range on the confusion scale. Five percent or more scored in the average range and no one scored in the very high range on the confusion scale.

These very consistent results showed that controllers beginning a work day felt reasonably good about themselves. Generally, they experienced an average amount of friendliness, elation and vigor. They experienced very little tension/anxiety, depression or confusion. A moderate number experienced average levels of anger and fatigue at the beginning of their work days. Hence, in general we can say that our field studies of physiological response to work were not confounded by extremely bad moods or extremely good moods, and further that the great majority of men began their work days in a clear state of mind without major affective states.

As one might expect with a state measure of mood there was some variability over the course of ten field studies in terms of the average level of the eight moods. Figures I-P display the changes in moods over the first ten field studies. They show, on the average, that controllers stayed within the average range for feelings of elation, friendliness, vigor, depression and anger. Although there were marked increases and decreases in these feelings over time, they stayed within the 40-60 average range. On the other hand, the controllers experienced less than average tension/anxiety, fatigue, and confusion over the entire course of the first ten field studies. These longitudinal results include variations in mood as a consequence of different controllers being studied at different times and of different work conditions over the ten field studies. By reference to Figures E-H in the previous section on Workload, it can be seen that there was no concurrent association of the controllers' moods and workload. That is, none of the mood variables showed the characteristic decrease over the first three visits and subsequent increase over the next seven visits that the workload measures showed.

TABLE 66

Descriptive Statistics
and
Distribution of Men Experiencing
Fatigue at the First Three Field Studies
according to the Profile of Mood States

Score	First Field Study	Second Field Study	Third Field Study
ery low (< 40)	53.9%	56.0%	56.5%
verage (40-60)	42.7%	40.7%	41.5%
ery high (> 60)	3.4%	3.3%	2.0%
mber of Subject	386	366	347
erall Mean	39.78	39.63	39.32
erall S.D.	7.69	7.76	7.18

Descriptive Statistics and Distribution
of Men Experiencing Depression
at the First Three Field Studies
according to the Profile of Mood States

Score	First Field Study	Second Field Study	Third Field Study
y low (< 40)	71.7%	65.8%	66.0%
rage (40-60)	27.8%	31.8%	32.6%
y high (> 60)	.5%	2.4%	1.4%
er of Subjects	385	366	347
rall Mean	40.18	40.48	40.57
rall S.D.	4.38	5.06	5.08

TABLE 67

Descriptive Statistics
and
Distribution of Men Experiencing
Vigor at the First Three Field Studies
according to the Profile of Mood States

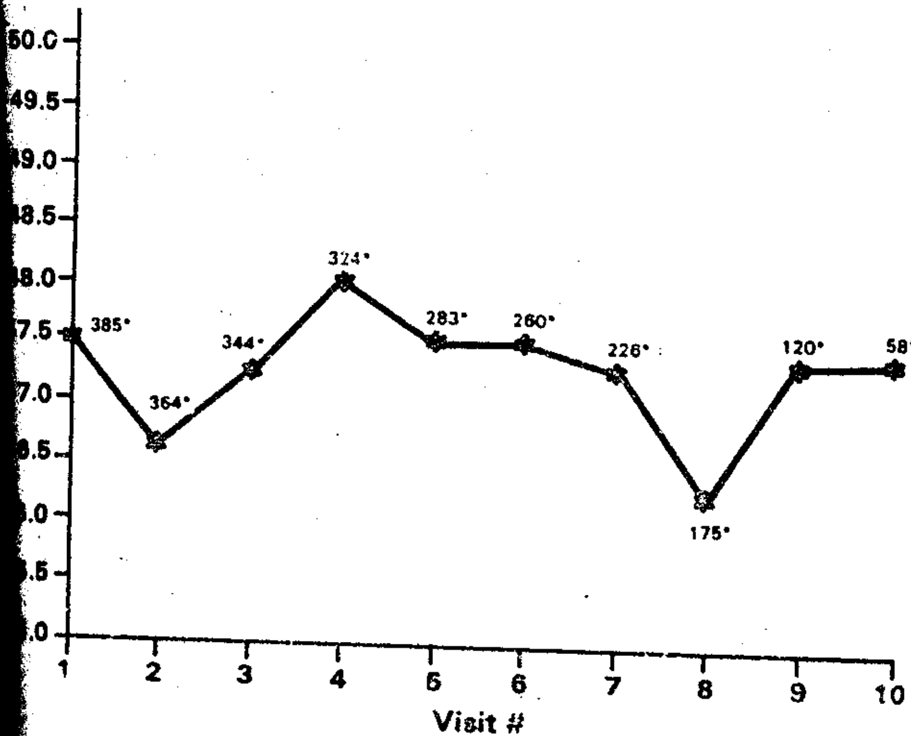
Score	First Field Study	Second Field Study	Third Feild Study
low (< 40)	15.5	15.9%	15.4%
age (40-60)	60.4%	65.1%	63.1%
high (> 60)	24.1%	19.0%	21.5%
er of Subjects	386	364	344
all Mean	51.33	50.08	50.94
all S.D.	10.57	10.37	10.56

Descriptive Statistics and Distribution
of Men Experiencing
Confusion at the First Three Field Studies
according to the Profile of Mood States

Score	First Field Study	Second Field Study	Third Field Study
low (< 40)	97.7%	94.8%	96.2%
e (40-60)	2.3%	5.2%	3.8%
igh (> 60)	0.0%	0.0%	0.0%
of Subjects	386	364	344
l Mean	28.32	28.50	28.44
l S.D.	4.60	5.12	4.60

FIGURE I

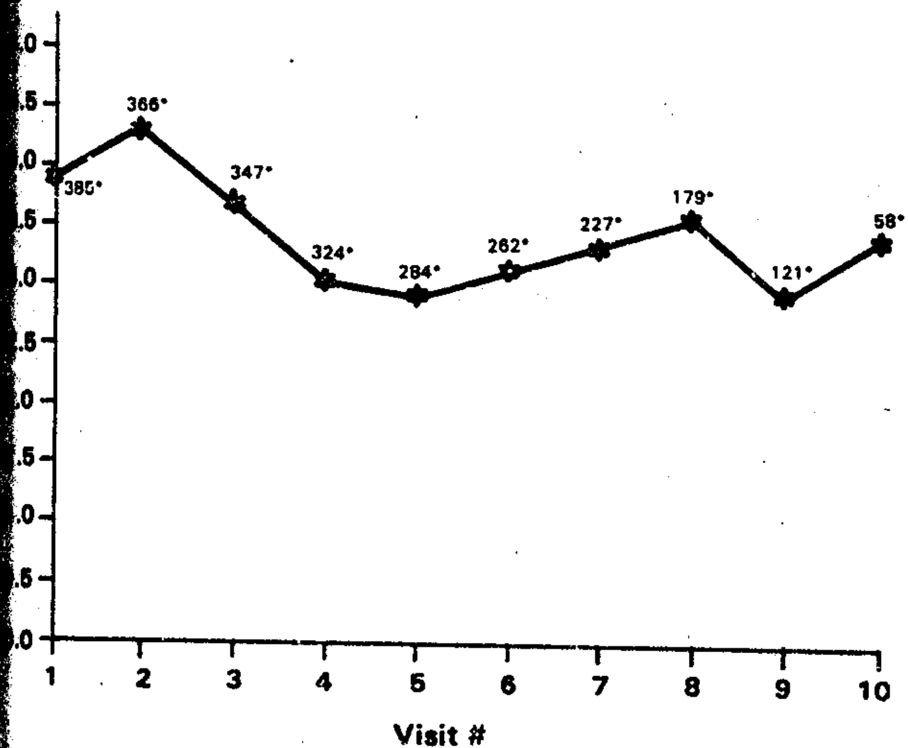
PROFILE OF MOOD STATES: FRIENDLINESS OVER 10 FIELD STUDIES



*Number of ATCs contributing to mean score at that visit

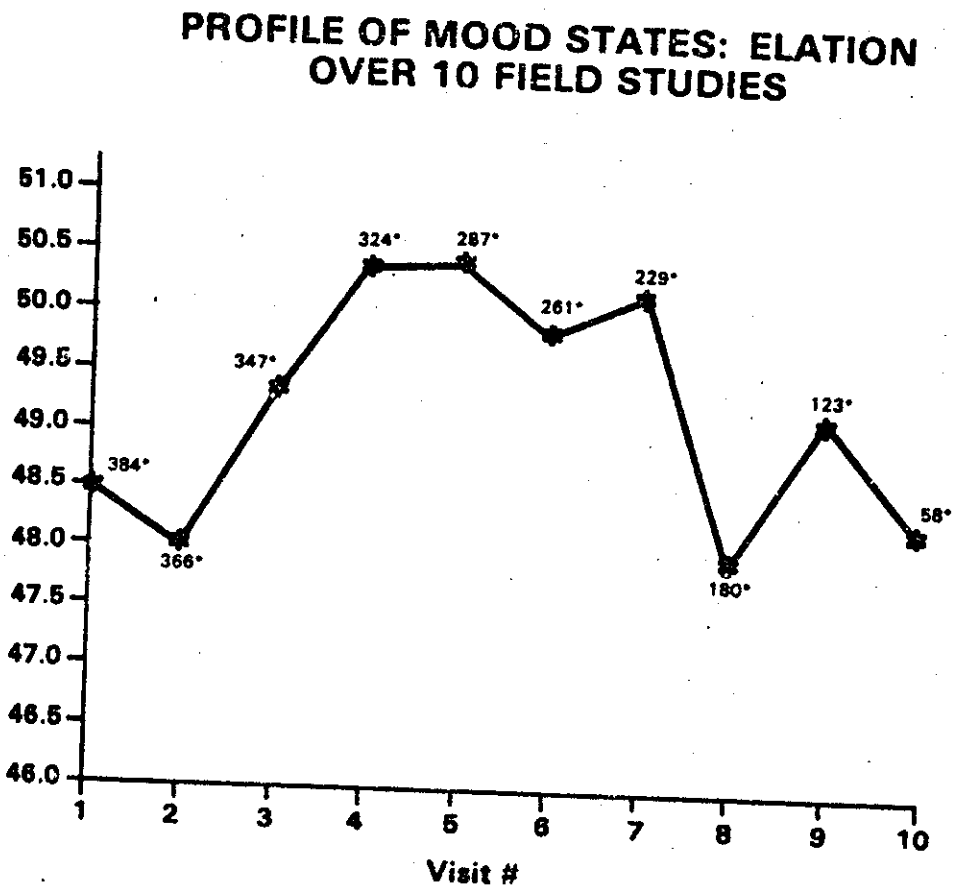
FIGURE J

PROFILE OF MOOD STATES: TENSION/ANXIETY OVER 10 FIELD STUDIES



*Number of ATCs contributing to mean score at that visit

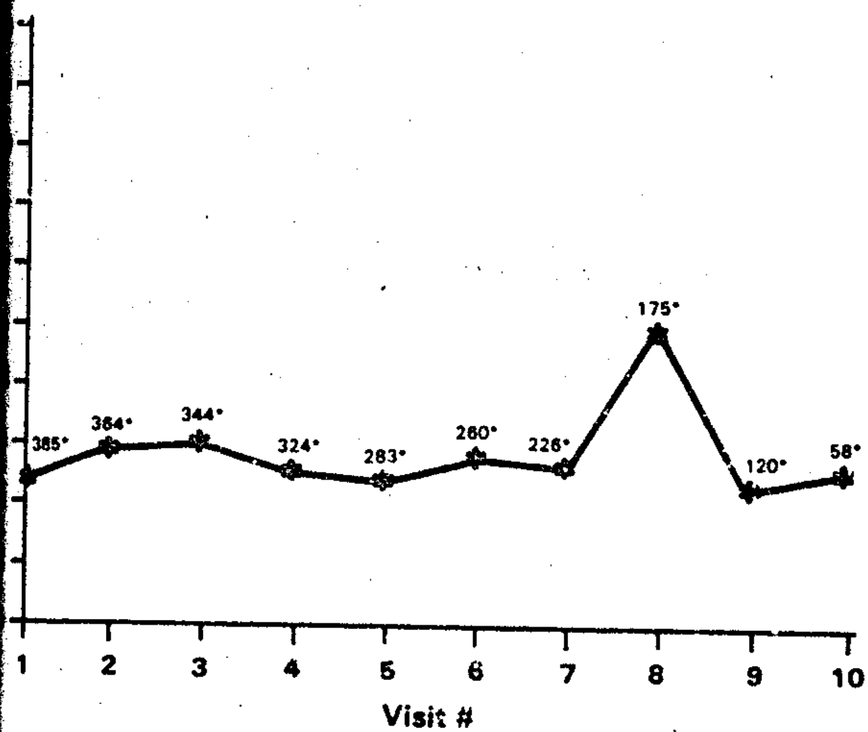
FIGURE K



*Number of ATCs contributing to mean score at that visit

FIGURE L

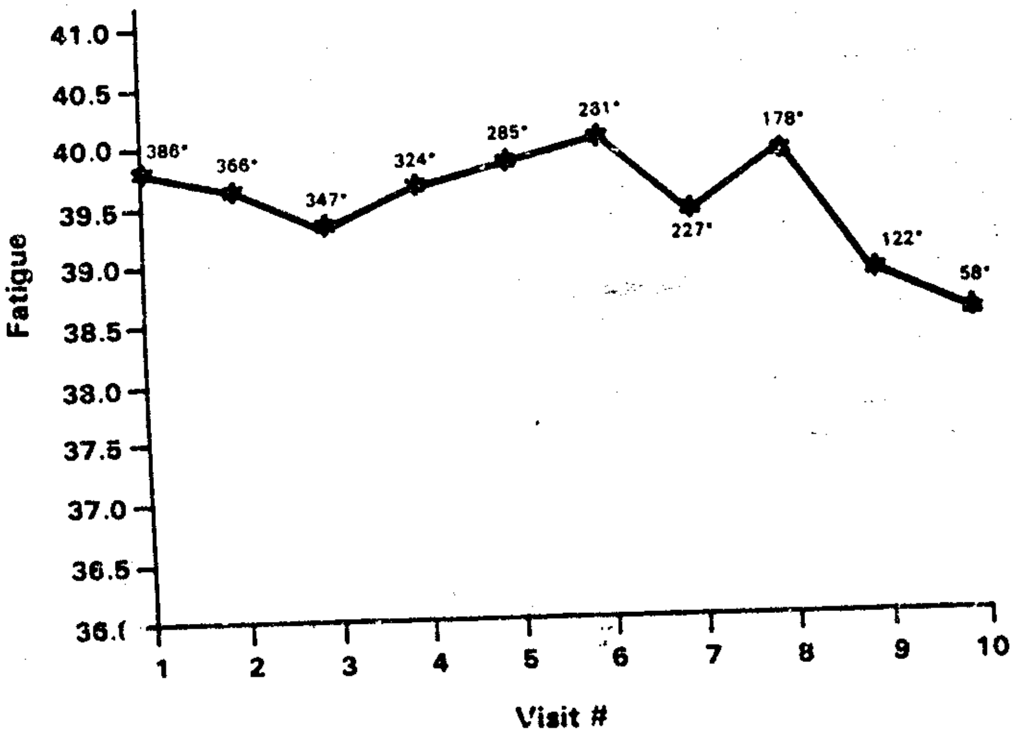
PROFILE OF MOOD STATES: ANGER OVER 10 FIELD STUDIES



*Number of ATCs contributing to mean score at that visit

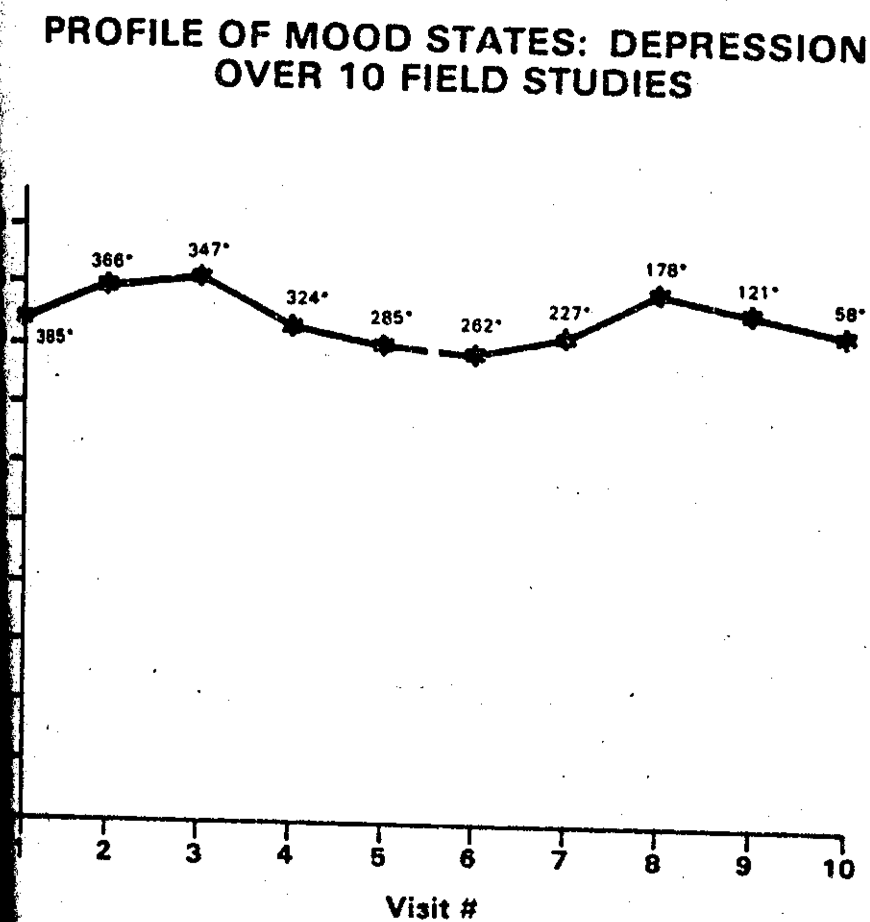
FIGURE M

PROFILE OF MOOD STATES: FATIGUE OVER 10 FIELD STUDIES



*Number of ATCs contributing to mean score at that visit

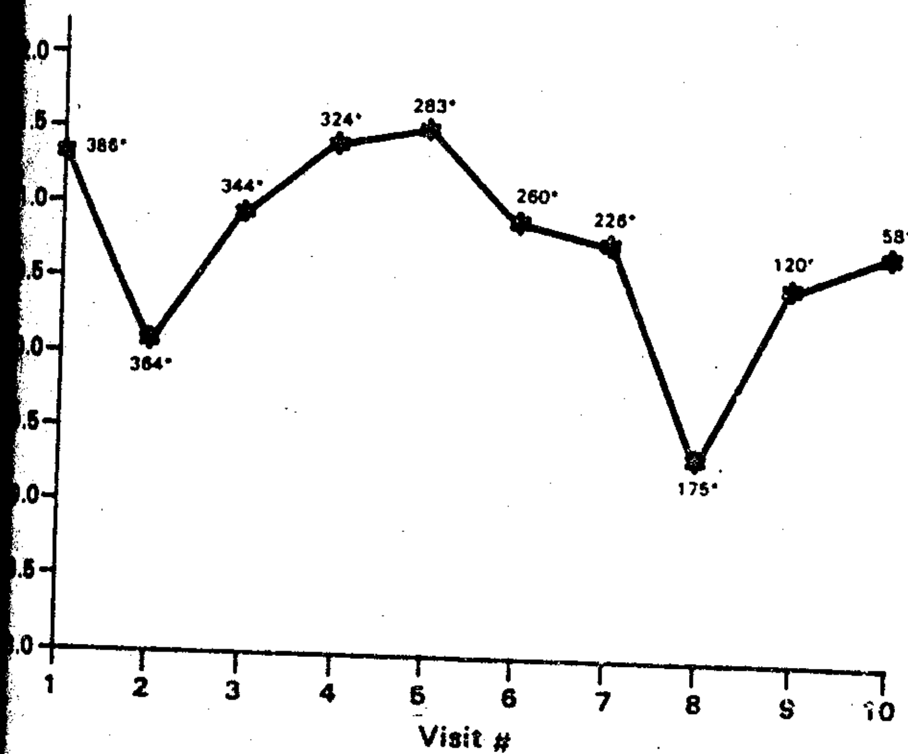
FIGURE N



*Number of ATCs contributing to mean score at that visit

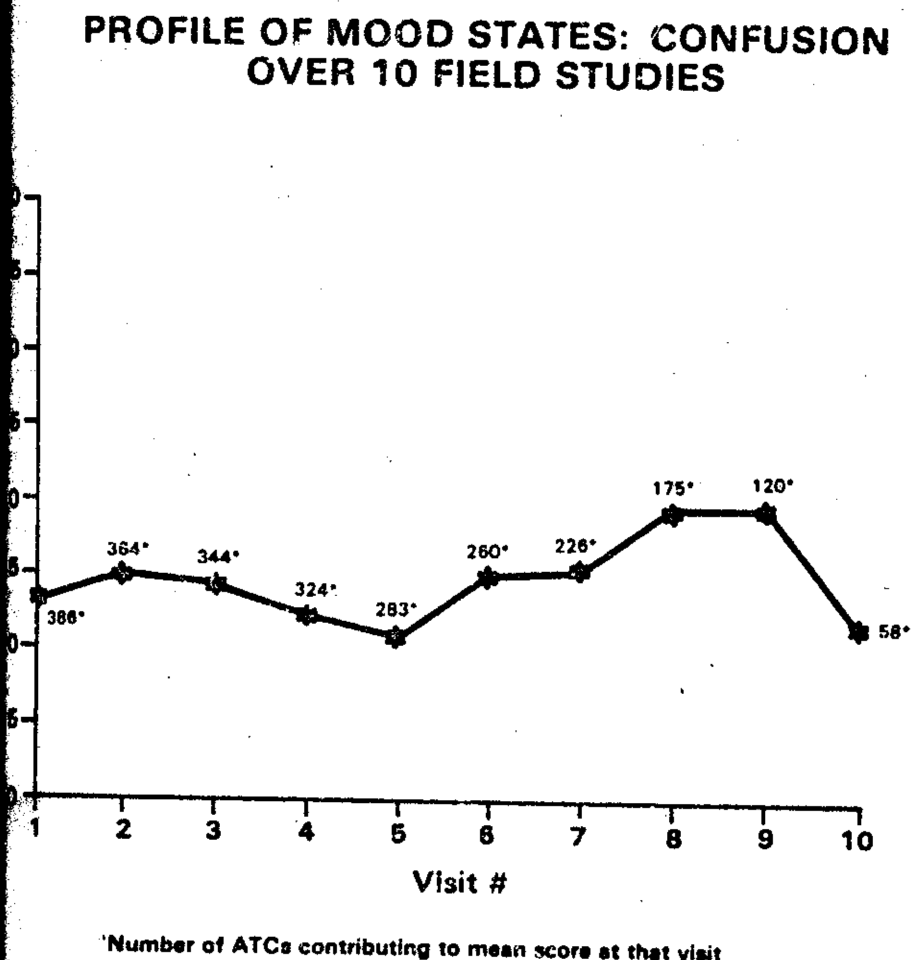
FIGURE 0

PROFILE OF MOOD STATES: VIGOR OVER 10 FIELD STUDIES



*Number of ATCs contributing to mean score at that visit

FIGURE P


PROFILE OF MOOD STATES: CONFUSION
OVER 10 FIELD STUDIES
386*
364*
344*
324*
283*
260*
226*
175*
120*
58*
1
2
3
4
5
6
7
8
9
10
Visit #
*Number of ATCs contributing to mean score at that visit

We also examined whether moods were different between facilities, age groups or experience groups. There was no consistent association between any of the moods and the facility within which a man worked, his age or his experience. There was a greater tendency for more experienced controllers to be in a friendly mood than less experienced controllers. This was the only mood associated with any of these groupings.

Table 68 displays the correlations between the eight moods and the five summary workload measures at each of the three first field studies. The table shows that there was very little, if any, significant association between any of the eight moods and the five summary workload measures. There were only five significant correlations in this entire set of results, about as many as one would expect to be significant at the .05 level among 120 correlations. Consequently we feel confident in concluding that the men's moods at the beginning of a study day had no effect on the amount of workload they subsequently performed. Results of studies on the relationship between moods at work and physiological, behavioral, and psychological responses to work are presented in later sections of the report.

In summary, by a highly reliable method of mood assessment, we determined that the controllers experienced very little negative affect at the beginning of the work days, and an average amount of positive affect at the beginning of their work days. Our results indicated that there were no differences between facilities, age groups or experience groups in the average levels of the eight moods, with the single exception that more experienced men tended to report more friendliness at the beginning of a work day. Finally, our results indicated that there was no significant linear correlation between any of the moods and our five summary workload measures.

ATC Behavior at Work

Our mood assessments at the beginning of the day were made as a prelude to assessments and measurements during the day of a controller's workload, behavioral responsivity, and physiological responsivity. In order to assess behavioral responsivity, however, we had to devise an appropriate scale of values to assign to observable behavioral response. We needed to develop a straight-forward instrument which (1) could be used easily by psychologically unsophisticated observers, (2) had a high level of face validity, (3) would be construct-valid in its relationship to physiological and psychological measures, (4) could be very reliable among multiple observers, and (5) would combine appropriate observable physiological responses with manifest behavior to indicate arousal.

An intense observation strategy was employed to provide a basis for assessing ATC behavior. Initially, a series of visits were used

TABLE 68

Correlations between Eight Moods
and Five Summary Workload
Measures at the First Three Field Visits

		Summary Workload Measure				
		Normalized Workload	Paceload	Timeload	Time Training	Work Training
endliness,	Visit 1*	.04	-.01	.03	-.01	.00
	Visit 2**	-.05	-.07	.01	-.08	-.07
	Visit 3**	-.04	-.10	-.05	-.02	-.03
sion/	Visit 1	.09	.05	.00	-.05	-.02
iety,	Visit 2	.00	.08	.08	.05	.04
	Visit 3	.03	.06	.10	-.02	-.04
tion,	Visit 1	.07	.08	.07	-.03	-.02
	Visit 2	-.04	-.03	.03	-.03	-.04
	Visit 3	.00	-.12	-.14	-.09	-.09
er,	Visit 1	-.06	-.07	-.04	-.04	-.03
	Visit 2	.03	.06	.06	-.07	-.06
	Visit 3	.06	.06	.06	.04	.01
igue,	Visit 1	.04	.04	-.02	-.04	-.01
	Visit 2	-.05	-.01	-.02	-.01	-.01
	Visit 3	-.04	.01	.05	.08	.06
ression,	Visit 1	-.02	-.10	-.13	-.04	.02
	Visit 2	.01	.00	.03	-.03	-.02
	Visit 3	-.01	.05	.10	.03	-.02
or,	Visit 1	.05	.01	.02	.02	.02
	Visit 2	-.01	-.02	.07	-.06	-.08
	Visit 3	.08	-.07	-.12	-.05	-.05
fusion,	Visit 1	.03	.02	-.08	-.05	-.01
	Visit 2	-.02	-.03	-.03	.00	.01
	Visit 3	-.03	-.01	.04	.00	-.03

Visit 1, N = 376
Visit 2, N = 352
Visit 3, N = 336

erlined values are significant at p <.05.

to allow the investigators to become familiar with the types of behavior to be seen on the job. A checklist of behaviors was then devised. The checklist used a three-point rating scale corresponding to low, moderate, and high behavioral arousal.

This initial behavioral scale was very easy to use and extremely reliable. However, we observed that the range of values and the complexity and meaning of specific behaviors was restricted. Therefore a new five-point scale was devised and field-tested by our medical technicians and the senior investigators of the study.

The five-point scale was tested in a total of 45 visits to air route traffic control centers. Approximately 200 controllers were observed working in light, moderate, and heavy air traffic conditions. The investigators observed any given controller for a half hour to three hours to provide adequate samples of behavior. The observers were provided with observation scales with spaces for recording additional behaviors not captured by the five-point scale definitions. Multiple discussions and feedback sessions yielded a seven-point scale embodied multiple dimensions of behavior and observable physiological responses.

The seven-point scale was tested in a three-day period at the New York Air Traffic Control Center with four medical technicians, two psychologists and a psychiatrist making concurrent observations. These concurrent observations permitted the evaluation of reliability. The final outcome of all of these preliminary studies was the ATCS Behavioral Rating Scale (BRS) displayed in Table 69.

A total of eight dimensions defined the interrelationship of the behavioral descriptors for each level of behavior. These eight dimensions were appearance, conversational orientation, attention, physical movement, joking/swearing, voice modulation, receiving assistance and coping with hand-off and pilot requests. Table 70 displays these dimensions with the descriptors for the highest and lowest levels of behavior for which they were associated.

To arrive at a global rating of behavior, an observer compared an ATC's behavior in a two-minute observation period with the descriptors for any given behavior level. At least one-half of the five to eight descriptors for each level had to be observed and the highest level satisfying this criterion was the overall behavioral rating assigned.

During our final checkout of the BRS, multiple observers made their ratings concurrently. The Pearson product-moment correlation coefficient of reliability between observers' ratings ranged from .85 to .98 over 69 observations of air traffic controllers. The coefficient of reliability for all combinations of raters over 99 observations was .91.

TABLE 69

ATCs Behavior Rating Scale

At least half of the behavioral descriptors in each set must be appropriate for the two-minute observation period prior to the physiological sampling and concurrent with the tabulation of other environmental events (e.g., peak traffic count, transitions, etc.). If in doubt between two or more ratings, please select the maximum applicable.

Level 1:

a.) Looks relaxed or bored or makes a comment to that effect.
b.) Jokes around a lot.
c.) Most, if not all, conversation is social in nature.
d.) Scans radar, strips, and/or airfield rarely.
e.) Physical movements are loose, spatially unrestricted, and usually unrelated to position activities.

Level 2:

a.) Looks relaxed and at ease, other than occasional fidgeting or looking around.
b.) Jokes around considerably.
c.) Much, but not most, conversation is social in nature.
d.) Scans radar, strips, and/or airfield occasionally.
e.) Physical movements are relaxed, spatially somewhat restricted, and usually restricted to position activities.

Level 3:

a.) Looks comfortable when involved in tasks.
b.) May joke around occasionally.
c.) Conversation is mostly task-related; little social conversation.
d.) Scans radar, strips, and/or airfield regularly.
e.) Physical movements are mostly restricted to position-related activities (e.g., moving "shrimpboats," marking strips, etc.).
f.) Engages in no task-related swearing.

Level 4:

a.) Conversation, if any, is generally restricted to controlling traffic; rarely initiates social conversation.
b.) Generally quiet and concentrating when not engaged in task-related communications.
c.) Physical movements are totally restricted to position-related activities.
d.) Voice is even or calm.
e.) Rarely or never jokes or swears.

Level 5:

a.) Exhibits signs of mild adverse physical arousal (such as muscular tenseness and mild perspiration).
b.) Responds to little social conversation.
c.) Generally busy and concentrating with few quiet periods.
d.) Physical movements are totally restricted to position-related activities and are somewhat constricted (i.e., quickened).
e.) Voice is stronger, more variable, and/or more clipped than normal.

TABLE 69 cont'd.

ainer momentarily overrides trainee.
gages in occasional relatively quiet, task related swearing.
arts doing partner's job despite being busy himself.

:

hibits quite obvious, but controlled, signs of adverse physical
ousal (such as tense movements, punching or otherwise striking at
ble or radar scope, sudden emotional gestures, etc.).
sponds very infrequently to social conversation.
enly refuses to accept additional traffic and/or refuses special
lot requests.
ysical movements are totally restricted to position-related ac-
vities and are considerably constricted (i.e., hurried and abrupt).
ice is much stronger, much more variable, and/or much more clipped
an usual but still restrained.
ditional staff appear on scene to observe or assist and/or trainer
kes over from traineee.
gages in considerable, easily audible, task-related swearing.

:

hibits signs of extreme adverse physical discomfort (such as
eating, becoming flushed, tremulous, etc.).
gages in absolutely no social conversation whatsoever, even when
costed.
fuses to accept additional aircraft and/or refuses even normally
ceptable pilot requests.
gages in obvious, agitated physical behavior (such as slamming fist,
rowing pencil, strip, or shrimpboat, etc.).
ice is loud and extremely variable.
quests additional help or relief and/or additional staff appear on
ene to assist and/or trainer takes over completely from trainee.
gages in loud and vehement task-inspired swearing (at aircraft, at
her ATC or supervisor, or at other person(s).

TABLE 70

Dimensions of Behavior being Rated at Highest and Lowest Levels

Overt Appearance
a. Level 1 - bored and relaxed
b. Level 7 - extreme adverse physical arousal

Conversational orientation
a. Level 1 - almost all social
b. Level 7 - all task oriented, refuses social

Attention
a. Level 1 - rarely looks at radar, strips, or airfield
b. Level 7 - completely attending to radar or airfield

Physical movements
a. Level 1 - loose, unrelated to job
b. Level 7 - agitated, aggressive behavior

Joking/Swearing
a. Level 1 - jokes around a lot
b. Level 7 - job related swearing loud and vehement

Voice quality
a. Level 1 - calm and even
b. Level 7 - loud, explosive, unrestrained

Receiving assistance
a. Level 1 - lets others do their jobs without paying attention to them
b. Level 7 - requests extra help and assistance

Coping with hand-offs and pilot requests
a. Level 1 - handles hand-offs and pilot requests without delay or strain
b. Level 7 - refuses hand-offs, refuses pilot requests even when reasonable

Finally, the psychologists and psychiatrist compared their ratings with the medical technicians. Since the psychologists and the psychiatrist had extensive training in the observation and evaluation of behavior, while the medical technicians' training was primarily that involved with the development of the behavioral rating scale, their different perspectives might have yielded substantially similar or substantially different behavior ratings. However, in a special trial covering 69 observations of controllers, the concurrent validity coefficient was .80 between the professionals and the medical technicians.

Thus we felt that the BRS was a reliable and valid assessment of air traffic controller behavior on the job. The very high levels of interrater reliability suggested that measurement error and different perspectives would account for only a very minor proportion of the differences in behavior ratings.

Table 71 shows the behavioral ratings ranged from 1 to 7 where 1 represented a completely relaxed, non-aroused condition, and 7, an extreme form of behavior with verbal and overt physiological arousal. We combined all of the observations made over the ten field studies to examine the distribution of behavior ratings in total. Table 71 displays this distribution of 25,568 behavioral ratings. Sixty-seven behavioral ratings were omitted in error, and 13,755 were not made because men were not on a working journeyman position. The results shown in Table 71 indicate that almost 12% of the observed behavior was at the minimum level. Only two observations, of 25,000, were rated at the most aroused level of behavior. Most air traffic controller behavior on the job was found to be in the moderate behavior range from 2 to 4 as defined by the BRS, suggesting low levels of arousal, concentration on the task of separating planes, but also a moderate amount of non-task activities.

Two summary measures of behavior were devised for indexing the total behavioral arousal of a man during a study day. One summary measure was the simple total of the behavioral ratings for a day - the grand total amount of behavioral arousal while working. The second summary measure was the range in behavior; that is, the maximum behavior rating minus the minimum behavior rating for the day.

In order to give a clearer picture of the meaning of the total behavior for the day, Table 72 displays a distribution of the average behavior across the first three field studies. Average behavior was defined as the total behavior rating divided by the number of times a man was observed during the day. These average behavior ratings were grouped into four descriptors.

It is immediately apparent from the table that no controller had a sustained behavior in the range from 4-4.99. Five to ten percent

TABLE 71

Distribution of Behavior Ratings over all Observations (N = 25635)

havior ting	Number of Observation	% of Observations	Cumulative % of Observations
1	3033	11.9	11.9
2	6749	26.4	38.3
3	11332	44.3	82.6
4	4129	16.1	98.7
5	304	1.2	99.9
6	19	.1	100.0
7	2	.0	100.0
otals	25568	100%	100%
al Missing	67		
al not on position	13755		
nd Total	39390		

TABLE 72

Distribution Of Average
Behavior Across The First Three Field Studies

vioral ge (Behavior Total/# of Readings)	First Field Study	Second Field Study	Third Field Study
9 (Relaxed, non-task oriented)	5.0%	6.4%	9.4%
9 (Comfortable with task oriented ity	53.3%	55.4%	51.3%
9 (Concentrating with some arousal; y task oriented)	40.2%	37.7%	38.1%
9 (Overtly aroused behaviorally erbally)	1.6%	.6%	1.2%
r of subjects	381	361	341
ll Mean	2.81	2.75	2.73
ll S.D.	.54	.55	.59

of the sustained behavior was in the range below level 2, and the great majority of behavior during the day averaged between 2 and 3.99.

The correlation of the average behavior with the total behavior for the day was .54 for the first visit, .64 for the second visit, and .65 for the third visit, indicating that the average behavior captured a substantial proportion, but not the entire amount of the variance in behavior observed for a whole day.

Table 73 displays the distribution of the range in behavior during each of the first three field studies. Between 1.6% and 3.8% of the men had no change in behavior ratings during the entire day. The modal range of behavior was two levels, such as from level 1 to level 3 behavior. Very few people experienced an intense range of behavior of four levels or more, as would be expected from the distribution of ratings already discussed. Sixty-five percent to 73% of the men experienced behavioral fluctuations of two or three levels as defined by the BRS.

Table 74 displays the correlations between the summary behavioral measures and the summary workload measures for the first three field studies. Total behavior observed for the day was most strongly related to the amount of time worked during the day, and to the paceload for the day. Relatively low but significant correlations were found between the total behavior for the day and normalized workload, time training, and work training. Thus, we concluded that the behavioral total score for the day was influenced primarily by how much time a man worked on a position, and the number of planes he had to handle.

On the other hand, behavior range was related to timeload and time training on the second visit, but it was not consistently related to these measures, nor was behavior range related to any of the other summary workload measures. The amount of fluctuation in an ATC's behavior was therefore not related to the amount of work imposed on him by his environment over the course of an entire day. This lack of a relationship was found in spite of the fact that both paceload and normalized workload had as one of their components the range in peak traffic or the range in normalized peak traffic respectively. Consequently, we concluded that the primary determinants of fluctuations in behavior were not in the workload as we assessed it in these five ways.

Not only had we hypothesized a relationship between moods and how much work a man would accomplish, but we also anticipated some possible relationship between behavior and mood. Table 75 displays the correlations between the two summary behavior measures and the eight moods over the first three field studies. The only consistent

TABLE 73

Distribution of
Behavioral Range On The Job
Across The First Three Field Studies

ral	First Field Study	Second Field Study	Third Field Study
)	1.6%	2.2%	3.8%
)	18.4%	25.9%	27.4%
(2)	40.9%	44.4%	44.2%
3)	32.6%	21.8%	20.6%
(4+)	6.5%	5.7%	3.8%
f subjects	386	367	346
Mean	2.24	2.04	1.94
S.D.	.88	.91	.89

TABLE 74

Correlations of Summary Behavior Measures with Summary Workload Measures, First Three Field Studies

ry Workload ure	Correlation with Behavior Measure at Respective Visits: Behavioral Total	Behavior Range	Number of Subjects
lized Workload, Visit 1	.18***	.05	381
lized Workload, Visit 2	.29***	.09	359
lized Workload, Visit 3	.23***	.07	340
oad, Visit 1	.68***	-.09	381
oad, Visit 2	.78***	-.04	359
oad, Visit 3	.78***	-.05	340
oad, Visit 1	.76***	.00	381
oad, Visit 2	.76***	.11*	359
oad, Visit 3	.79***	.05	340
Training, Visit 1	.15**	.13*	381
Training, Visit 2	.24***	.02	359
Training, Visit 3	.22***	.08	340
Training, Visit 1	.17***	.07	381
Training, Visit 2	.29***	-.02	359
Training, Visit 3	.26***	.03	340

5
.01
.001

TABLE :5

Correlations of Behavior
with Eight Moods
for The First Three Field Studies
(Visit 1, N=376; Visit 2, N=352; Visit 3, N=336)

	BEHAVIOR TOTAL, VISIT 1 BEHAVIOR TOTAL, VISIT 2 BEHAVIOR TOTAL, VISIT 3	BEHAVIOR RANGE, VISIT 1 BEHAVIOR RANGE, VISIT 2 BEHAVIOR RANGE, VISIT 3
iness	.00	.02
	-.06	.02
	-.07	-.00
/Anxiety	$\underline{.10}$	-.05
	$\underline{.10}$	-.04
	$\underline{.11}$	.05
	.06	.02
	-.02	-.01
	-$\underline{.13}$	-.06
ostility	-.03	-.03
	.08	-.01
	.07	.03
	-.01	-.02
	-.01	-.03
	-.00	.08
ion	-$\underline{.10}$	-.01
	$\underline{.04}$	.01
	$\underline{.11}$	.08
	.01	.04
	.01	.02
	-.07	-.08
on	-.03	-.02
	-.03	-.02
	.00	.04

ned values are significant at p <.05, two tailed.

and significant relationship was a very low positive correlation between a mood of tension/anxiety at the beginning of the day and the subsequent total amount of behavior exhibited. All other relationships between behavior and moods were non-significant or inconsistent over the three visits. Thus the mood of air traffic controllers at the beginning of the work day had very little systematic effect upon either the total behavior exhibited while working or the range that occurred in behavior.

In summary, the Behavioral Rating Scale was devised to attain a reliable assessment of the controllers' behavior on the job. Our reliability studies indicated that the instrument served this purpose very well. In general most of the behavior exhibited by air traffic controllers on the job was in the low moderate range of activity and arousal. Very few controllers experienced intense behavioral arousal. Our findings also indicated that the total behavior exhibited during the day was highly related to the amount of time a man had to work and the paceload to which he was exposed during the day. There were low magnitude correlations between the total behavior and normalized workload, time training, and work training. On the other hand, the range in behavior exhibited during the day had no consistent relationship to the amount of work performed. Finally, the behavior of an air traffic controller during a study day was affected very little by any of eight moods at the beginning of the day. In fact we found that tension/anxiety at the beginning of the day was the only mood related consistently and significantly, though weakly, to total behavior. The range in behavior over a day was not systematically affected by any of the moods.

Behavior as a response to workload and other environmental determinants will be discussed in Section V.

<u>ATC Subjective Difficulty</u>

In addition to the assessment of eight moods at the beginning of the day and behavior during the day, we assessed an air traffic controller's subjective perception of the difficulty of his work on that day. After a man completed his shift on a study day, he filled out the ATC Subjective Difficulty Questionnaire.

This questionnaire contained seventeen items that were pilot tested at the Oakland ARTCC in 1973. The items are listed in Table 76. The response choices for 9 items were ordered negatively such that a low number indicated more difficulty and eight items had their responses ordered such that a high number indicated more difficulty. The items were counterbalanced in this fashion to preclude response sets. The items with negatively ordered response categories had their scoring reversed so that higher scores represented more subjective difficulty.

TABLE 76

Questions in the
ATC Subjective Difficulty Questionnaire

Overall, how difficult do you feel your assignments have been today? (1 = very easy...7 = very difficult)

Considering all of the tasks that composed your various assignments, how good a job do you feel you have done today? (1 = best I've ever done...7 = worst I've ever done)

How heavy has been today's traffic in terms of number of aircraft handled? (1 = highest ever...7 = lightest ever)

What kind of aircraft mix have you had? (1 = totally commercial... 7 = totally military and general aviation)

How much did weather affect the ease of working traffic today? (1 = made extremely easier...7 = made extremely difficult)

Compared to other times, what was the quality of help provided by your data man and handoff man? (1 = terrific...7 = the worst ever)

How many breakdowns or serious impairments of function (quality of return, fruiting, degradation, etc.) did your radar and communications equipment have today? (1 = greatest ever...7 = fewest ever)

How many failures and impairments of function did the communications equipment on the aircraft have today? (1 = greatest ever...7 = fewest ever)

How did your supervisor contribute to your performance today? (1 = helped tremendously...7 = hindered completely)

How many interpersonal conflicts occurred for you today? (Such as those arising from FAM flights, administrative policy, and other non-ATC action) (1 = most ever...7 = fewest ever)

How many potential traffic conflicts occurred today compared to "normal" times? (1 = most ever...7 = fewest ever)

How many changes from peak to slow controlling conditions occurred today compared to "normal" times? (1 = most ever...7 = fewest ever)

How many times did you feel you were about to "go down the pipe?" (1 = most ever...7 = fewest ever)

How much did your general mood prior to coming to work affect the difficulty of your job today? (1 = made job much more difficult...7 = made job much more easy)

If you gave training (formal or informal) today, how much of a burden was it? (1 = lightened job extremely...7 = extremely burdensome)

TABLE 76 cont'd.

16. How much did the requirements of this study contribute to the difficulty of your job today? (1 = made very much easier...7 = made very difficult)

17. Overall, in your experience as an ATC at this facility, how difficult were your assignments compared to most other assignments that you might have had? (1 = very easy...7 = very difficult)

Table 77 displays a comparison of the average item ratings in the Subjective Difficulty Questionnaire for the first three field studies. A score of four represented the exact mid-point for all of the items. In general this middle category indicated that the particular component of subjective difficulty was neither more difficult than average nor less difficult than average.

Table 77 indicates that only two facets of the work day were considered above average in difficulty: giving on-the-job training and participating in the field study itself. The controllers also consistently indicated that the sector at which they worked and the number of times they felt they were losing control ("going down the pipe") were much less than average difficulty for the first three studies. Aircraft communication outages and interpersonal conflicts also were ranked at a relatively low level in contributing to the subjective difficulty experienced on the study days.

A number of statistical analyses were conducted to determine if a sub-group of items focused most clearly on the overall subjective difficulty for a day. In particular, factor analyses were conducted on random split-halves across all man days. They revealed that seven items accounted for most of the variance between all of the items in the Subjective Difficulty Questionnaire. The seven items were replicated in the split-half random samples, and therefore we were confident that they represented the core dimension of subjective difficulty for an air traffic controller.

Table 78 displays the item composition and reliability of the summary subjective difficulty score eventually used. The scoring for each item is indicated in the direction of more subjective difficulty, though on the original form half of them had their responses ordered in the opposite direction. The seven main components of subjective difficulty were the man's overall assessment of the difficulty of the assignment, the perception of how heavy air traffic had been, the perception of potential separation conflicts, the number of times a man had to cycle from peak to slow controlling conditions, how many times he felt that he was about to "go down the pipe" (lose control) having to give training and finally, perception of the difficulty of this assignment relative to all other assignments. Responses to these items were added together to form a summary subjective difficulty score.

Table 78 also displays the internal consistency reliability for the summary subjective difficulty score. For the first three administrations the internal consistency reliability ranged from .82 to .86. These reliability figures were quite

TABLE 77

Comparison of Average Item Ratings for Subjective Difficulty over First Three Field Studies

Item Content	Average at First Visit[1]	Average at Second Visit[2]	Average at Third Visit[3]
1. Overall sector difficulty	2.87	2.77	2.87
2. Self-rated competence	3.73	3.73	3.65
3. Traffic load*	3.41	3.48	3.46
4. Type of air traffic	3.82	3.87	3.90
5. Weather	3.41	3.49	3.60
6. Quality of help	3.78	3.71	3.89
7. Radar outages *	3.80	3.63	3.61
8. Aircraft comm. outages*	3.22	3.18	3.2[illegible]
9. Supervisory help	3.90	3.90	3.84
10. Interpersonal conflicts*	3.29	3.32	3.26
11. # of potential aircraft conflicts*	3.39	3.38	3.38
12. Peak to slow traffic changes *	3.59	3.51	3.62
13. Feeling loss of control*	2.68	2.68	2.69
14. Mood prior to work*	3.78 (N=106)	3.84 (N=96)	3.78 (N=96)
15. Training	4.69	4.58	4.58
16. This study	4.22	4.30	4.30
17. Overall assignment difficulty	3.56	3.29	3.34

Scoring reversed such that high scores = more difficulty.

N = 389, except as noted.

N = 362, except as noted.

N = 346, except as noted.

TABLE 78

Item Composition And Reliability Of Summary Subjective Difficulty

Item Content

Overall, how difficult do you feel your assignments have been today? (scored from 1 = very easy to 7 = very difficult)

How heavy has been today's traffic in terms of number of aircraft handled? (Scored from 1 = lightest ever to 7 = highest ever)

How many potential traffic conflicts occurred today compared to "normal" times? (Scored from 1 = fewest ever to 7 = most ever)

How many changes from peak to slow controlling conditions occurred today compared to "normal" times? (Scored from 1 = fewest ever to 7 = most ever)

How many times did you feel you were about to "go down the pipe?" (Scored from 1 = most ever to 7 = fewest ever)

If you gave training (formal or informal) today, how much of a burden was it? (Scored from 1 = lightened job extremely to 7 = extremely burdensome)

Overall, in your experience as an ATC at this facility, how difficult were your assignments compared to most other assignments that you might have had? (Scored from 1 = very easy to 7 = very difficult)

Internal Consistency Reliability*

First administration, reliability = .82
Second administration, reliability = .85
Third administration, reliability = .86

*Coefficient alpha (Nunnally, 1967)

high and consistent. Therefore we concluded that the summary subjective difficulty score was a coherent assessment of the overall difficulty experienced by a man as he reflected upon the work day he had just completed.

Table 79 displays the distribution of summary subjective difficulty scores over the first three field studies. Slightly more than 17% of the men thought that the day they were studied was much below average in difficulty while the great majority (70%-75%) thought the day was average. A small proportion (8%-11%) considered the day above average in difficulty.

We were interested in whether or not the subjective difficulty of a day was related to the objective difficulty of a day assessed by our summary workload measures. Table 80 displays the correlations between summary subjective difficulty score and the summary workload measures for the first three field studies.

Summary subjective difficulty was related most closely with the normalized workload measures, yielding moderate but highly significant correlations. Paceload also was moderately correlated with summary subjective difficulty. The correlation between paceload and summary subjective difficulty increased over subsequent visits. Timeload had only a marginally significant relationship to summary subjective difficulty. Finally, the two training workload measures had moderately low correlations with subjective difficulty at the three studies. Thus, a man's perception of the difficulty of his day was moderately related to objective measures of how much work he accomplished.

The consistent and moderate correlation between normalized workload and subjective difficulty allowed us to compute another measure of subjective response to work. This measure, called psychological response to work, was equal to the difference between the actual subjective difficulty and the subjective difficulty that would be expected from the objective workload. Hence a high score on the psychological response measure indicated that a man experienced the day as more difficult than the objective workload warranted. Men who responded in this way were classified as augmentors. On the other hand, a low score indicated that a man perceived his day to be much less difficult than would be expected from the objective normalized workload. Men who responded in this fashion were classified as reducers, inasmuch as they underestimated how much work they had done.

The psychological response to work variable was calculated from our post hoc findings, and should be differentiated

TABLE 79

Distribution of Summary Subjective Difficulty over the First Three Field Studies

mary lective Difficulty	Percentage Scoring at First Field Study	Second Field Study	Third Field Study
al Score ≤ 14 h less than average ficulty	18.5%	17.7%	17.6%
al Score 15-28 rage difficulty	70.2%	71.3%	74.6%
al Score 29+ ve average ficulty	11.3%	11.0%	7.8%
l Subjects	389	362	346
rall Mean	20.77	20.32	20.62
rall S.D.	6.37	6.23	6.19

TABLE 80

Correlations Of Summary Subjective Difficulty With Summary Workload Measures, First Three Field Studies

MMARY RKLOAD	CORRELATION WITH SUMMARY SUBJECTIVE DIFFICULTY AT RESPECTIVE VISIT NUMBER	NUMBER OF SUBJECTS
rmalized Workload, Visit 1	.44***	381
rmalized Workload, Visit 2	.42***	355
rmalized Workload, Visit 3	.43***	339
ceload, Visit 1	.33***	381
ceload, Visit 2	.37***	356
ceload, Visit 3	.43***	340
meload, Visit 1	.03	381
meload, Visit 2	.18*	356
meload, Visit 3	.18*	340
me Training, Visit 1	.27***	381
me Training, Visit 2	.27***	356
me Training, Visit 3	.31***	340
rk. Training, Visit 1	.30***	381
rk Training, Visit 2	.26***	356
rk Training, Visit 3	.33***	340

* $p < .01$
* $p < .001$

from the summary subjective difficulty. The predictive utility of the psychological response to work, as well as of subjective difficulty and objective workload are discussed in Section V.

Finally, we were interested in the relationship between moods, behavior and the summary subjective difficulty measures. Table 81 displays the correlations between the summary subjective difficulty and the summary behavioral and mood measures at the first three field studies. Only one mood was consistently, though weakly, related to the subjective difficulty reported at the end of the day: friendliness. The correlation was negative, so that greater friendliness at the beginning of the day related to reports of less difficulty during the day. No other correlations between moods and summary subjective difficulty were consistent and replicable across the first three field studies.

Table 81 also displays the correlation between the summary behavior measures and the summary subjective difficulty for a day over the first three field studies. Total behavioral arousal (total behavior) during the day was significantly and positively correlated with reported overall difficulty (summary subjective difficulty) at the end of the day. Similarly we found that the range in behavior displayed during the day was positively related to the subjective difficulty reported at the end of the day. The correlation between behavior range and summary subjective difficulty, however, was much less than that between the total behavior and subjective difficulty, indicating that fluctuations in behavior were less important in the perception of a day's difficulty than was the total amount of arousal he experienced.

An average behavior rating also was calculated for illustrative purposes. The average behavior manifested during the day was moderately and significantly related to summary subjective difficulty. This result added further credence to the suggestion that the variability in behavior was less related to experiencing difficulty for a day; rather, the total continued amount of behavioral arousal was more important to men when they recalled the difficulty of their day. Correlations of average behavior rating and subjective difficulty decreased as the field studies progressed, whereas correlations of total behavior and subjective difficulty increased. We did not, therefore, continue to use the average measure since the total behavior measure displayed more consistent and increasingly strong associations with objective and subjective workload. The results reported in Table 81 illustrate this phenomenon over the first three studies.

TABLE 81

Correlations of Summary Subjective Diffculty with Summary Behavior Measures and Eight Moods at First Three Field Studies

Profile of Mood States	Summary Subjective Difficulty Visit 1	Visit 2	Visit 3
Friendliness	-.13*	-.14*	-.11*
Tension/Anxiety	.19**	.03	.00
Elation	-.11*	-.10*	-.07
Anger/Hostility	.02	.00	.02
Fatigue	.09	.05	-.03
Depression	.04	-.01	-.01
Vigor	-.06	-.10*	.04
Confusion	.10*	-.01	.00
Summary Behavior			
Total Behavior	.26***	.34***	.36***
Behavior Range	.21**	.13**	.10*
Average Behavior	.44***	.40***	.39***

* $p < .05$
** $p < .01$
*** $p < .001$

In summary, the ATC Subjective Difficulty Questionnaire was devised to assess the components of workload which contributed to the perception of a difficult day.

The difficulty of a day as reported was moderately and significantly associated with the amount of work performed for the day. Only one mood was consistently related to subjective difficulty at the end of the day: a friendly mood in the morning was related to a report of less than average difficulty for the day. Total behavioral arousal during the day was significantly correlated with the total subjective difficulty reported by men at the end of the day. The range in behavior manifested during the day was significantly, but slightly, related to subjective difficulty. The average behavior manifested was moderately related to subjective difficulty reported by men at the end of the day.

SUMMARY

Cardiovascular Measures at Work

Levels and variability of blood pressure and heart rate at work were studied to ascertain whether or not the air traffic control work was producing increased cardiovascular response. Portable automatic recording sphygmomanometers were used to measure and record the ATCs' blood pressure every 20 minutes during a 5-hour study day while they were actually controlling traffic and while they were off position.

The following findings emerged:

1) The men's average blood pressures at work were highly correlated with their average blood pressures in the physician's office (.64 for both systolic and diastolic);
2) Daily average systolic blood pressure was 129 ± 13 mm Hg; daily average diastolic blood pressure was 87 ± 8 mm Hg; daily average heart rate was approximately 78 ± 10 beats per minute;
3) There was a slight tendency for blood pressures to be higher while men were on-position compared to when they were off-position;
4) On the average, systolic pressures ranged 35 ± 12 mm Hg over a day, and diastolic pressures ranged 27 ± 8 mm Hg;
5) There was a significant difference of 3 mm.Hg in the diastolic blood pressure averages of the oldest group of controllers, aged 39-49 at entry to the study vs. the younger controllers;
6) Working on the radar position had a consistently significant effect on range in blood pressure. Neither working other positions nor working any sector had a significant effect;
7) Controllers at Logan Tower, younger than controllers at other facilities, had highest systolic blood pressure values;
8) When each man's changes in blood pressure were calculated on the basis of his own record, on-position measures were significantly higher for most men than off-position measures;
9) On the average, level of systolic blood pressure increased 4 to 5 mm Hg and diastolic blood pressure increased 2 to 3 mm Hg under conditions of higher workload when days of varying workload were compared;
10) Variability of blood pressure was not affected by changing amounts of workload, but was influenced by time on-position, which did not influence average pressure.

For predictive studies, a number of measures of cardiovascular responsivity to work were derived in addition to measures describing blood pressure levels and variability for each man.

9. Cardiovascular Measures at Work

Introduction

The cardiovascular measures that we studied were levels of blood pressure and heart rate, and variability of blood pressure and heart rate. Systolic blood pressure, diastolic blood pressure and heart rate were used to assess the nature of men's cardiovascular responses while they were controlling air traffic as well as when they were on work breaks and carrying out other duties.

A standard protocol was followed for all observations. Systolic and diastolic blood pressures were recorded using portable automatic recording sphygmomanometers that measured and recorded pressure in the cuff and recorded Korotkoff sounds on magnetic tape during deflation of the blood pressure cuff. The sphygmomanometer was operated by a medical technician, and recordings were made every twenty minutes. During the 2-minute period just before each measurement of blood pressure, a number of other assessments were made, such as traffic conditions and men's behavior. These data were collected and recorded three times an hour for a period of five hours beginning at 7:30 during the morning shift and at 15:00 during the afternoon shift. During each 5-hour period of measurements, fifteen observations were made for each man on each day he was studied. It should be noted that men's average blood pressures while at work were highly correlated with their average blood pressures from the office visits (.64 for both systolic and diastolic pressures).

Cardiovascular Measures and Descriptive Statistics

Almost all of the men were studied during work on at least one day. Table 82 shows the number of men with complete blood pressure studies. Three hundred and eighty-two men (92% of our sample) had at least one blood pressure study. The majority of men (69% of our sample) were studied for five hours on three or more different days. Over the course of the study, we collected data during a total of 1,421 man-days and 7,105 hours of observations. In all, we collected over 21,000 blood pressure and heart rate measurements while men were at work.

Levels of systolic blood pressure, diastolic blood pressure, and heart rate were computed for each man every day he was studied at work. On the average, the systolic blood pressure measured over the whole day was 129 ± 13 (mean ± S.D.) mm Hg and the average diastolic blood pressure was 87 ± 8 mm Hg. The average heart rate measured over the whole day was approximately 78 ± 10

TABLE 82

Number Of Men Completing Blood Pressure Studies

		% of 416
1st	382	92%
2nd	340	82%
3rd	288	69%
4th	249	60%
5th	162	39%

Total Man-days	1421
Total Hours	7105
Total Number Observations	21,315

beats per minute. Tables 83 and 84 present the average systolic and the average diastolic blood pressures respectively for each of the first three visits. These tables list the mean across all men for the average pressures, the standard deviations of those averages, and the maximum and minimum average blood pressure. Additional tables presenting values for heart rate are included in Appendix I. Other tables listing similar statistics for the maximum systolic and diastolic pressures also are included in Appendix I.

Table 83 also shows that levels of blood pressure were not different when data from men receiving antihypertensive medications were included. In both groups of men, with and without those receiving antihypertensive medications, there was a slight tendency for blood pressures to be higher while men were on-position compared to when they were off-position. Furthermore, values for blood pressure for all men were slightly lower on subsequent days than during the first day they were studied. On the average across all men and across all visits, blood pressures were approximately 1 mm Hg lower each visit but the differences were not clinically significant. These minor reductions were evident in the all-day measures, the on-position measures, and the off-position measures.

The amount that blood pressure varied during the day was more impressive than the levels of blood pressure. Whereas average levels of blood pressure were within the limits of pressures recorded in other groups of subjects studied by other investigators, the ranges were much greater than we expected. Tables 85 and 86 present population statistics for systolic range and diastolic range. We calculated the range as the difference between the highest and lowest pressures observed for each man on one day. On the average, systolic range measured over the whole day was 35 $\pm$ 12 mm Hg and the average diastolic range was 27 $\pm$ 8 mm Hg. During the working day, systolic blood pressure rose on the average above 150 mm Hg and fell on the average below 115 mm Hg at least once. Diastolic blood pressure rose on the average above 100 mm Hg and fell below 75 mm Hg. This variation is of great clinical significance.

These measures of blood pressure levels and variability were calculated in three ways. First, we evaluated these measures over all assessments made for a day. These are referred to as daily measures. Secondly, these same measures of level and variability were calculated only for those times a man was working a journeyman position (see Section IIIB7).

TABLE 83

ATC

FIELD STUDY

CARDIOVASCULAR MEASURES

Average Systolic At Work

ALL SUBJECTS

AVERAGE SYSTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	130.83	129.30	128.59	12.49	12.91	13.34	176.33	185.90	174.20	101.08	99.30	96.50	382	340	288
ON POSITION	131.32	129.46	128.87	13.48	13.47	14.12	181.33	185.36	178.22	98.43	98.00	97.75	375	338	286
OFF POSITION	130.63	128.72	127.84	13.16	13.24	13.46	172.00	187.50	173.83	102.67	92.75	94.25	376	336	287

EXCLUDING SUBJECTS ON ANTI-HYPERTENSIVE MEDICATIONS

AVERAGE SYSTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	130.93	129.33	128.75	12.50	13.09	13.46	176.33	185.90	174.20	103.79	99.30	96.50	360	320	269
ON POSITION	131.41	129.44	129.00	13.51	13.68	14.26	181.33	185.36	178.22	100.91	98.00	97.75	353	318	268
OFF POSITION	130.70	128.80	128.01	13.22	13.39	13.57	172.00	187.50	173.83	102.67	92.75	94.25	354	316	268

TABLE 84

ATC

FIELD STUDY

CARDIOVASCULAR MEASURES

Average Diastolic At Work

ALL SUBJECTS

AVERAGE DIASTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	87.39	86.64	86.71	8.46	8.39	7.60	113.92	109.00	110.40	65.54	66.40	67.60	382	340	288
ON POSITION	88.32	87.15	87.16	9.06	8.62	7.97	120.00	111.00	110.10	66.11	66.30	68.17	375	338	286
OFF POSITION	86.46	85.82	85.75	9.18	9.09	8.39	115.40	114.75	112.00	55.00	65.00	64.50	376	336	287

EXCLUDING SUBJECTS ON ANTI-HYPERTENSIVE MEDICATIONS

AVERAGE DIASTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	87.45	86.67	86.77	8.47	8.4[illegible]	7.68	113.92	109.00	110.40	65.54	66.40	69.40	360	320	269
ON POSITION	88.36	87.15	87.22	9.03	8.64	8.0[illegible]	120.00	111.00	110.10	66.11	66.30	68.86	353	318	268
OFF POSITION	86.54	85.78	85.79	9.26	9.03	8.35	115.40	114.75	112.00	55.00	65.00	64.50	354	316	268

TABLE 85

ATC

FIELD STUDY

CARDIOVASCULAR MEASURES

Range Of Systolic At Work

ALL SUBJECTS

RANGE - SYSTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	37.38	34.64	33.84	12.36	11.42	12.39	100.00	91.00	75.00	11.00	13.00	11.00	382	340	288
ON POSITION	29.77	28.62	28.43	12.35	11.72	12.41	71.00	75.00	70.00	0	0	5.00	375	338	286
OFF POSITION	24.17	22.11	20.84	14.25	11.80	12.60	100.00	65.00	75.00	0	0	0	376	336	287

EXCLUDING SUBJECTS ON ANTI-HYPERTENSIVE MEDICATIONS

RANGE - SYSTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	37.45	34.77	34.07	12.43	11.56	12.53	100.00	91.00	75.00	11.00	13.00	11.00	360	320	269
ON POSITION	29.77	28.64	28.82	12.29	11.87	12.41	71.00	75.00	70.00	0	0	5.00	353	318	268
OFF POSITION	24.39	22.17	20.71	14.31	11.91	12.74	100.00	65.00	75.00	0	0	0	354	316	268

TABLE 86

ATC

FIELD STUDY

CARDIOVASCULAR MEASURES

Range Of Diastolic At Work

ALL SUBJECTS

RANGE - DIASTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	26.86	27.05	25.70	8.92	8.11	7.37	78.00	55.00	50.00	8.00	12.00	10.00	382	340	288
ON POSITION	20.93	21.89	21.61	8.94	8.65	8.11	57.00	55.00	50.00	0	0	1.00	375	338	286
OFF POSITION	17.67	18.25	15.25	9.76	9.36	9.09	65.00	55.00	47.00	0	0	0	376	336	287

EXCLUDING SUBJECTS ON ANTI-HYPERTENSIVE MEDICATIONS

RANGE - DIASTOLIC	MEAN VISIT 1	MEAN VISIT 2	MEAN VISIT 3	STD. DEVIATION VISIT 1	STD. DEVIATION VISIT 2	STD. DEVIATION VISIT 3	MAXIMUM VISIT 1	MAXIMUM VISIT 2	MAXIMUM VISIT 3	MINIMUM VISIT 1	MINIMUM VISIT 2	MINIMUM VISIT 3	N VISIT 1	N VISIT 2	N VISIT 3
DAILY	26.91	27.13	25.70	8.92	8.12	7.40	78.00	55.00	50.00	8.00	12.00	10.00	360	320	269
ON POSITION	21.05	22.02	21.66	8.71	8.67	8.19	50.00	55.00	50.00	0	0	1.00	353	318	268
OFF POSITION	17.69	18.20	15.13	9.83	9.41	9.09	65.00	55.00	47.00	0	0	0	354	316	268

These measures are referred to as calculated only for those times in a day that a man was not working a journeyman position. These times included when a man was on break, working a data position, or in a supervisory position (see Section IIIB7.). These assessments are referred to as off-position measures in our tables. We were primarily interested in men's experiences as journeyman controllers. Therefore, to investigate blood pressure responsivity to work, we used the on-position measures. We could examine thereby the possible effects of changes in traffic on blood pressure during the time a man was actually working a journeyman position.

Factors Other Than Workload Affecting Blood Pressure

Before investigating blood pressure responsivity to work-related phenomena, we checked to see whether a number of other factors affected blood pressure. First, comparisons of the morning and afternoon readings of blood pressure across the first three visits showed no consistent differences in blood pressure between the morning and afternoon readings. Thus, we could collapse across morning and afternoon visits.

Other factors checked for their effects on blood pressure measures were age and amount of experience as an air traffic controller. No consistent differences were found between experience groups. The only significant differences attributable to age were in average values for diastolic blood pressures. Table 87 shows that the group of oldest controllers, men age 39-49 at the beginning of the study, had significantly higher diastolic averages than the controllers in the younger age groups across all three visits. The difference between the oldest group and the youngest group was 3 mm Hg.

Other factors examined were the possible effects of working different journeyman positions and working in different types of sectors. Working on the radar position did have a consistently significant effect on men's range in blood pressure across the first three studies. However, there were no consistent effects of working a particular type of sector.

Significant differences between facilities for all three visits were observed. These results are summarized in Table 88. The measures that consistently discriminated between facilities were systolic range and systolic maximum. Men studied at Logan Tower had the highest values for systolic blood pressure among the controllers during two of three visits. Diastolic average and maximum, significantly different on one of three visits, were also highest at Logan Tower. Controllers at Logan were younger than controllers at other facilities.

TABLE 87

Differences in Age in Blood Pressure Response To Work

	Younger	Middle	Older		
Age	25-34	35-38	39-49		
# subjects	115	105	132		
Diastolic Averages				F	p
Visit 1	86.9	87.9	90.2	3.72	.0246
Visit 2	85.3	87.2	88.9	5.16	.0065
Visit 3	86.4	86.1	88.9	3.42	.0330

FACILITY		ISLIP	NASHUA	COMMON IFR	LOGAN TOWER	QUONSET, OTIS WINDSOR, PROV.	F	P
VISIT 1	N =	(152)	(84)	(57)	(27)	(32)		
Systolic Aver.		130.0	130.3	133.2	139.1	132.3	3.15	.0144
Systolic Range		39.3	33.8	36.2	40.2	36.9	3.27	.0115
Systolic Max.		152.2	147.2	152.3	159.8	153.9	3.29	.0115
VISIT 2	N =	(140)	(79)	(49)	(24)	(26)		
Systolic Range		34.4	33.1	37.5	39.8	31.8	2.76	.0275
Systolic Max.		147.5	146.6	153.1	154.9	144.2	2.52	.0403
Diastolic Aver.		86.4	88.8	84.4	89.3	89.2	2.99	.0188
Diastolic Max.		99.2	100.0	98.4	104.5	103.1	2.72	.0291
VISIT 3	N =	(125)	(73)	(35)	(19)	(16)		
Systolic Aver.		131.1	122.8	131.5	136.2	126.8	6.04	.0003
Systolic Range		35.1	32.1	35.7	40.1	24.8	4.29	.0026
Systolic Max.		149.3	141.2	149.7	156.7	137.7	5.45	.0005

Therefore, these high pressures cannot be explained by age. On the average, controllers at the Common IFR Room had the second highest pressures.

Interrelationships Among Cardiovascular Measures

The interrelationships among the daily cardiovascular measures are presented in Table 89. In order to determine responsiveness to variations in traffic, we derived measures to describe men's variability and level of blood pressure at work. Both range and level of blood pressure might be expected to be affected by variations in traffic. Because the range correlated highly with the standard deviation (0.94 and 0.93) the standard deviation measure was not included in our analyses. As presented in Table 89, the average systolic blood pressure correlated 0.35 with the range, and the average diastolic blood pressure correlated 0.15 with its range. While there was considerable independence between the range and the average blood pressure (indicated by low correlations), the relationships between these measures were statistically significant. Men with higher levels of blood pressure tended to have more variability in blood pressure during the course of the day.

A more accurate reflection of variability in blood pressure independent of average level was obtained by calculating a residual systolic and a residual diastolic range. The residual range measures were calculated as the difference between the actual range for a man and the range predicted from his average blood pressure. Therefore, a residual range of 4 means that a man's actual range in blood pressure was 4 mm Hg higher than that expected from his average blood pressure. Similarly, the maximum blood pressure measurements were significantly correlated with the average blood pressures (0.86 for both systolic and diastolic). Men with higher levels of blood pressure had higher maximum blood pressures. Therefore, for the purposes of analysis, we also calculated a residual maximum measure for each man which is interpreted in the same way as the residual range measures. That is, a residual maximum of 4 means that a man's maximum blood pressure exceeded that expected from his average blood pressure by 4 mm Hg.

Although blood pressure levels measured on three different days were highly correlated, ranges observed varied on different days. Table 90 shows how stable the blood pressure measures were across three studies for the men who were studied on three occasions. The average blood pressure measure was the most stable measure.

Specifically, the average correlated approximately 0.5 to

TABLE 89

Correlations Among Daily Cardiovascular Measurements Of Air Traffic Controllers At Work: First Study

N* = 358-360

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
YSTOLIC												
. Average	1.00											
.Range	.35	1.00										
.Maximum	.86	.73	1.00									
.Standard Deviation	.36	.94	.71	1.00								
ASTOLIC												
Average	.57	.08	.44	.11	1.00							
ange	.25	.26	.28	.26	.15	1.00						
aximum	.57	.23	.50	.26	.86	.54	1.00					
tandard eviation	.27	.26	.28	.28	.13	.93	.49	1.00				
T RATE												
erage	.25	.22	.26	.22	.18	.06	.17	.08	1.00			
nge	.07	.26	.19	.26	.13	-.01	.11	-.01	.35	1.00		
ximum	.23	.27	.27	.27	.21	.06	.19	.07	.89	.68	1.00	
ndard viation	.05	.18	.13	.19	.13	.03	.13	.02	.40	.89	.68	1.00

uding subjects on anti-hypertensive medications; some correlations are based on r subjects than others because for some subjects there were too few observations alculate a measure. (e.g. there is no range when only one measurement was made).

TABLE 90

Correlations Among Daily Cardiovascular Measurements Of Air Traffic Controllers At Work: Across Three Studies

N* = 256-320

	First vs Second Study	First vs Third Study	Second vs Third Study
SYSTOLIC			
Average	.59	.57	.49
Range	-.02	.14	.24
Maximum	.41	.44	.39
Standard Deviation	.03	.20	.25
DIASTOLIC			
Average	.61	.47	.52
Range	.00	.12	.04
Maximum	.46	.30	.36
Standard Deviation	.04	.15	.09
HEART RATE			
Average	.54	.47	.51
Range	.04	.10	.25
Maximum	.40	.36	.44
Standard Deviation	.08	.07	.19

*Excluding subjects on anti-hypertensive medications; some correlations are based on fewer subjects than others because for some subjects there were too few observations to calculate a measure. (e.g. There is no range when only one measurement was made). In addition, the correlations could only be calculated for subjects having a particular measure for all three studies.

0.6 across the first three studies for both systolic and diastolic pressures. But while the correlations indicated considerable stability of level of pressures, the correlation coefficients were not so high as to obviate the need for multiple studies. Maximum blood pressure also correlated significantly across studies for both systolic and diastolic (0.39 to 0.44 for systolic and 0.30 to 0.46 for diastolic). However, the maximum was a less stable measure than the average pressure. Finally, range was the least stable measure (correlations ranged from - 0.02 to 0.24 in systolic and 0.0 to 0.12 for diastolic).

Additional tables in the Appendix present the correlations among the on-position measures. Three of these tables list the correlations among all the on-position measures for Study 1, Study 2, and Study 3. The fourth table lists the correlations among the on-position measures across the three studies. The daily measures were highly correlated with the on-position measures, the average correlation coefficient being 0.95. Thus, for a given individual's values, the daily measures and the on-position measures were interchangeable. However, the on-position measures were generally more stable over time than the daily measures. This may be due to the fact that men's activities while working a journeyman position were fairly consistent. In comparison, when off-position and on break, for example, men could be walking, eating, playing sports and carrying out other activities.

Relationships Between Work and Cardiovascular Measures

Turning now to the relationship between workload and blood pressure, we saw differences between the on-position and off-position blood pressure measurements across all men. These differences were slight, but suggested that the differences would be more substantial if we compared measures within individuals. Thus each man would serve as his own control.

As shown in Table 91 we found that the on-position measures were significantly higher for most men than their off-position measures. For example, on the third visit, the maximum systolic and diastolic blood pressures were 3 and 4 mm Hg higher on-position than off-position. Similar changes were observed in heart rates. Furthermore, there was a combination effect evident in that average blood pressures decreased from visit 1 to visit 3. This decrease was much greater for the off-position averages than for the on-position averages.

Heart rate measures also were significantly different between the times when a man was on-position and the times he was off-position. However, there were a number of confounding factors that affected heart rate. Men who smoked cigarettes had higher average heart rates and higher maximum heart rates on-position than men who were former smokers or who did not smoke. On the other hand, smoking did not have a significant effect on blood pressure and therefore we concentrated our analyses on blood pressures.

The higher blood pressures recorded on-position compared to off-position required further analysis. We found that, on the average, blood pressure increased under conditions of higher workloads and we undertook extended analyses to determine the strength and sources of these effects.

Two methods were used to investigate the differences in individual responses to increasing workload. Method A included individuals whose workload fell in the highest quartile on one day and in the lowest quartile on another. That is to say, the work they performed on the high day was in the highest 25% of all work done by individuals for that visit, and on the low day, their work fell into the lowest quartile of all values for that visit. One hundred and twenty-three men fulfilled these criteria. The comparisons of their blood pressures on their high day and their low day of workload are presented in Table 92. The average and maximum systolic and the average and maximum diastolic blood pressure levels were significantly higher on the high workload day than on the low workload day. The average systolic difference on these two days was 4 to 5 mm Hg while the average difference in diastolic was 2 to 3 mm Hg. These results were all statistically significant at the .005 level or better. These data showed that, on the average, men who were exposed to very different workloads on two different days had significantly increased blood pressure responses on the high workload day.

Method B compared men's values for both blood pressure and workload to each man's own averages of blood pressure and workload across all of his studies. The grand average for workload included the amount of work done while blood samples were withdrawn. This yielded an average of 7 to 8 days that we could use to calculate each man's average workload. We then selected individuals who exceeded their workload average on one day by 10%, and on another day fell below their workload average by 10%. One hundred and sixty-one men fulfilled these criteria. Results of these comparisons are shown in Table 93. Again, the differences in blood pressures between the two days were measured by comparing the values on each day to an individual's own grand average for all blood pressure studies. This permitted a parallel analysis of both the independent and dependent variables.

TABLE 92

Comparison Of Men With Themselves When They Had Very High Workload On One Day And Very Low On Another Day
(N = 123)

METHOD A: Work Compared to Population Values

Men in the Highest Quartile of Normalized Workload

On Day One and in the Lowest Quartile on Day Two

	HIGH WORKLOAD	LOW WORKLOAD	DIFFERENCE	t	p
erage Systolic mm Hg	131.85	127.51	4.34	3.73	.0005
erage Diastolic mm Hg	88.73	85.80	2.93	4.32	.0005
ximum Systolic	147.59	141.88	5.71	3.29	.005
ximum Diastolic	99.30	96.45	2.85	3.13	.005

TABLE 93

Comparison Of Men With Themselves When They Had Very High Workload On One Day And Very Low On Another Day
(N = 161)

METHOD B: (Ipsative) Work Compared to Man's Own Average For 5 Field Studies. High Day Exceeds 10% Of Own Average Workload by 10%, Low Day Falls Below Own Average Workload by 10%

	HIGH WORKLOAD	LOW WORKLOAD	DIFFERENCE	t	p
DAY AVERAGE MINUS GRAND AVERAGE FOR ALL BP STUDIES					
Systolic BP mm Hg	+ 2.09	- 2.32	4.41	4.01	.001
Diastolic BP mm Hg	+ 1.76	- 2.41	4.17	6.69	.001
Maximum Systolic	2.64	- 1.66	4.30	2.90	.01
Maximum Diastolic	1.61	- 1.52	3.13	4.00	.001

The results were impressive whether Method A or Method B was used. The difference in systolic average blood pressure between a high workload day and a low workload day was approximately 4 mm Hg, and the difference in the maximum systolic pressure for high and low workload days also was approximately 4 mm Hg. The difference in diastolic values was 3 to 4 mm Hg. All of these results were statistically significant. Although only a subsample of the population fulfilled the criteria of having significantly different workloads on their study days, it was clear that on the average, systolic blood pressure increased for that subsample by approximately 4 to 5 mm Hg on days of increased workload compared to days of significantly less workload. There was a similar increase of 3 to 4 mm Hg in diastolic blood pressure, for both maximum and average values.

In contrast to levels of blood pressure, variability of blood pressure was not affected by changing amounts of workload. The residual ranges for systolic and diastolic blood pressures were the same under high and low workload conditions. However, residual range was influenced by time on position whereas the average pressure was not. Time-load, our summary measure, was the sum of maximum consecutive time on position and the total time on position during the course of one day. We found that the residual range measures were significantly related to the number of on-position blood pressure observations a man had for that day. For this analysis, we chose to include only the men whose measurements of range were based on more than five observations. Sixty-five men had residual range measures based on six or more on-position observations for the days when they were in the highest quartile of time-load and for the days when they were in the lowest quartile of time-load. Table 94 shows that for both systolic and diastolic pressures, men tended to exceed the amount of variability expected from their level pressures by about 3 mm Hg on extremely high time-load days. They tended to have almost 1 mm Hg less variability than expected from their level pressures on extremely low time-load days. These results were significant at or beyond the 0.05 levels.

In summary, Tables 92-94, demonstrate that blood pressure values were increased or decreased in accord with different workload conditions. Men's average and maximum pressures were significantly higher under extremely high workload conditions compared to men's average and maximum pressures under extremely low workload conditions. Further, when comparing timeload, defined as time on position summed with maximum consecutive time on position, there were substantial differences in men's blood pressure variability. These differences were all statistically significant and seem to represent a real physiological difference.

TABLE 94

Comparison Of Systolic And Diastolic Residual Range Measurements
Under Conditions Of High Temeloa𝑑 And Low Timeload
For
Days Including At Least Six Measurements On Position

ressure ement	N^2	High Timeload Mean (S.D.)	Low Timeload Mean (S.D.)	r	t**	p
c Residual	65	3.17 (9.30)	-.58 (11.57)	-.05	1.99	$p < .05$
ic Residual	66	3.12 (7.76)	-.82 (7.62)	.26	3.43	$p < .01$

d is the sum of maximum consecutive time on a working position and the
time on a working position. High timeload represents the first occurrence
top quartile of timeload over five field studies. Low timeload represents
st occurrence of the bottom quartile of timeload over five field studies.

anti-hypertensive medication during any of the first three field studies are ex-
In addition, men who never experienced the top quartile of timeload
se who never experienced the bottom quartile are excluded. Further, men
e not studied at least six times are excluded. Finally, the blood
e measurements were calculated while men were on a working position. There-
en who were never on a working position are also excluded.

l range = Actual Range minus Predicted Range given his average
ressure.

t for correlated means

Cardiovascular Predictors for Future Health Change

The next major question was whether blood pressure responsivity to work influenced future health change. The previously described results led to the definition of three variables that captured blood pressure responsivity to work. Specifically, we computed the difference in average level pressures between the high and low workload days. However, when comparing the two days, the men who had the least amount of change in blood pressure levels tended to have had higher values on the low day than the men who experienced larger increases in blood pressure levels. Therefore, to get a more meaningful value of blood pressure responsivity to work, we again calculated residual measures, this time holding the initial level on the low day constant. These measures were:

1. Difference in systolic average between high and low mormative man work days, controlling for initial level on law days.

2. Difference in diastolic average between high and low normative man work days, controlling for initial level on low days.

3. Difference in mean pressure average between high and low normative man work days, controlling for initial level on low days.

For the purposes of predicting future health change, we also devised a series of summary blood pressure measures that averaged individual values for the day across three or more blood pressure studies. To obtain characteristic and stable values for individuals, we eliminated the data for the men who had been studied only one or two times. Further, we excluded data for the men receiving any antihypertensive medications. The eight summary measures that were used as potential predictors of future health change were:

1. Grand average of systolic average for the day.
2. Grand average of diastolic average for the day.
3. Grand average of systolic maximum for the day.
4. Grand average of diastolic maximum for the day.
5. Grand average of systolic residual maximum for the day.
6. Grand average of diastolic residual maximum for the day.
7. Grand average of systolic residual range for the day.

Further discussion of the significance of our findings of blood pressure response to differences in work, in terms of its influence on development of health change can be found in Section V.

SUMMARY

Endocrine Measures at Work

Cortisol and growth hormone measurements, determined by radioimmunoassay of blood samples taken while the men were at work showed that:

1) the controllers as a group were secreting more cortisol than other groups reported in the literature;
2) there was considerable variability in cortisol secretion among and within individuals over time;
3) cortisol response did not relate consistently to any measures of work load;
4) there was a tendency for maximum levels of cortisol to be higher while men were on position;
5) levels of cortisol were not related to body weight, age, cigarette or coffee consumption;
6) other factors that may explain or relate to cortisol secretion are not yet known;
7) only 2% to 3% of blood samples taken showed significant levels of HGH secretion;
8) maximum HGH secretion during the day was not consistent within individuals;
9) maximum HGH secretion for most men was so small as to be undetectable in most analyses, confirming that most men are not responsive in HGH secretion to challenging tasks,
10) characteristics associated with the relatively few high HGH secretors were: lower weight, pulse rate and blood pressure, less psychiatric symptomatology (as measured by the PSS), less use of physical activity as a coping mechanism.

10. Endocrine Measures at Work

Among possible variables that might be studied for their ability to predict future health, we selected two endocrine measures that had been suggested in the scientific literature as responsive to stress - cortisol and growth hormone (see Section II). We hypothesized that individual differences in these hormonal responses, particularly with respect to workload, would be related to health change. Our procedures for determining the amounts of these hormones secreted by the men, and the methods of analyzing the data, are related below, together with the results.

Procedure

Over the course of the study, we collected blood for cortisol and growth hormone determinations during 1,156 participant-days. Similar to the procedure during measurement of blood pressure, we observed the work the man was doing, noted his behavior, and in addition, before the start of the work day, individuals filled out the Profile of Mood States Questionnaire and at the end of the day, the Subjective Difficulty Questionnaire. They also noted the nature of their sleep and any time off schedule in the past month, as well as any change in sleep pattern. As a consequence of differences in work schedules, 369 men, or 89% of our population, had at least one day in which blood was withdrawn, 316 of these 369 men were studied 2 days, 243 men were studied 3 days, 158 men were studied 4 days and 63 men were studied 5 days.

During each day of blood collection, a small non-thrombogenic catheter was implaced in the vein of each man's forearm, and this catheter was attached to a sigmamotor pump, permitting continuous collection of blood. Five ml samples were collected every 20 minutes to a total of 15 samples or 300 minutes of observation. This five-hour period of study began either at 0800, continuing until 1300, or at 1500, continuing to 2000. A total of 16,792 samples of blood were withdrawn from the volunteers over the course of the study, and the same number of observations were made of workload and behavior.

Blood samples were collected in heparinized tubes, and within ten minutes of being collected, they were centrifuged and the plasma was withdrawn and frozen with dry ice. All tubes to be used in a given day for an individual were labeled with a master label identified by color coding as well as numerically. In addition, a label was placed on the data sheet containing the observations for that individual for the day and another was placed in the master log next to the name of the man on whom the samples were collected. An open I.D. system was used in storing the blood samples, as no attempt was made to introduce secret I.D.s for the blood collection procedure. This was different from the procedure used with blood pressure measurements. Because of the potentially sensitive nature of the blood pressure data, a code procedure was used to mark the tapes on which the blood pressure readings were recorded.

The frozen plasma samples were kept stored in a freezer at the field facility and then transported in batches to the laboratory at Boston University where they were stored in a specially designed freezer that maintained the temperature at -25°C. This freezer was protected with dual compressors, as well as an auxiliary generator for power when the electricity was interrupted. The freezer never defrosted during the course of the study, and no samples were thawed except those removed for analysis.

Both cortisol and growth hormone were determined by radioimmunoassay. Cortisol was stripped from the corticosteroid-binding globulin (CGS) by adding 40 microliters of plasma to a methanol-burate buffer and heating for 30 minutes at 60°C. Fifty microliters of this mixture were then removed and transferred to assay tubes for binding. No further chromatography or separation of cortisol was performed, as the antibody would provide adequate specificity. A 1:2500 dilution of the antibody (made to cortisol hemisuccinate-BSA conjugate) was added to the tubes and incubated for three hours at room temperature. Separation of bound and free cortisol from the antibody was accomplished by adding 250 microliters of saturated ammonium sulphate solution and centrifuging at 3,000 rpm for 10 minutes. A 400 microliter aliquot was removed from each tube to a scintillation vial, 10 milliliter scintillation solution added, and counted to 2% statistics. The standard curve was made up in triplicate with values ranging from one to 50 nanograms per 100 milliliters with a sensitivity equivalent to detecting one nanogram per 100 milliliters.

Nonspecific binding for standard tubes ranged from 4% to 6% and for the samples, 9% to 12%. The nonspecific binding value was used to correct the counts in calculating the sample values from the standard curve.

There was excellent reproducibility of the cortisol assay. The coefficient of variation (standard deviation over the mean x 100) for duplicates, or within assay reproduction, was 2.96%. The coefficient of variation between or across assays was 6.41%.

Human growth hormone was determined by double-antibody radioimmunoassay. Samples were determined in duplicate, using 200 microliters of plasma in each determination. Purified growth hormone standard was provided by Dr. A.F. Wilhelmi from the National Pituitary Agency in the National Institute of Arthritis and Meta-

2

lic Diseases (NIAMD) (Lot #1652C). The I^{125} growth hormone s purchased from Abbott and antisera that was used was pro- ced in rabbits and used at final dilution of 1:600,000 (v:v).

The assay was carried out at room temperature for 24 hours, th each tube containing 200 microliters of the sample plasma, ,000-14,000 DPM of I^{125}-HGH, 50 microliters of growth hormone tibody, diluted with normal rabbit sera-human serum albumin-BPS ffer. Separation of bound from free HGH was accomplished by dition of a 1:20 dilution of goat anti-rabbit gamma globulin d incubation for 20 hours at room temperature. At this concen- ation, the percent binding equaled 58.5 $\pm$ 2.7 percent (n = 30).

Standards were made up in triplicate, with eight values rang- from 0 to 32 nanograms per milliliter in blank plasma, which assayed to contain 0.25 nanograms per milliliter.

Sensitivity of the assay varied between 0.25 and 0.5 nano- ms per milliliter, and the contribution from nonspecific bind- averaged between 3% and 5%. Correction was made for this in calculations.

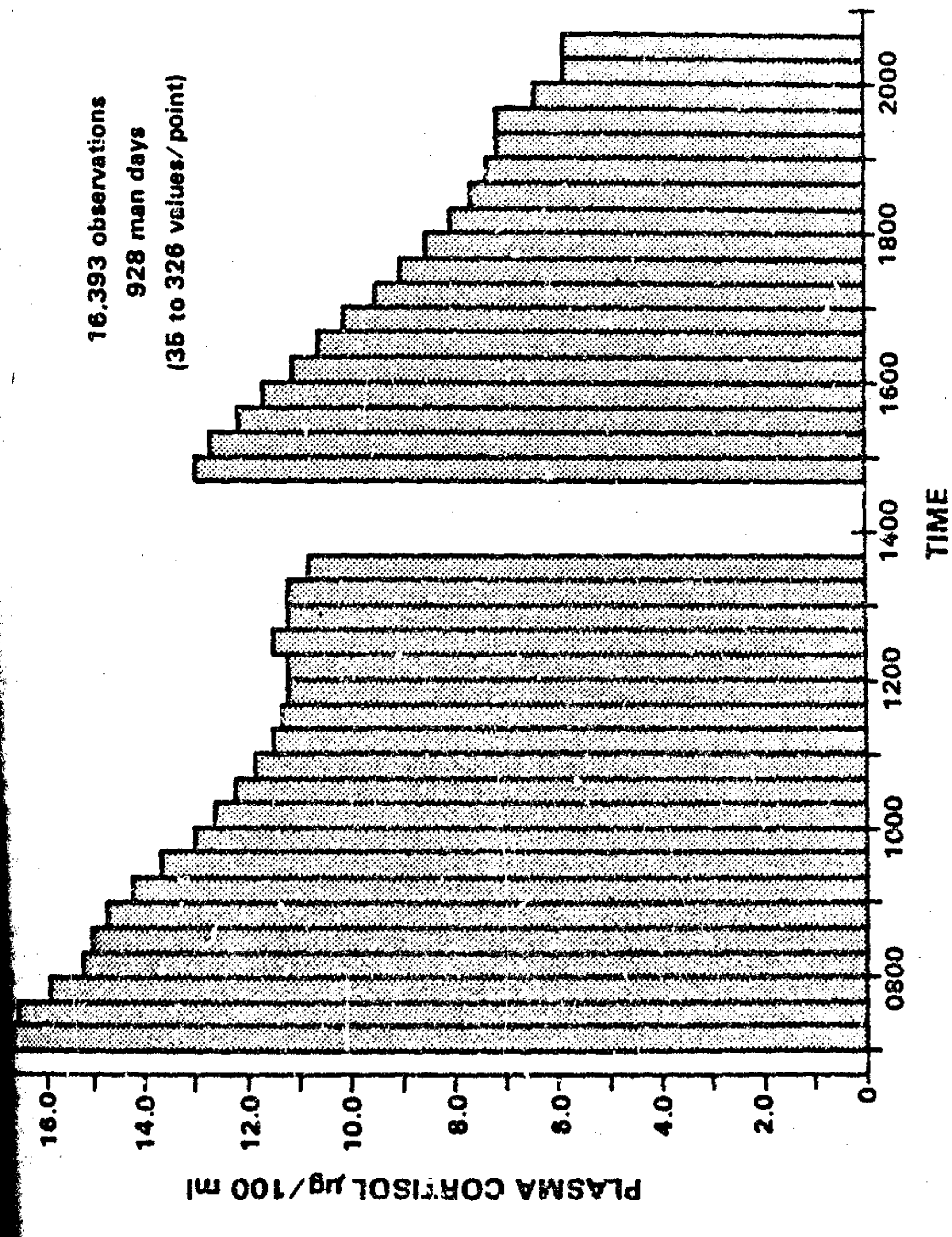
As with the cortisol assay, reproducibility of the growth mone assay was excellent. The coefficient of variation within ay (duplicates) equaled 2.20% and between assays was 2.92%.

tisol Results

The cortisol results for 12,393 values are plotted in Figure Because there was some variability in the time at which men ted both morning and afternoon collections, our data covered s from 0700 to 2040. The resulting curve was similar to that rted in the literature showing the clear diurnal trend both he morning and afternoon. Several things were noteworthy. t, the values were high in comparison with those reported in literature. The values averaged 5 to 7 µg/100 ml than values rted by other investigators using instantaneous collections. , in comparison with either values of integrated collections alues of instantaneous collections, the controllers as a group secreting significantly more cortisol during the morning and afternoon than other groups reported in the literature. These r groups, however, were largely patients in hospitals or others differed from our ATC group in that blood collections were not while the other groups were actively working.

We evaluated the effect of venipuncture by comparing the morn- nd afternoon samples of 214 men who were studied at both times. here was some variability in the exact time that people started, mpared the samples collected during the first hour, i.e., es 2-4 with those samples collected during the second hour of

FIGURE Q


16,393 observations
928 man days
(35 to 326 values/point)
PLASMA CORTISOL µg/100 ml
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0
0800
1000
1200
1400
1600
1800
2000
TIME

observation, i.e., samples 5-7. Sample I was eliminated from this analysis as it was an instantaneous sample obtained when the catheter was first inserted and represented only the value of cortisol at that particular moment rather than the integrated value over 20 minutes. In all cases the samples collected during the first hour were higher than the samples collected during the second hour with time of day held constant. These data suggested, as was reported in the literature, that there was a moderate effect of venipuncture apparent in the first hour, producing an average increase in cortisol from 1 to 2 μg/100 ml.

Selection of Measures for Analysis

As noted, samples 2-15 were similar in that they represented values obtained for an integrated 20-minute sample, while the first sample was an instantaneous value. Therefore, for all analyses using cortisol results, we eliminated the first sample. We also wanted to obtain a measure of daily cortisol for an individual free of the influence of venipuncture present during the first hour of sample collection. However, because the venipuncture effect was of relatively small value and was not a constant across all individuals, we also chose another measure which would not eliminate values obtained during the first hour.

Four measures were chosen to reflect cortisol secretion during the day. The first measure was the total amount of cortisol secreted, the sum of all values for samples 2-15. Individuals with less than 11 samples for a day were dropped from the analysis, leaving a total of 1,106 man-days of studies. More than 90% of the men remaining in the analysis had all 15 samples during the day they were studied.

The second measure calculated for each man-day was related to whether or not an episode was observed. An episode began, by definition, with a value that exceeded the previous value by 2 μg/100 ml. This was a conservative estimate as previous investigators used the 2 μg/100 ml cutoff with instantaneous samples. When a 20-minute sample differs from a previous 20-minute sample by 2 μg, it is highly likely that sometime during that 20-minute period, the cortisol may have increased by as much as 6 or 8 μg/100 ml. Episodes were calculated from samples 5 to 15, eliminating the first four samples in order to free the analysis of any influence of venipuncture. An episode ended when a subsequent sample was equal to or lower than the sample defining the start of the episode.

The third measure of cortisol was the number of peaks that occurred during the day, and these ranged from zero to three with an average of approximately .8 peaks for all samples observed.

The fourth measure was the maximum value of cortisol observed either during the entire 5 hours or while a man was on position.

Normalization of Cortisol Data

One of our major goals in collecting the cortisol data was to be able to divide men into responders and non-responders according to cortisol values. We planned to compare men in terms of the magnitude of cortisol response to varying workload over the numerous times we studied them. We were able to study 243 men on three different occasions during the three years they were in the study. Thus, 58% of the men could be assessed in terms of cortisol responsivity using three days of data and 316 or 76% of the men could be assessed on the basis of two days of data. However, since men were studied both in the morning and the afternoon, it was essential that we find a way to combine the results for these two periods of time. It is now well established that cortisol exhibits a very strong diurnal trend, with values in the afternoon averaging approximately 4 to 5 μg/100 ml lower. This effect is so strong that one finds a small but significant correlation between the exact time in which a study was started in the morning or afternoon and the total amount of cortisol for that day. It was therefore decided to compare every cortisol value determined for each individual with the average of all cortisols observed for that exact period of time. The individual value was thus normalized with respect to the average for that exact period of time of the day. A total of 13,480 cortisol values were used in 46 different time blocks to calculate normalized values. The results were expressed in terms of deviation from 100 such that if the normalized value equaled 100, it was equivalent to the raw mean of that period of time. We therefore could combine morning and afternoon visits, controlling for the different times during the day when plasma was collected for cortisol determination.

Relationships Between Cortisol Measures

The four measurements of cortisol responsivity during the day were significantly correlated with one another. The average correlation between total area and peak area for visits 1, 2 and 3 was .49. The average correlation between total area and number of peaks was .29, and the average correlation between the number of peaks and the peak area was .62. Maximum cortisol also showed an average correlation of .84 with total cortisol and .50 with peak cortisol.

Despite the consistency within a given visit between the four measures of cortisol response, there was relatively low reliability across visits. The only measure of cortisol response that had a significant between-visits correlation was that for normalized total cortisol. Neither peak area nor number of peaks were consistent between visits. The correlation of normalized total area from visits 1 to 5 ranged from a low of .23 to a high of .40 with an average correlation across all 5 visits of .29. However, the first visit did correlate .71 with the average of 3 visits, and visits that were closer to one another in time were generally more highly correlated. Cortisol response on any single visit was obviously not a trait phenomenon, and it was therefore decided to average the cortisol values to obtain a more stable or characteristic value for each man.

Cortisol Responses as Predictors

We decided to include only those individuals who had at least three studies in the field involving blood collection and for whom we had at least eleven of a total of fifteen possible samples during the five hours they were studied. These criteria were fulfilled by 231 men. As many of these same men had been studied four or five times, the average number of studies performed on these men was 3.9, approaching a more stable or valid picture of the cortisol responses at work for these subjects.

We assessed three different measures of cortisol response for these 231 men: average normalized total cortisol; average normalized peak cortisol; average normalized maximum cortisol while on position. Table 95 divides the men into low, middle and high groups and translates the normalized values into raw cortisol levels for AM and PM separately for total and maximum cortisol measures, and Figure R shows high, middle and low groups throughout the day. As noted, values tended to be higher than reported in the literature, especially since these values represented a 20-minute average level rather than instantaneous values.

The table also translates total cortisol into an average value for the AM or PM to facilitate comparison with published data. This was done by simply dividing total for AM or for PM by the average number of samples (12).

It is apparent that there was considerable variation in average daily cortisol output among the men studied. The high AM group range was 13.9 to 21 μg/100 ml for the five hours studied, which represents very considerable cortisol secretory levels, as does the high PM group range from 10.2 to 17 μg/100 ml. Similar variation is seen for maximum cortisol on position with impressively high levels seen for the high AM and PM groups. Also, as indicated in Figure S and T, the differences between high, middle and low groups were not due to differences in the response to venipuncture. Levels during the first hour, and the slope of the curves do not differ significantly in visits 1, 2 and 3 as they would if the groups had differing responses to venipuncture.

TABLE 95

Average Total and Maximal Cortisol Across All Visits For Men Who Had Three or More Studies

 age for Day

	Group Mean	Low Group	Middle Group	High Group
alized value	100	77-95	96-103	104-140
N.	231	77	77	77
Cortisol (µg/100ml)				
tudy	13.0	8.7-11.9	12.1-13.7	13.9-21.8
tudy	9.4	4.1 - 8.2	8.6- 9.8	10.2-17.0
um Value sition				
lized	120	88-112	113-125	126-180
ortisol (µg/100ml)				
udy	16.5	10.3-15.8	15.8-17.8	18.3-29.0
udy	13.1	6.7-11.0	11.6-13.1	13.5-25.1

AVERAGE PLASMA CORTISOL AT WORK

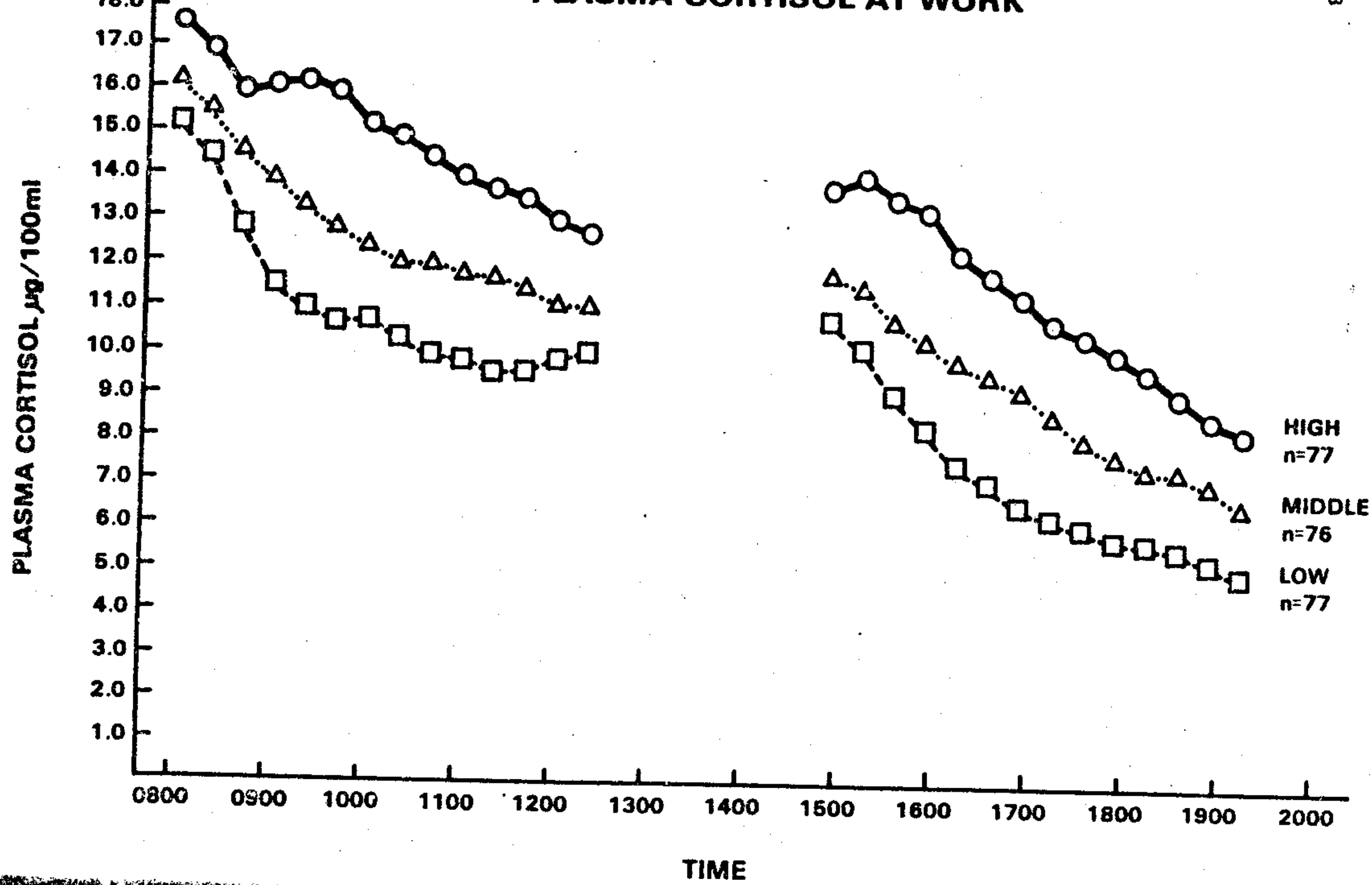


FIGURE 5

HIGH VS. LOW CORTISOL RESPONDERS
A.M.

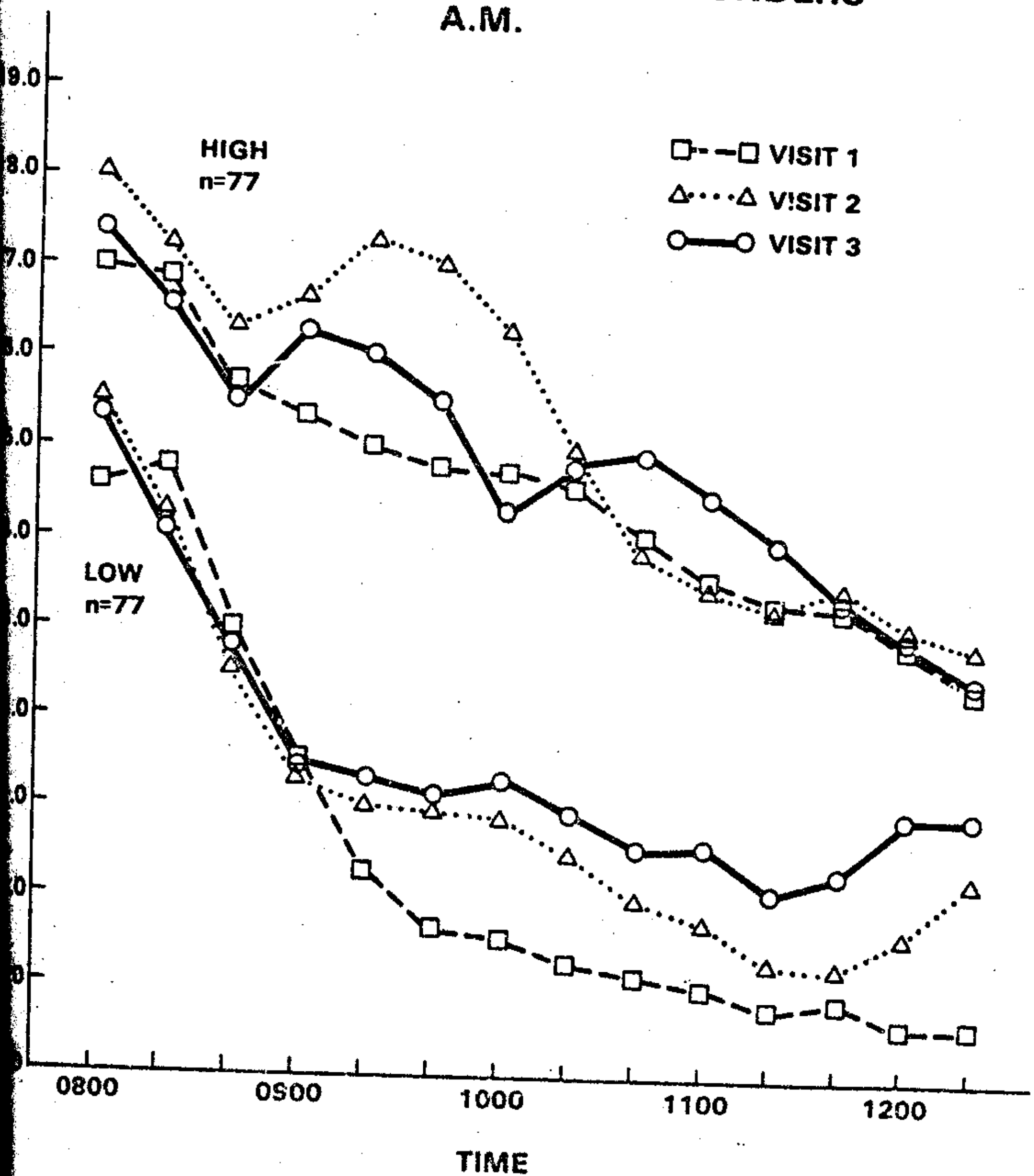
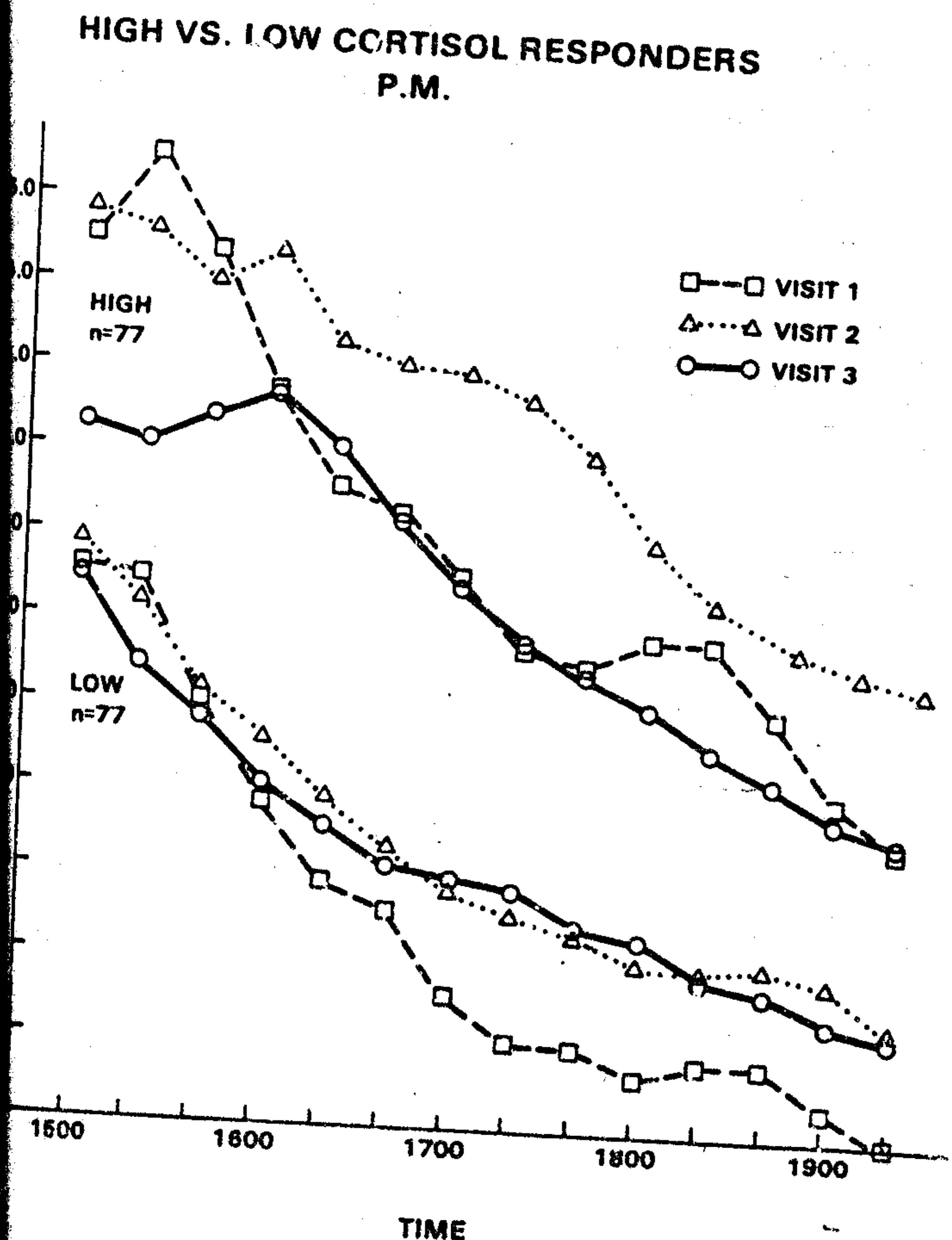


FIGURE T


HIGH VS. LOW CORTISOL RESPONDERS
P.M.
HIGH
n=77
LOW
n=77
VISIT 1
VISIT 2
VISIT 3
1500
1600
1700
1800
1900
TIME

The average total, peak and maximum cortisol values for these 231 men were studied as potential predictors or correlates of psychological and physical health change. These studies are reported in Section V.

Cortisol Levels and Workload

Extensive analyses were performed to assess any relationship between cortisol and various measures of work (e.g., number of peak planes, range of peak planes, time on position, etc.). Initially each visit was studied separately and correlations were calculated between a man's cortisol responses and various measures of workload. No consistent relationships across the three visits were found between cortisol levels and various measures of work. This inconsistency reflected in turn the separate inconsistencies of each of the variables being correlated, i.e., men neither showed a consistent level of cortisol over time nor were they particularly consistent across studies in the number of planes they controlled or the total time spent on position. The nature of work performed is discussed in Section III B 7, Measurement of ATC Workload.

In order to control for differences across visits, we also analyzed the relationship between work and cortisol within a visit. As shown in Table 96 the maximum cortisol while on position was significantly higher than the value off position for all three visits. Off position included data position as well as being on break or assigned to other details for a short time. On position included radar, handoff, training, check code monitor. Although statistically highly significant, the actual difference in cortisol levels between on versus off position averaged only five normalized units or approximately 1 µg/100 ml of cortisol. However, as is discussed in the predictive section, men showed significantly different responses to increased workload.

Total cortisol for the day, peak cortisol or number of secretory episodes were not significantly higher while on position than while off position.

We also were able to compare men on two different visits with respect to varying workload. We selected 90 whose workload on one day was in the lowest quartile and on another day was in the highest quartile. This method of selection maximized the opportunity for observing any influence of changing workload on cortisol. The results are summarized in Table 97. Maximum cortisol was significantly higher during high workload. The difference in actual cortisol levels between high and low workload days was 1.44 µg/100 ml (7.20 normalized units). However, total cortisol or peak cortisol were not significantly higher on high workload days.

TABLE 96

Comparison of Cortisol Levels When Working On Position Versus Off Position

Maximum Cortisol Observed, Normalized (controlling) For Time Of Day

N = 227

	Visit 1	Visit 2	Visit 3
ax. On Position	117.76	120.71	120.33
ax. Off Position	113.47	117.23	113.07

Average Maximum On Position	119.60
Off Position	114.59
Diff	5.01

Way Analysis of Variance / Repeated Measures	On vs. Off	$F = 38.51$ (df 2,226)	$p < .0001$

TABLE 97

Comparison Of Maximum Cortisol Response In Men During Very High And Very Low Workload

N = 90

	Highest Quartile Workload	Lowest Quartile Workload	Difference
Maximum Cortisol (Normalized)	125.54	118.34	7.20

One-way Analysis of Variance for Repeated Measures

F - 4.52 (df = 1,88) $p < .05$

We were surprised that cortisol levels were not more directly related to workload. Several qualifiers must be added to this point. We have focused our efforts primarily at whole day analyses, with the exception of comparing on versus off position within a given visit. These analyses did support the conclusion that maximum levels were higher while working. However, the effect was not large. We were not able within our report deadline to analyze more carefully within a day, for example, whether or not cortisol rises following a major rise in traffic. We also were not able to control for mood at the beginning of work, amount of sleep in recent nights, or subjective responses to the day's work. Nevertheless, a rather unexpected finding emerged and is discussed under predictors of mild and moderate illness. Men who consistently showed lower cortisol responses to work had significantly more illness than those with middle or high cortisol levels.

In summary, it is apparent that on the average our participants showed very substantial differences among themselves with respect to total daily cortisol, peak cortisol and maximum cortisol levels while working. Levels of cortisol were not related to body weight, age, cigarette or coffee consumption. We found that there was a tendency for maximum levels to be higher while men were on position. Cortisol responses at work as a possible predictor of health change were examined by comparing the frequency of illness among those with low, middle and high cortisol levels when we studied them at work. This comparison is reported in Section V.

Human Growth Hormone (HGH) Results

All blood samples from the first and second rounds of blood collections in the field were also analyzed for their growth hormone content, using the laboratory methods already described. In the first round of studies, 368 men participated, yielding a total of 5,349 blood samples whose HGH content was determined. In the second round of studies 299 men participated for a five-hour day, with blood samples taken every 20 minutes. A total of 4,361 of these samples were analyzed for HGH content. As the sensitivity of the laboratory assays varied, but was assured at the two nanograms per milliliter level, we decided to consider all HGH values at or below 2 mg/ml as a single category. In this way, we avoided the possible errors introduced in attempting to distinguish between very small, clinical unimportant levels of HGH. For many of the growth hormone data analyses, four levels of secretory activity were identified: 0 to 2 ng/ml, 2.1 to 4.9 ng/ml, 5.0 to 9.9 ng/ml, and 10.0 or greater ng/ml.

The distribution of growth hormone levels for all plasma samples of visits one and two are displayed in the table below:

Distribution of HGH Levels for All Plasma Samples of Visits 1 and 2

Visit	Less than 2.0	2.1 to 4.9	5.0 to 9.9	10.0 plus
1	91.1%	6.2%	2.0%	0.8%
2	93.9%	3.8%	1.8%	0.5%

It can be readily seen that the vast majority of blood samples contained minimal or even undetectable levels of growth hormone. Only 2% to 3% of all samples showed levels of 5.0 ng/ml or greater, the threshold commonly used in the growth hormone literature to indicate a significant level of secretion.

The following table shows the maximum level of growth hormone experienced by the men in the course of a study day:

Percent of Men at Different Levels of Maximum Daily HGH Secretion

Visit	Less than 2.0	2.1 to 4.9	5.0 to 9.9	10.0 plus
1 (368 men)	60.1%	22.8%	11.3%	5.7%
2 (299 men)	68.6%	16.4%	10.9%	4.0%

Although 90% of the growth hormone samples had concentrations of 2.0 ng/ml or less, only 60% to 70% of the men completed an entire five-hour shift without having a value above that level. There was also a tendency for higher HGH values on Round 1 than on Round 2. This was consistent with the hypothesis that anticipated pain or discomfort, or simply the novelty of the initial day of blood sampling created stresses that may have raised HGH levels.

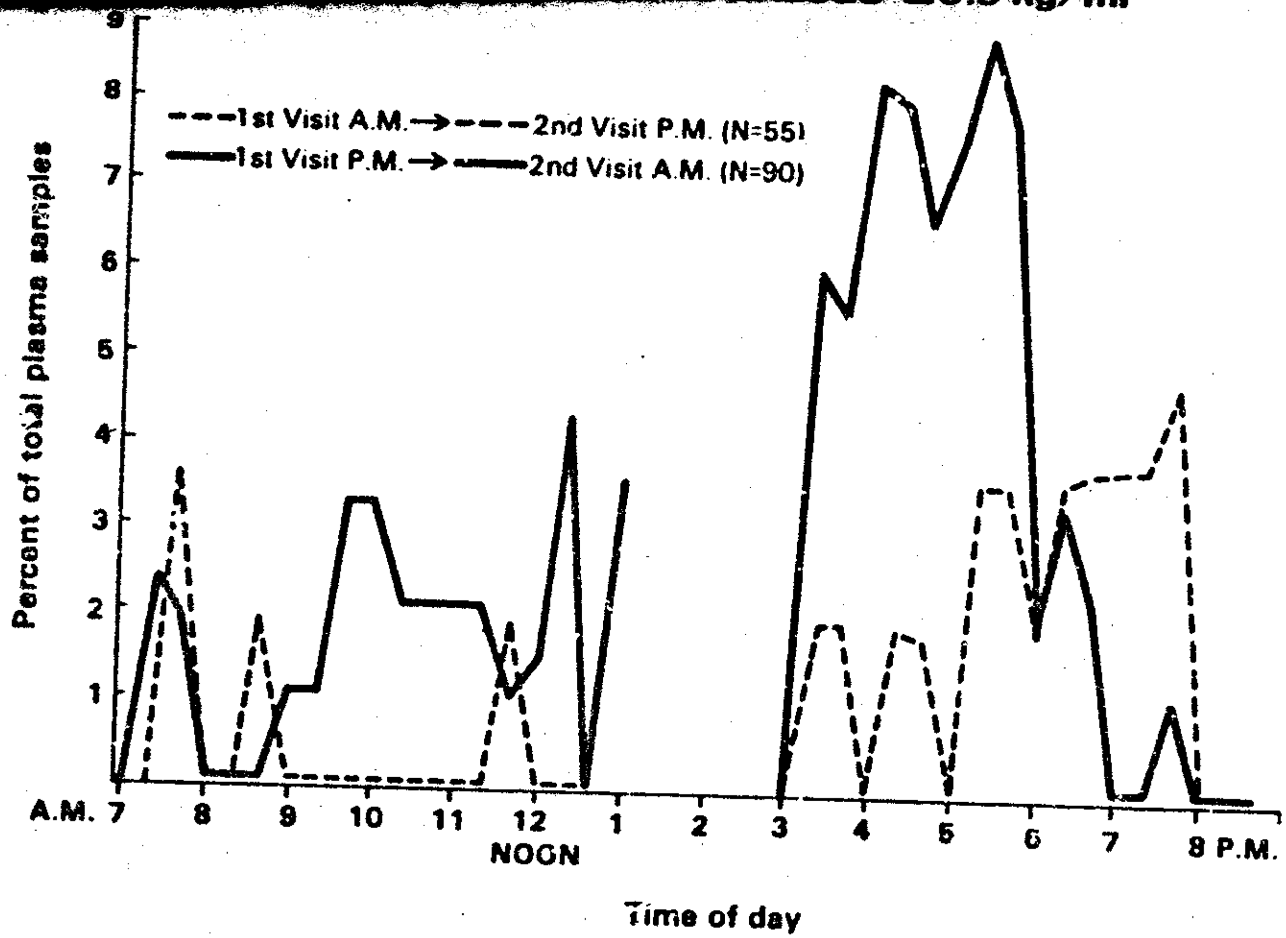
Catheterization Effect

To examine for the possibility of a catheterization effect, all samples collected were sorted by their ordinal position during the day for each man studied. Thus, regardless of the time of day when the catheter was inserted, the first sample was labeled #1 and the sample drawn one hour later, #4. For the entire first round of studies only 3.8% of the samples had a HGH level greater than 4.0 ng/ml. However, for samples 3 through 7 of the 15 samples drawn on each man during his first day, the percentage of values greater than 4.0 ranged from 6.1% to 8.5%, the latter being the frequency of elevated readings on sample 5. It should be noted that the third sample was gathered 40 minutes after insertion of the catheter and represents a moving average of HGH levels from the period 21 to 40 minutes after catheterization. Only 3.6% of the blood samples representing the accumulation from 1 to 20 minutes after catheterization were greater than 4.0 ng/ml. What is surprising is that the level of modest and higher elevations in HGH continued through the 7th sample which represented blood gradually gathered from 101 to 120 minutes after insertion of the catheter. It would seem, therefore, that if in fact these elevations represented a catheterization effect, the responsiveness of the growth hormone system (1) must have taken at least 20 minutes, (2) was highly variable from subject to subject and (3) occurred in only a minority of all subjects undergoing this procedure. This elevated frequency of higher growth hormone levels from 21 to 120 minutes after catheterization observed for visit 1 did not replicate in visit 2. For this visit only 2.8% of all samples had HGH concentrations greater than 4.0 ng/ml, and the highest percentages among the five samples 3 through 7 were 3.6% and 3.7%. The two ordinal sample numbers having the highest percentages of HGH greater than 4.0 ng/ml were samples #10 and #15 which occurred toward the end of the day. This leaves the question of whether or not HGH is normally elevated as a function of a catheterization effect with rather equivocal results. The modest, occasional, somewhat delayed set of elevations during visit 1, which might be construed as a catheterization effect, clearly did not replicate in visit 2.

Diurnal Variation

The next issue we investigated was whether or not there was a diurnal effect on HGH secretion, such as that consistently observed for cortisol. It was necessary to control for visit number in order to test the difference between morning and afternoon, inasmuch as it has been shown that first visit HGH values tended to be higher than those at the second visit. We therefore selected two groups of men: those whose first study was during a morning and second study was during an afternoon, and the second group whose first study was during an afternoon and whose second study was during the morning. For each of these men in the two groups, HGH levels were plotted by time of day according to the same 20-minute intervals at which blood was gathered. Figure J shows the percent of total samples drawn at each 20-minute period with HGH values of 5.0 ng/ml or greater.

The following inferences can be drawn from the graph: (1) More high values of HGH were observed in the afternoon than in the morning; (2) there were no trends hour-by-hour within the morning or within the afternoon; (3) for the two groups selected by these definitions, visit number was not important. The previously observed "visit effect" must have come from persons who were studied in the afternoon on both days or in the morning on both days or those who had only one study and refused participation on later days; and, (4) these two groups differed in total potential of HGH secretion, in that the group whose first study was in



1st Visit A.M. → 2nd Visit P.M. (N=55)
1st Visit P.M. → 2nd Visit A.M. (N=90)
Percent of total plasma samples
9
8
7
6
5
4
3
2
1
A.M. 7
8
9
10
11
12
NOON
1
2
3
4
5
6
7
8 P.M.
Time of day

the afternoon and second study was in the morning tended to have more high HGH secretions no matter what time of day they were studied. This may have been either a random phenomenon, "the luck of the draw," or it may have been due to some self-selective process related to who was willing to have their first study performed during a given shift.

Consistency of HGH Responsivity Over Time

Did the men who showed HGH spikes during the first day of study also have the same secretory pattern on other days? HGH assays are time consuming and expensive and therefore were performed only on blood samples of field visit days 1 and 2, with a small additional analysis from visit 3, described later. Two hundred and ninety-nine men had completed HGH data for 2 study days. Table 98 shows the consistency of the maximum value for these men on each of these two days. It will be noted that consistency is very low and that the maximum reading for most men on most days is so close to zero as to be undetectable in most lab analyses. Even for the 35 men who had one or more HGH values between 5.0 and 9.9 ng/ml on the first study day, 60% had all values less than 2.0 on the second study day. Only for the 17 men with a value of 10.0 ng/ml or greater on the first day did a majority also have at least a 5.0 ng/ ml secretion on the second day. The 59 men who had at least one HGH value of 5.0 ng/ml or greater on one (but not both) of the two study days were checked to see what their secretory pattern was on their third study day, if in fact a third day's blood samples had been drawn. Complete sets of frozen plasma samples from the third study day were located for 54 of the 59 men who had on one occasion showed one or more samples with HGH levels of 5.0 ng/ml or greater. Of these 54, only 5 had one or more samples again at that elevated HGH level for the third day. This relatively low yield again underscored the strong tendency for usual day-time levels of HGH activity to remain very low in most persons most of the time despite mentally demanding tasks. The finding also underscored the low consistency of response patterns on different study days.

Physical Correlates of Elevated HGH Secretion

The next major set of analyses concerned the search for characteristics of those persons showing the potential for secretory bursts of HGH. To those 19 men who had HGH values of 5.0 ng/ml or more on both of the first two study days, we added the 5 men among the 54 whose third study day plasmas were analyzed and who thus qualified as having one or more high HGH levels on two of three days. This high secretory group (N = 24) was then compared with 131 men who had no values exceeding 2.0 ng/ml for the first days, ninety men who had one or more values greater than 2.0 ng/ml but

TABLE 98

Human Growth Hormone

Consistency of Maximum Value on Two Study Days

with Fifteen Blood Samples Analyzed Each Day

Maximum HGH Value (in nanograms)
Second Day

Maximum HGH Value
First Day

	≤2.0	2.1 - 4.9	5.0 - 9.9	10.0+	Total
2.0	131* 73.6%**	31 17.4%	14 7.9%	2 1.1%	178 100%
2.1-4.9	46 66.7%	13 18.8%	6 8.7%	4 5.8%	69 100%
5.0-9.9	21 60.0%	4 11.4%	8 22.9%	2 5.7%	35 100%
10.0+	7 41.2%	1 5.9%	5 29.4%	4 23.5%	17 100%
TOTAL	205	49	33	12	299

* Number of Men

** Percentage of men at each First Day's Maximum Level having stated Maximum on Second Study Day.

less than 5.0 ng/ml on either of the two days, and 54 men who had a value of 5.0 ng/ml on only one of the three days for which data were available for them.

A substantial number of physical exam findings were associated with the tendency to have at least occasional secretory episodes of HGH during waking hours. The 24 high secretors were substantially lower in weight at intake than the rest of the men. Their average weight was 170 pounds compared with 183 pounds for the entire group participating in field studies (P = .02). Their pulse rates were also considerably lower at 68 bpm vs. 72.3 bpm for the entire group (P = .02). They also were closer to ideal weight, being only 2% overweight, whereas the 131 men who never had a measurable elevated HGH averaged 12% overweight (P = .007). These same findings were replicated at the second round physical examination and in addition, resting casual blood pressures were studied and found to be lower in the HGH secretors. This group had average blood pressures, 124/86, whereas the nonsecreting group had pressures of 130/90. The systolic blood pressure difference was marginally significant, (P = .07), while the diastolic blood pressure difference was clearly significant (P = .02).

Psychosocial Correlates of Elevated HGH Secretion

Several psychosocial characteristics of the various HGH secretion groups were also found to be of marginal or clear significance. For most analyses of variance, the bulk of the F statistic for significance came from the distinction between the high secretors (N = 24) and the other three groups with varying levels of low HGH. The men with demonstrated capacity for high growth hormone secretions showed less pathology on the PSS at Round 2. They had no instances of significant alcohol abuse (P = .04) and no instances of serious work role impairment (P = .06). Nevertheless, they tended to show somewhat more depression on the POMS at intake into the study (P = .04).

Men with higher HGH secretions tended to report more life changes at the Round 2 examinations. This only reached significance for the self-ratings of how distressing these changes had been (P = .03). The high secretors also were the least likely to use physical activities as a coping mechanism (P = .03). A number of other findings were statistically significant or marginally so, but some of these represented curvilinear associations which were essentially uninterpretable.

In summary, men in this study who were HGH responders tended to be close to ideal weight, to have somewhat lower pulse rates and blood pressures, to show more depression, and to be more distressed by their recent life changes. Despite this tendency to emotional sensitivity, they did not let these unpleasant affects interfere with their work role performance or lead them into alcohol abuse problems.

SUMMARY

Special Interview Ratings of Investment, Coping and Burnout

On the basis of answers to questions regarding the ATCs' investment in their occupational identity, their methods of coping with heavy workdays and their assessment of their possibility of burning out, two psychologist interviewers provided ratings for each ATC on these dimensions.

It was found that most men were highly invested in their ATC identity and were coping moderately well with heavy days. Less than 5% were rated at very high levels of burnout.

The interviewer ratings tapped somewhat different facets of the investment and coping constructs than the self-ratings and provided an evaluation of these aspects.

11. Special Interview Ratings of Investment, Coping and Burnout

Because a large number of our psychological predictor variables were self-reports, we added a special series of interviewer-rated characteristics. These characteristics were: investment in the ATC identity, coping with the stresses of the job, coping with very heavy workdays, coping with training duties, overall coping, and overall burnout. In each area a key question or two was asked, and then the interviewer followed up responses to clarify his clinical impression of the rating to be made.

The interview and ratings were made by the psychologist who administered the PSS to an ATC. The two psychologists who made all of these ratings, Dr. Thomas Hefele and Dr. Michael Hurst, conducted 20 interviews together and made separate ratings. The inter-rater reliability was over .90 and ratings differed by no more than plus or minus one unit on a scale from one to seven.

Two questions were used to develop a clinical sense of investment. Controllers were asked how important it was to them personally to be an ATC. Their answers were rated from one (not at all important) to seven (extremely important). A large percentage responded that it was very important because of the money and the status, but we inquired further to determine whether their reasons were purely pragmatic, producing lower ratings, or whether they involved the controller's sense of who he was as a person, resulting in higher ratings.

For the second investment question, we asked the controller to provide six words that would help answer the question, "Who am I?" The words were recorded in their exact order of presentation. If an ATC failed to mention the word "controller" or "ATC," we asked where he would put it among the six he presented.

The interviewers then considered the importance a subject attached to being an ATC and the rank order of his placement of "ATC" in response to the question, "Who am I?" A high rating was given if the importance was high and "ATC" was the first or second word in the list of six. A low rating was given if the importance was low and "ATC" was mentioned last or not at all. Middle ratings were given for responses between these extremes.

The coping ratings were based on the responses to three built-in questions, followed up with whatever questions the interviewer considered necessary to make the rating with confidence. The three questions were:

1. How do you handle the stresses of the job?

2. What do you do when you have a very heavy day?

3. How do you handle training duties?

The interviewers made a rating from one ("no coping mechanisms") to seven ("very effective mechanism") for each of these questions. High ratings were given for active, problem-solving and effective approaches. Low ratings were given for drug use, explosive behavior, and no problem-solving activity. For example, a high rating would be accorded to someone who did something relaxing during a break, and a low rating would be given to someone who obsessed over a difficulty. Similarly, one who would take a walk and work out the difficulty was given a high rating whereas one who tried to avoid thinking about the problem because he would get upset was given a low rating. The main criterion used to make ratings was whether or not the coping activities were goal-oriented, effective, and healthy behavior.

The three coping ratings then were used as the basis for assigning an overall coping rating. Again the rating was from one to seven and was a total clinical assessment combining all the information obtained from lead-in and follow-up questions.

Four general questions were used to provide a basis for making a burnout rating:

1. When things are starting to go "down the pipe" for you, how do you react?

2. Do you find yourself worrying about your controlling ability when this happens?

3. Do you expect to last until normal retirement age?

4. If not, why not?

For these questions, the interviewers concentrated on comparing past behaviors and expectations with present ones. Those who got more upset in the present than in the past, who did not expect to last until retirement, and who anticipated medical/psychiatric difficulties were given high burnout ratings, with "seven" the highest burnout rating and "one" the lowest. Men who were calmer, more confident, and hopeful about working to retirement were given low burnout ratings.

Table 99 displays the distribution of ratings for each area assessed by the clinical interviewers. Most men were highly invested in the ATC identity in the personal psychological sense. Only 16% had very low investment. The majority of men were rated in the moderate range on the coping indices with only 9% to 15% rated very highly. Finally, less than 5% were rated at very high levels of burnout.

Table 100 displays the correlations of the interviewer ratings with the most similar self-rated scales from the ATC Questionnaire. Interviewer ratings of investment were only slightly related to self-ratings, indicating very little overlap between the two. Burnout ratings were highly related to the bounceback-burnout self-rated factor, suggesting that an external evaluator's assessment can be quite similar to an individual's own assessment of burnout. The interviewer-rated coping variables had little or no relationship to an individual's own rating of how effective his coping mechanisms were.

Since the investment and coping factors were multi-dimensional, the interviewer ratings provided independent evaluations of some dimensions that were different from those self-reported. The lack of correlation between self-reports and interviewer ratings of these factors was therefore understandable and the comparative utility of the interviewer-ratings and self-ratings can be evaluated from our predictive findings reported in Section V.

TABLE 99

Distributions Of Interview Ratings
Of Investment, Coping, And Burnout At Intake

Qualitative: / Quantitative:	% Of ATCs Rated At Each Level: Very Low 1 - 2	Moderate 3 - 5	Very High 6 - 7	Number Rated
Investment	16.2%	39.0%	44.8%	415
Coping with stresses on the job	32.8%	56.7%	10.5%	411
Coping with heavy work-days	26.8%	56.6%	16.6%	415
Coping with training duties	26.0%	64.3%	9.7%	415
Overall coping	21.7%	69.2%	9.1%	415
Burnout	65.1%	20.3%	4.6%	415

TABLE 100

Correlations Between Interviewer And Self-Rated Investment, Coping Effectiveness And Burnout*

terviewer Rating	Correlation with	Self-rated scale on the ATC Questionnaire
vestment	.10 ($p < .05$	Investment Factor
rnout	-.62 ($p < .0001$	Bounceback (High)- Burnout (Low) Factor
ing with stress on the job	.05 (N.S.)	Coping effectiveness scale
ing with heavy kdays	.09 (N.S.)	Coping effectiveness scale
ng with training	.08 (N.S.)	Coping effectiveness scale
ll coping tiveness	.08 (N.S.)	Coping effectiveness scale

402 with all measures used to compute correlations.

SUMMARY

Methods of Measurement

The data collection system devised to ascertain health changes included the following:

1) Health History Form
2) Monthly self-reports via a checklist of symptoms mailed to the ATC and back (The Monthly Health Review)
3) Monthly self-report of anxiety and depression via a checklist of symptoms (Zung Anxiety and Depression Scales)
4) Physical examinations by a physician at approximate 9-month intervals at Boston University Medical Center
5) Psychological examinations during the same visits to Boston University Medical Center by two psychologists
6) Audiological examination at Boston University Medical Center
7) Laboratory tests (blood chemistry, urinalysis)
8) Serologic tests for syphilis
9) X-rays of the chest
10) Electrocardiogram
11) Tests of pulmonary function.

1. Met.ods of Measur(ment

The essential central consideration for the Air Traffic Controller Health Change Study (ATC HCS), or any other project involving the continuing surveillance of health states, was the establishment of a system for ascertaining the full gamut of illesses and injuries during the surveillance. Many longitudinal tudies of the health of industrial groups have been performed, nd it was expected that we would be able to select from among everal well-established health reporting systems one which s well suited to this project. A sufficiently comprehensive alth ascertainment system with established validity was not und, but we were able to profit by prior work in our efforts assemble such a system for this study. This section describes general approach to this task, indicating some weaknesses and blems to be avoided. Later sections describe in detail each the several components of our final health ascertainment sys-.

"Health change" is indeed a broad concept. It includes nges of a positive as well as a negative nature. In view of generally good health enjoyed by all of the ATC HCS particis at the beginning of our project, and taking into account direction of emphasis usually present in the medical model diagnosis and treatment, our study concentrated on negative th changes rather than positive ones.

ition of Health Change

The determination of a change from a "healthy" or "normal " is quite simple for clearcut instances of serious disease jury. The task of determining a change from a healthy state s more problematic, however, when one attempts to include ll range of human ills, particularly minor conditions.

hich conditions are considered illness or diseases and are considered merely variations of the "normal state" is, final analysis, determined by society. Different social vary in what they consider to be illness. This holds, not r other cultures, but also within our own society where nt social classes and the two sexes differ sharply in what ons they consider to be "symptoms of disease" and what circes lead them to take sick leave or to consult a doctor. nsulting a variety of sources, including the work of the Center for Health Statistics of the Department of Health, n and Welfare (US DHEW, PHS Publication No. 1000, May, he study team defined negative health change as any condiing one or more of the following characteristics:

(a) it creates unequivocal sustained pain or discomfort to the subject;

(b) it causes a loss of ability to perform one's usual activities for a day or more;

(c) it is determined by medical examination or laboratory test to pose a threat for the future continued well-being and normal functioning of the subject.

If the deleterious health change was attributable to an identified physical insult, whether the agent was mechanical, thermal or chemical, the health change was termed an injury. Otherwise, it was termed an illness.

The issue of distinguishing between true illness and complaint behavior is a critical one for any study of health change. The ideal circumstance would be to be able to diagnose disease completely independently of verbal report of the study subject. This can only e done for a few conditions, such as hypertension. For the great ajority of genuine health changes, however, verbal report by the tudy subject plays an essential part in the diagnosis. The larger he part played by more objective findings from medical examination r laboratory, the more convincing the diagnosis decision. Nevereless, even if no objective findings are possible, if the comaint behavior leads to assuming the "sick role" with its attennt reduction in occupational and social functioning, we must conude that we are dealing with genuine disease in the historic and cial definition of the term. The only question remaining is ether the disease is somatic, psychosomatic, or more purely emonal. For these reasons, as well as the consequence of our criia of disease (above), we included as health changes not only atic disorders, but psychiatric disorders as well, providing t they created unequivocal sustained pain or discomfort, or sed a loss of usual function, or posed a threat to the future -being of the subject. Our range of procedures for ascertainhealth change was constructed to be appropriately broad to r this full range of forms of pathology.

Health Change Data Collection System

The system designed for ascertaining health changes in the CS had the following attributes: it relied both upon frequent report and periodic in-depth examination by a physician and chologist; it employed laboratory tests to detect subclinical tions in somatic health and, in like manner, used quantified logical procedures to determine subclinical changes in emoand social adjustment; it probed systematically for indicaof physical illness, injuries and emotional problems; it screened monthly in a superficial manner for the more common signs and symptoms of health change in order to insure relatively complete detection of those conditions which were mild and of short duration, and hence, likely to be forgotten after a month or more; it examined at nine-month intervals for more serious or chronic conditions. This schedule represented an optimum solution to three different needs: the need to monitor health change closely, to keep examination costs within feasible limits and to maintain participant motivation to continue in the project. The balancing of frequency and depth of data collection on health change was guided by research from the National Center for Health Statistics. Reports from NCHS showed that the relative volume of illness reported per unit of time declined as the length of recall period increased. One study found that in sample of persons known from physician's office data to have visited the doctor within a two-week period, 36% of doctor's visits failed to be reported to survey interviewers, even though other relatively sensitive health data were often volunteered. More serious medical events had a higher probability of being recalled over longer periods (US DHEW, National Center for Health Statsitics, Series 2, No. 7, July, 1965). Although weekly or biweekly reporting by mailed questionnaire would have been desirable in the ATC HCS from the point of view of reducing forgetting, it was decided that a request for more frequent comprehensive reports would probably lead to greater subject resistance to the study, and thus, to more carelessness in reporting and a greater failure to respond. Similarly, the scheduling of the comprehensive physician examinations was derived by balancing the costs of repeated examinations, the yield of newly discovered events as a function of interval, the desire to maintain high rates of cooperation, and the likelihood of failing to discover important if transient health changes if the interval between examinations were too long.

The data collection procedures which comprised the system for ascertaining illness and injuries in the ATC HCS are listed in Table 101 The table specifies what data were collected each mouth, those collected at periodic medical examinations, and those special sources of information sought out at irregular intervals from hospitals, doctors, and other sources when participants gave indication (usually through the Monthly Health Review) that they had in all probability incurred a health change and had consulted one of these sources.

Instruments and Methods Used in Data Gathering

The non-copyrighted data-gathering instruments listed in Table 01 are appended as Exhibits to this report. A brief description each follows:

							MONTHLY	PER INCIDENT
Medical Questionnaire	Past medical history	X						
	Family history	X	X	X	X	X		
	Smoking, eating, drinking habits, medications, and review of systems	X	X	X	X	X		
2. Physical Examination	Physical findings, BP	X	X	X	X	X		
	Timed vital capacity	X		X		X		
	EKG	X		X		X		
	Chest films	X				X		
	Urinalysis, CBC, ESR	X	X	X	X	X		
	Serology	X		X		X		
	Audiological examination			X				
3. Psychiatric Interview	Psychiatric Status Schedule	X	X	X	X	X		
4. Health Checklist	Physical signs and symptoms as episodes, continuing problems, or isolated events						X	
	Accidents and injuries						X	
	Days lost from work and days below par						X	
	Visits to physician or hospitalization						X	
5. Mood Checklist	Zung anxiety and depression questionnaires						X	
6. Physician and Hospital Reports	Records of illnesses, diagnoses							X
7. Special Diagnostic Tests & Consultations	As required to make diagnosis							X
8. Sleep Questionnaire	Characteristics of sleep complaints							once
9. Headache Questionnaire	Characteristics of headache complaints							once

Review of Health History

Usually, the first task facing each ATC participant at each of his five examination visits at nine-month intervals to Boston University Medical Center was to complete this comprehensive health history form. Early in the study the questions were displayed and responses made at a computer video terminal. Later, however, it was found more efficient for the participant to fill out a printed form and take this with him to the internist at the time of his physical examination. This permitted immediate review by the internist of all replies, questions about uncertain points, and greater emphasis on certain parts of the physical examination and medical history guided by positive answers on the health history form. After the participant's visit, the health history, as clarified by discussion with the internist, was entered into computer files, enciphered, and added to the rest of the participant's health data. The health history consisted of a 12-page form which dealt with the following categories of data in sequence: his own and his family history of major diseases (after the intake round, only changes in this family history were recorded), smoking history, regularity of meals, consumption of caffeinated beverages, sleep patterns, use of "over the counter" and prescription medications, a symptoms checklist reviewing the following systems: ophthalmic, otolaryngologic, respiratory, cardiovascular, gastrointestinal, musculoskeletal, genitourinary, hematologic, endocrine, dermatologic, and neurologic, allergies, exposure to toxic substances, and surgical interventions.

These data were stored on computer tape and were used in a number of analyses dealing with later development of symptoms in various organ systems. The ability of the data to predict later health change was also analyzed. We were particularly interested in the development of respiratory, gastrointestinal, and cardiovascular conditions — the categories which have proved to be most frequently involved in the health changes in this study population. The main immediate use of these data, however, was to help the physician to gather a comprehensive primary and interval medical history, to suggest the need for more specific non-routine diagnostic procedures, and to assist in making diagnoses. Diagnoses of medical conditions derived from this data-gathering instrument, as well as all others, were recorded on diagnostic health change forms by the physician and entered into a separate computer file. The fact that the health history contributed data to the making of the diagnosis is recorded where applicable on the diagnostic summary. The protocols and procedures for the diagnosis of health change are described in sections of the report to follow.

Physical Examination

The physical examination made at nine-month intervals was guided by an extremely detailed checklist form which reviewed 19 systems of anatomy and physiologic function in a sequence compatible with the usual thorough medical examination. All systems were examined at the initial and final visits; the third visit omitted rectal examination; the briefer Round 2 and Round 4 examinations included only heart, lungs and abdomen, in addition to recording vital statistics. Positive findings were checked on the pre-coded forms, or occasionally, for unusual findings, written in space provided. These data were then entered into participants' files in terms of a possible vocabulary of over 900 six-digit entries, each signifying a specific type of clinical observation. These were, in turn, enciphered, and in addition, the identification number of the participant was enciphered so that a double protection of confidentiality was assured. The bulk of these data files contained negative findings reflecting the general good health of the study group. The relatively few positive irregularities noted reflect a great variety of usually minor deviations from the norm. Many of these deviations appear to have no long-term implications for health, but they were recorded, nevertheless, in the interests of completeness and thoroughness. The data within the physical examination which are most amenable to statistical treatment are a series of continuously distributed variables recorded on the face sheet of the exam form. These include height, weight, pulse, respiration, and repeated measures of blood pressure in standing, sitting and supine positions. Findings from the physical examination and medical history at intake were carefully recorded and referred to periodically to determine whether new medical findings represented a continuation or exacerbation of a condition existing at intake or whether they represented a new health change.

Laboratory Studies

At each of the five visits participants made to Boston University Medical Center for physical and psychological examination, comprehensive laboratory studies were performed. These tests, performed on fresh specimens obtained at each visit, were erythrocyte sedimentation rate, complete blood count which included automated analyses of hemoglobin, hematocrit, red blood cell count and indices (mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration), white blood cell count, and differential count of leukocyte varieties. Also performed was a urinalysis, including a description of the sediment of the centrifuged specimen.

At the first and last visits, serologic tests for syphilis were made, as well as posterior-anterior and lateral roentgenograms of the chest, standard 12-lead electrocardiogram, and two tests of pulmonary function — total vital capacity and first-second forced expiratory volume. Stool specimens obtained at the time of digital examination of the rectum were tested for the presence of blood.

A chemical analysis of the blood was also made at the initial examination. This was the standard package, "SMA-12" (calcium, phosphate, glucose, urea nitrogen, uric acid, cholesterol, total protein, albumin, bilirubin, alkaline phosphatase, lactose dehydrogenase, and glutamic-oxalecetic transferase), plus glutamic-pyruvic transferase. For Rounds 3 and 5, an expanded package became available and was utilized -- "SMAC," including all of the above analyses and triglycerides, iron, sodium, potassium, chloride, carbon dioxide content, creatinine, and calculated albumin/globulin ratio.

The hematologic examinations and urinalyses provided not only specific data about primary blood cell disorders (e.g., anemias and leukemias) and urinary tract diseases (e.g., nephritis, cystitis), but also general information about the presence of other problems (e.g., allergies, diabetes, infections). The erythrocyte sedimentation rate is a non-specific measurement made frequently because of its value as a clue to potentially significant systemic problems needing investigation. The group of chemical tests provides a scan for the presence of dysfunction involving liver, kidney, endocrine, and other systems, dysfunction that in many cases has not yet created symptoms. The pulmonary function tests seek to identify individuals with restrictive or obstructive lung diseases, such as emphysema.

This schedule of tests had a variety of aims. The initial group sought to help establish baseline health status with a comprehensive array of both specific and non-specific measurements, looking for the presence of current conditions or diseases as well as for risk factors that might presage illness. The finding of abnormal results (unless they reflected already-known, stable conditions) led to repeated testing and, where appropriate, special investigation and consultation.

The routine followed for all subjects reflected the kinds of changes in health status most likely to develop in this population of men, and the frequency of examination was adjusted so as to minimize study costs and inconvenience to participants while at the same time assuring maximal ascertainment of health changes.

Monthly Health Review

Most minor short-term health changes which occur during a year cannot be discerned by physical examinations or by laboratory studies performed six to twelve months apart. Many of these health changes do not receive medical attention, and hence records of consulting a physician or clinic are also an inadequate approach. Finally, it has been shown that even self-report information gathered at six to twelve-month intervals is severely diminished by forgetting (Jenkins, Hurst and Rose, in press). Therefore, a method for more frequent ascertainment of minor health changes was needed. To complete this part of the clinical picture of health events, two self-report forms were sent to each participant each month. The rate of completion and return of these forms was exceptionally high, greater than 90% complete data for more than 80% of the participants.

The first of these self-report forms, called the Monthly Health Review, was a checklist of 30 specific symptoms of ill health, which included the most common signs and symptoms of dysfunction in the physiological systems most frequently afflicted by disease, particularly respiratory, gastrointestinal and musculoskeletal. Each month each respondent was asked to report which, if any, of the 30 health symptoms had occurred and whether these had occurred as part of an illness episode (a cluster of symptoms occurring together at about the same time and recognized as part of the same disease), as isolated symptoms occurring occasionally apart from illnesses, or as continuing chronic problems present on more than half the days of the month. In addition, participants were asked whether each illness episode or aggregate of isolated events or continuing problems was brought to medical attention and how many days below par and days of reduced physical activity accompanied each health change.

The Monthly Health Review also included a separate section in which to record injuries. Five types of trauma were listed so that the specifics of the injury could be reported. Severity of the injury was inferred by level of medical care received (self, medical professional, inpatient hospitalization) and the number of days of restricted activity.

Each month all Monthly Health Reviews which contained checks indicating an illness episode, more than occasional isolated symptoms, or a non-trivial injury were reviewed individually by the project internist. Where indicated, telephone calls were made to the participant to obtain further information, and such health changes as were determined to be present were recorded on a diagnostic summary form and given a numeric diagnosis according to the International Classification of Diseases, Adapted.

Further description of the Monthly Health Review, evident of its validity, and descriptions of the seasonal prevalence of a variety of symptoms therein reported are presented in a paper by Jenkins, Kreger and Rose (submitted for publication).

SUMMARY

Indexing Health Changes by Levels of Severity

Three criteria were applied to define a negative health change and to determine its level of severity: pain or discomfort, restriction of usual activities, and/or extent of threat to future well-being.

Four levels of severity of illness or injury were established, from Level 0, considered trivial, to Level 3, considered severe. Computer assignment of level of severity was made by reference to tables indicating criteria for assignment of each ICDA code to a level of severity.

2. Indexing Health Changes by Levels of Severity

The value of the ATC Health Change Study, both for administrative and scientific purposes, was greatly dependent on a system involving not only the specific diagnosis of particular diseases and health changes, but also a method for consistent rating of the severity or seriousness of each health change episode. The characteristics that make an illness undesirable, both to individuals and society, are its effects on feelings, functioning and future. In fact, if a condition does not make one feel bad and does not impair function and also does not have implications for future welfare or survival, it is not considered a disease or illness. Our system for sorting health changes into levels of severity was based on the same criteria used to determine health change: pain, dysfunction, and threat to future well-being.

We were concerned, however, that differences in reporting discomfort or pain might have an undue influence on judgments regarding the presence of disease or its severity. Persons whose psychological style was to complain most loudly might then appear to have the most frequent and most serious health changes. We were skeptical, therefore, about the reporting of pain or discomfort and were intent on "validating" such reports by additional evidence, e.g., that the patient sought medical care and, preferably, received a medical diagnosis from his own physician or the ATC study physician. Thus, the receipt of medical care, the giving of a diagnosis, the loss of time from usual activity due to illness, and the possible long-term deleterious implications of a medical diagnosis all went into the assignment of a "level of health change." Specific definitions of severity relating to each specific medical diagnosis (based on the number assigned it from the International Classification of Diseases, Adapted) were decided upon, taking into account three possible levels of medical care received (self-care, ambulatory medical care, inpatient hospitalization), and number of days restricted from one's usual activities due to illness. These definitions were entered as tables into the computer and each diagnosis rendered by the project physician, accompanied by all of the above data (also entered into the computer), was assigned a health-change-level score by the computer. Four levels of health change, described below, were identified. The specific rules for assigning health change levels to specific illness episodes or chronic conditions are included in Appendix I, Section III C.

Levels of Health Change

Level Zero was reserved for symptom complaints such as headache, pain in the limb, nausea, and other symptoms to which a more specific medical diagnosis could not be given. Level Zero was also assigned to rather specific physical signs, e.g., deformity of foot, excessive coughing, certain skin and dental conditions which appeared to represent genuine physical findings whether or not brought to the attention of a health professional, but which were attended by no days of restricted activity.

Level One was assigned to brief self-limiting illnesses. These often resulted in one to six days of restricted activity, but if medically attended, in many instances were not associated with specific numbers of "days down." Peripheral nerve disease, gout, otitis media, upper respiratory infection and gastroenteritis are typical examples. Level One health changes rarely involved seven or more days of restricted activity, and never involved hospitalization. The average upper respiratory infection, non-specific viral infection, or occasional exacerbations of conditions like allergy or back problems are typical of Level One health changes. Most injuries treated by a physician, possibly in an emergency room (but not involving overnight hospitalization) were also considered Level One.

Level Two health changes were more extended or aggravated illnesses and injuries, all of which received medical attention. They also all involved some limitations of activity, usually a week or more. A number of exceptions included conditions which had greater-than-average chances for complications or future health problems, such as urinary calculus, hospitalizations for cellulitis, short episodes of acute hepatitis, recurrence of pain plus days of restricted activity in a previously diagnosed ulcer of the stomach. Despite its lack of pain and absence of hospitalization, glaucoma was always rated Level Two. Had conditions like these developed into more serious diagnoses, a Level Three would have been recorded.

Level Three illnesses were serious ones requiring medical attention, and having in addition at least one of the following attributes: (a) more than 30 days limitation of activity (this occurred a number of times with moderate illnesses which developed severe complications, and with serious injuries); (b) hospitalization for a life-threatening or career-threatening disease; (c) longstanding or permanent disability from injuries; and (d) emergence of any disease which typically involves a serious threat to longevity or future functioning (even without hospitalization), such as discovery of a malignancy, coronary heart disease, hypertension, gastric ulcer, diabetes.

As might be expected in a healthy population screened annually to meet job requirements, the number of Level Three health changes encountered in the ATC HCS was relatively small, and Level Two and Level One diseases were progressively more frequent. The frequencies of the most common illness or injury conditions diagnosed by level of health change are presented in Section III C, Prevalence and Incidence of Physical Health Changes.

SUMMARY

Illness by Etiological Category: Frequency of Specific Conditions

The most common medical problems among the ATCs at entry into the study were hypertension (32.5%), hemorrhoids (30.5%), hay fever (22.6%), and a combination of skin conditions (boils, tinea, pilonidal cysts and sebaceous cysts) that together affected almost 39%. Chronic bronchitis was prevalent among 12% of the men at intake, and active peptic ulcers were prevalent in 6.5%.

In order to use the data on health changes gathered in the ensuing three years of health surveillance, it was necessary to have a summary index that would represent each man's illness experience and that could be used in analyses of possible predictors. Such an index was calculated for each man by summing his Level 1 and 2 illness and injuries during each interval, weighting the Level 2 events by counting each such event as two, then annualizing the interval data to produce a comparable period at risk for all men, and finally, averaging the annualized rate for the four intervals. This "compressed, annualized incidence rate" was used in all further analyses of moderate illness and injury. Level 3 illnesses and injuries were analyzed separately.

The most common diagnoses during the 3 years after intake were upper respiratory infection (20% of the total), sprains, common cold, non-specific viral disease and gastroenteritis.

On the average, participants in the study had 2.5 (compressed, annualized) health changes per year.

Of Level 3 events, 21 severe injuries were incurred by 18 men, 36 men were newly diagnosed as hypertensive, 25 other Level 3 diagnoses were in various categories of illness and were diagnosed in 23 men.

a. Illness by Etiological Category: Frequency of Specific Conditions

Three paramount objectives of this study were: 1) to determine the frequency, severity, and types of health changes within our sample of air traffic controllers, 2) to establish whether or not these health changes were related to work facts, and 3) to ascertain whether or not health changes among our study population could be predicted by psychological, work, or other factors.

As described in IIIC, a comprehensive system was established to monitor and record health changes and to index their severity. In this section we will report first on physical (somatic) disease or illness conditions existing at entrance into the study (prevalence at intake) and then on physical health changes diagnosed during the 36 months of surveillance following intake into the study (incidence of illness). The two most frequent acute physical health changes, respiratory illness and injuries, are described in separate sections following the general report of findings. The most common chronic disease, hypertension, is also examined separately. Psychological health outcomes are reported in Section D.

Physical health states and changes were classified by physiological-anatomical system into the following categories: Respiratory; Nonspecific Viral; Acute Gastrointestinal; Non-infectious Gastrointestinal; Hemorrhoids; Skin; Bones, Muscles, and Joints; Allergies; Eye; Ear; Dental; Genitourinary; Hypertension; Injuries; and Other. Within these categories, diagnoses were coded according to the International Classification of Diseases, Adapted.

Conditions existing at the intake examination, as well as illness in the past, were recorded by whether the condition was: 1) a single event in the past; 2) a chronic condition inactive for at least the past two years; 3) a chronic condition active at the time of intake, or, 4) a problem newly discovered at intake. Physical health changes developing after intake were recorded by their level of severity as determined by computer algorithms taking into account ICDA code, type of care received, and days of restricted activity. This system is described in IIIC, Indexing Health Changes by Levels of Severity.

A total of 6,217 diagnoses of physical illness conditions were rendered: 31% (1,916) of them were conditions observed at intake or reported at that time as past events, while the remaining 69% were illness episodes and conditions that developed after the beginning of the study.

Prevalence of Illness Conditions at Intake

The illness conditions recorded at entry were categorized by their chronicity and current status, and were distributed as follows: single illness events in the past (approximately 22% of the 1,916 intake diagnoses), conditions reported as formerly chronic in the medical history but inactive for at least two years prior to the initial examination (6%); chronic conditions active at the time of the initial examination or recently active (48%) and problems newly found at the initial examination (24%).

We were interested in the illness conditions experienced by the men before entrance into the study primarily because of their possible value in predicting future illness. The illness conditions existing, and those discovered, at the intake examination also provided a base measure of the general health status and the prevalence of specific conditions among ATCs at entry into the study. This permitted us to distinguish between continuing health difficulties and the development of new disease in the ensuing three years.

Certain illnesses and illness conditions were prominent in the profiles of the ATCs at entrance into the study. These conditions are shown in Table 102 with the number and percent of the study participants who reported them as "currently inactive conditions" (i.e., single events or chronic but inactive conditions) and the number and percent experiencing them as chronic and active, or new-found at intake.

Examination of Table 102, "The Most Common Medical Diagnoses Among 416 ATCs At Intake Into The ATC HCS," reveals a substantial excess of chronic active and newly discovered conditions as contrasted with currently inactive conditions. Careful methodological studies have shown substantial underreporting of illness conditions in the past, particularly of minor self-limiting illnesses. Furthermore, many minor conditions which do not have a name used in common parlance (e.g., varicocele) are unlikely to be named even if their prior existence was recollected. Although the information about past illnesses or conditions inactive at time of entry into the study, to whatever extent it was recalled, was of interest to us for our predictive analyses of future health change, the prevalence of each condition, by definition, was determined by those conditions which were documented by the examining physician as present (chronic or newly-found) at the time of intake into the study.

	ICDA	Condition (#)	Condition (#)	Total Reported (#) Men	(%) Men N=416
RESPIRATORY					
Chronic Bronchitis	490-491	1	49	50	12.0%
NON-INFECTIOUS GASTROINTESTINAL					
Peptic Ulcers	531-533	13	27	40	9.6%
Chronic Enteritis	563.9	--	15	15	3.6%
Irritable Colon	564.1	1	13	14	3.4%
Liver Inflammation	573.	2	21	23	5.5%
HEMORRHOIDS	455.	25	102	127	30.5%
SKIN					
Boils (all sites)	680.	16	27	43	10.3%
Tinea pedis, cruris	110.1, 110.9	--	40	40	9.6%
Pilonidal cyst	685.	23	11	34	8.2%
Sebaceous cyst	706.2	24	21	45	10.8%
BONES, MUSCLES, JOINTS					
Varicose veins, legs	454.9	1	26	27	6.5%
Low back pain	728.7-.9	--	40	40	9.6%
ALLERGIES					
Asthma	493.	2	17	19	4.6%
Hay Fever	507	4	90	94	22.6%
Adverse Effect:					
Penicillin	960.	1	13	14	3.4%
GENITOURINARY					
Gonorrhea	980	43	1	44	10.6%
Varicocele	456.1	--	30	30	7.2%
Prostate Problems	600-602.	11	30	41	9.9%
OTHER SYSTEMS					
Inguinal Hernia	550.	27	5	32	7.7%
HYPERTENSION	401	3	132	135	32.5%

The most common medical problem prevailing among ATC study participants was hypertension. By the diagnostic criterion adopted for this study, 135 men (32.5%) were diagnosed as hypertensive at intake. As our determination of criteria for diagnosis of hypertension was complex and involved comparisons with other studies, a complete discussion of our diagnosis of hypertension among the study sample, comparisons with other populations, and the implications of the findings are presented separately following the discussion of general findings.

The second most prevalent chronic medical problem among these men was hemorrhoids. Approximately 31% of participants suffered from this problem at intake, with the majority of these reports verified by physical examination. Hay fever, reported by almost 23% of the ATCs at intake, was next most common. Hypertension, hemorrhoids, and hay fever were the only specific conditions with an intake prevalence of more than 20%. Taken together, four dermatologic conditions (boils, tinea, pilonidal cyst, and sebaceous cyst) affected almost 39% of the sample.

Chronic bronchitis among our sample of air traffic controllers was more prevalent than we had expected, with 12% of subjects thus diagnosed. Although comparisons of the rates of chronic respiratory diseases among our sample with those of other studies would have been desirable, they are problematic because of the wide variety of diagnostic procedures and criteria used. A review of the epidemiology of chronic respiratory disease by Dr. I.T.T. Higgins (mimeograph, Department of Epidemiology, University of Michigan School of Public Health, December, 1970, p.9), stated, "Epidemiologic studies have indicated that the prevalence of chronic respiratory disease of all degrees of severity varies from under 10% to over 40% of the population, depending on age, sex, and the definitions of chronic respiratory disease which are used."

The earlier Cobb-Rose report (1973) found peptic ulcers to be excessively prevalent in air traffic controllers when compared to a group of licensed airmen. In the present study, gastric and duodenal ulcers active at intake had a combined prevalence of about 6.5%, and an additional 3.1% of ATCs in the group reported ulcers that were no longer active. The combined reported lifetime prevalence at intake was, therefore, just under 10% (96/1,000) for these men aged 25-45 years. Of the 27 currently active ulcers, nine were of the duodenum, and the remaining 18 were of unspecified site (ICDA 533).

These findings suggested an elevated morbidity for these conditions among the air traffic controllers when compared also with prevalence data from the general U.S. population. U.S. National Health Survey data derived from interviews during 1975 indicated a combined gastric and duodenal ulcer prevalence of 23 per thousand males aged 17 to 44 (U.S. DHEW, PHS communication of unpublished data). The prevalence of 65 per 1,000 of currently active peptic ulcer conditions among our study sample appears, therefore, to have been almost triple the national rate. However, the national figure would be expected to be somewhat lower because the national survey included men from 17 to 25 years of age who would be expected to have a considerably lower prevalence of ulcers. Since rigorous diagnostic criteria and cross-examination by the physician were integral to the diagnoses in our study, whereas the national diagnostic data were based on less stringent interview criteria, biases may exist in this comparison. To the extent that survey respondents in the general public may overreport peptic ulcer disease, or include under that label a variety of other gastrointestinal disorders, the likelihood is increased that our ATCs may really have had more than the 3 to 1 relative risk of peptic ulcer disease suggested by the data cited above.

Incidence of Physical Health Changes After Intake

Physical health changes (illnesses and injuries) were monitored for each participant from the date of his entry examination into the ATC HCS until his final examination approximately three years later. The first ATCs entered the study in February, 1974 and the last ones were mustered out in October, 1977. A total of 4,300 health change events were certified in the interim by the study physician. The information from which these diagnostic decisions were made came from a variety of sources. For 94% of the diagnoses, information came in whole or in part from the Monthly Health Review. For 6% of the changes recorded*, the ATC himself was involved in providing information, either by telephone or at the time of his regular examination. Additional sources of information included the physical examination itself, laboratory findings and the participants' physicians. The 4,300 new health changes comprised 3,623 acute illness and injury events (84% of the total) and 433 recurrences or reactivations of previously diagnosed conditions (10% of the total). The remaining 244 events were almost entirely the clinical emergence or discovery of chronic disease problems.

Of the total health changes developing after intake into the study, 61% were cared for by the participant himself without professional assistance. Thirty-seven percent received physician or

These percentages total more than 100% because some diagnoses were made on the basis of several sources of information.

clinic attention on an outpatient basis, and the remaining 2% (90 events) were treated by admission to a hospital.

The distribution of days of limited activity attributable to these health changes is displayed in Table 103. The median event for which a specific number of restricted days could be ascertained cost the participant between three and six days of disability. For 34% of events, days of restricted activity could not be ascertained from the information received.

The health changes were categorized into broad diagnostic groups representing the anatomic-physiologic system involved and/or its etiologic type. The distribution of health changes after intake among these diagnostic categories is displayed in Table 104. In the ATC HCS, as in most other extended studies of illness experience in normal populations, the most common health problem was respiratory disorders. These accounted for 36% of all health changes. Injuries were second most common, with 12.5% of the total of all diagnoses. Acute gastrointestinal problems, and complaints regarding bones, muscles, and joints were the next most common health changes. The frequencies displayed in this table include all 4,300 recorded health changes, including 1,445 "trivial" (Level 0) changes, 2,523 minor health changes (Level 1), 250 moderate health changes (Level 2), and 82 severe health changes (Level 3). (See Section IIIC, Indexing Health Changes By Levels of Severity, for explanation of criteria used.) It was often difficult to discriminate between complaint behavior and actual physical health changes in those diagnoses classified as trivial. This is documented by the fact that only 72% of Level 0 events were self-treated and 91% were designated as having indeterminate or no days of restricted activity. In addition, there was a profound decrease in the frequency of certification of Level 0 changes as the study progressed. We therefore decided that the Level 0 health changes were not medically important and not consistently reported; hence they were omitted from all calculations of incidence of the various health changes and from predictive analyses.

Examination of the specific diagnoses entering into Level 1, 2, and 3 health changes, and the probabilities of receiving medical care and hospitalization associated with each diagnostic entry for each level of health change revealed that the 2,773 Level 1 and 2 health changes were composed of a similar set of conditions. The main difference between Level 1 and Level 2 was the severity of the episode, particularly in terms of days of disability. In contrast, Level 3 illness events and Level 3 injuries were qualitatively quite different, posing either a threat to a normal life expectancy or a lengthy dis-

TABLE 103

Distribution Of Days Of Restricted Activity For All Diagnosed Health Changes In The ATC HCS

# of Days	# of Events	Percentage
None	879	20%
1-2	807	19%
3-6	760	18%
7-14	232	5%
15-30	107	3%
More than 30	41	1%
Undetermined	1474	34%
	4300	100%

TABLE 104

Distribution Of ATC Health Changes After Intake Among Diagnostic Categories

Etiological Group	Absolute Frequency	Relative Frequency
RESPIRATORY	1557	36.2%
ACUTE GASTROINTESTINAL	358	8.3%
NON-SPECIFIC VIRUS	236	5.5%
NON-INFECTIOUS GASTROINTESTINAL	100	2.3%
HEMORRHOIDS	79	1.8%
SKIN DISORDERS	194	4.5%
ONES, MUSCLES, JOINTS	317	7.4%
LLERGIES	172	4.0%
YE	47	1.1%
AR	110	2.6%
NTAL	219	5.1%
JURIES	536	12.5%
NITOURINARY	66	1.5%
PERTENSION	36	0.8%
HER	273	6.3%
	4300	100.0%

ability.

For these and other reasons, it was decided to combine Level 1 and Level 2 health changes into a single series of data analyses for both descriptive and predictive purposes. The 82 Level 3 health changes were treated separately and further divided into three diagnostic groups: hypertension (36 new cases), severe injuries (21 cases), and all other serious somatic illnesses (25 cases).

The basic summary statistic used in analyses of slight and moderate health changes (Levels 1 and 2) was "the average annualized compressed illness rate." This lengthy term describes how this illness index was calculated. Level 1 and Level 2 illnesses and injuries were compressed into a single index by counting 1 point for each Level 1 health change and 2 points for each Level 2 health change. This was done on the basis that the median days of disability associated with Level 1 changes was 3-6 days, whereas the median days of disability for Level 2 changes was 7-14 days, an approximate ratio of 2 to 1. Similarly, Level 1 health changes received medical attention in 39% of cases, whereas Level 2 events received medical attention in 85% of cases, again close to a 2 to 1 ratio.

The intervals between successive medical examinations at Boston University differed considerably across men for the four prospective intervals of follow-up which were terminated by Examinations 2-5, respectively. In order to compare illness rates across men for different periods at risk, each interval's illness experience was adjusted to an annualized rate. Similarly, although the rate of return of Monthly Health Reviews - the basic information source for health change decisions - was 90 95% (more fully described later), we felt that some adjustment was necessary for individuals who returned fewer than the full number of reports, and thus were less at risk for having acute illness events reported and certified. Accordingly, the rate of return of Monthly Health Reviews as well as the period of risk was taken into account in the calculation of annualized illness rates.

Examination of annualized illness rates by each of nine major diagnostic categories plus the total of all fifteen diagnostic categories revealed uniform correlations between illness rates for corresponding diagnostic categories across intervals, for the total annualized illness rate across categories, and for each of these with the grand total of all illnesses for the entire study period (Table 105). This led to the conclusion that each interval of observation (i.e., each period at risk between successive medical examinations) was an unbiased

TABLE 105

Correlations Of Average Annualized Illness Incidence Rates (Totals For All Diagnostic Categories) For Each Interval With The Average For The Entire Study Period And The Number Of Intervals Each Man Participated Actively

	Annualized Incidence For:					Number of Intervals
	Int. 1	Int. 2	Int. 3	Int. 4	Entire Study	
Annualized Incidence for:						
Interval 1	---	(320)*	(284)	(262)	(358)	(385)
Interval 2	.372	---	(293)	(272)	(339)	(339)
Interval 3	.308	.375	---	(273)	(298)	(298)
Interval 4	.319	.415	.348	---	(276)	(276)
Average Incidence for Entire Study	.759	.765	.712	.711	---	(378)
Number of Intervals of Participation	.085	.034	.087	-.009	.083	---

*Numbers in parentheses represent sample sizes from which correlation coefficients in lower left triangle were computed.

estimate of the total illness experience for each man. Therefore, the most stable, unbiased estimate of his total illness-proneness by diagnostic category and overall, would be the annualized illness rate averaged for four intervals. This statistic became our major dependent variable for slight and moderate health changes. In addition, it was found that the number of intervals in which a man was active in the study did not correlate with average annualized illness rates. Therefore, no bias was introduced by including men with only one or two scorable intervals in the analyses along with men who were full participants in the study throughout all four inter-examination intervals.

A further word on the completeness of the data is in order. Men were sent Monthly Health Reviews each month that they were active in the study. Mailings continued until participants were promoted, transferred, medically disqualified from ATC work, or until they gave notice of their desire to withdraw from the study for personal reasons. The amount of drop-out was very low, and the reasons for it are described elsewhere. Nearly 90% of all MHRs mailed out were returned to the study. There was no decrement in cooperation as the study progressed (see Table 106). There were, nevertheless, a few men who tended to be remiss in their return of Monthly Health Reviews. We decided that for men with data missing on occasional months, a fair representation of total illness experience could be obtained by prorating the reported illness experience of the reported months. We decided further that at least half the months in an interval had to have data or proration would be subject to too great a possibility of error. Since the average interval between examinations was nine months, we set a lower limit of five returned MHRs to qualify a man as "scorable" for annualized illness rates for a given interval. Men returning fewer than five MHRs in an interval were not scored for that interval.

Thirty-eight of the 416 entrants into the study had no interval in which they returned 5 or more MHRs. They thus had zero scorable intervals and were excluded from all analyses using average annualized illness rates. The remaining 378 men had one or more scorable intervals. Since number of scorable intervals did not correlate with average annualized illness rates for any diagnostic category or for the total of all categories, all of these men were included in these analyses.

The amount of proration that took place for men scored for each interval was very small. Table 106 shows the rate of MHR return, from 94.3% to 95.8%, for all scorable men in each interval. Thus, for the "fully active participants in

TABLE 106

Rate of Return of Monthly Health Reviews

Segments of Longitudinal Study	Percentage of MHRs Returned: From All Participants Receiving MHRs	Percentage of MHRs Returned: From Active Participants Contributing to the Annualized Illness Rates
Interval 1	87.8%	95.8%
Interval 2	88.5%	95.0%
Interval 3	90.8%	95.7%
Interval 4	89.6%	94.3%

the study," - the 378 men who entered into the data analyses for average annualized illness rates - the illness rates are based on 95% complete data from the primary source from which slight and moderate health changes were ascertained.

The Most Common Health Change Diagnoses

The study physician who certified the 4,300 health changes after intake utilized 318 specific diagnostic codes to classify the diagnoses. Many of these marked only one or two events each, and only 16 types of illness or injury occurred more than 50 times among the subjects (including Level 0 health change). These 16 conditions are displayed in Table 107, with number of men affected at each level of severity and total number of diagnoses. The most common diagnoses were upper respiratory infection, accounting for 20% of the total, sprains (all locations collectively), common cold (nasal symptoms only), non-specific viral disease, and gastroenteritis.

As stated before, Level 0 health changes were considered medically unimportant as well as inconsistently reported. Level 3 health changes were of sufficient severity and rarity to warrant separate analysis. We then calculated average annualized compressed illness rates based on the 2,773 diagnoses categorized Level 1 or 2. This became the summary measure of morbidity for minor and moderate illnesses and injuries. The following findings emerged: respiratory disorders were the most common health change, occurring, on the average, 1.33 times per person per year. Eighteen percent of the men reported no respiratory problems at all, while 24% reported an average of two per year or more. Injuries were the next most common, with a rate of 0.40 per man per year, with 60% of the ATCs remaining free of injuries throughout the study while the remaining 40% averaged from 0.2 to 4.3 injuries per man per year. Fourteen percent averaged 1 injury per year, or more. Acute gastrointestinal problems occurred to the average man 0.29 times per year. Sixty-one percent had no gastrointestinal problems during the entire 3 years, while amost 7% averaged at least 1 per year. Non-specific viral disease struck 25% of the men each year on the average. For 6% of the men, one episode of viral disease was recorded for each of the 3 years of participation. For 64%, no events occurred. Allergies, though common among the men at intake, had a tendency to more frequent recurrence and were observed 0.17 times per year for the entire population. This amounted to an average of 0.71 times per year in the 90 men who were subject to any allergy at all during the course of the study. Problems with bones, muscles, or joints of Level 1 or Level 2 severity occurred an average of 0.12

TABLE 107

Incidence Of The 16 Most Frequently Diagnosed Physical Health Changes After Intake Into The ATC Study By Diagnostic Code And Number Of Men Diagnosed At Each Level Of Severity*

	ICDA Code	Number of Men With Health Change By Level of Severity 0	1	2	3	Total Number Of Events
SPIRATORY						
Common Cold	460.0	108	73	4	-	249
Sore Throat	462.0	101	51	-	-	186
Upper Respiratory Infection	465.0	-	296	-	-	866
'Flu-like Syndrome' and Flu	470 - 472	-	138	-	-	207
PERTENSION	401.0	-	-	-	36	
UTE G I						
Diarrhea	009.1	62	50	2	-	142
astrcenteritis	009.2	-	141	5	-	213
SPECIFIC VIRAL DISEASE	079.9	-	139	13	-	236
ORRHOIDS	455.0	44	12	2	-	79
ES, MUSCLES & JOINTS						
ow Back Pain	728.7	69	8	1	-	104
ack Pain	728.9	34	21	8	-	73
ERGIES						
ay Fever	507.0	-	71	3	-	125
arache	384.9	50	20	-	-	84
TAL						
oothache	525.9	108	7	-	-	171
URIES						
l Sprains	840 - 848	112	62	52	6	255
l Superficial Injuries	910 - 918	49	53	4	-	124

individual ATC for whom repeated episodes of the same condition were gnosed at different levels of severity will be represented in the t at each level of severity for which he received a diagnosis.

times per man per year. All other diagnostic categories occurred much less frequently.

When all of these categories are considered together, a total illness incidence rate can be computed for Levels 1 and 2 events compressed and annualized as described. On the average, participants of this study had 2.5 such health changes per year. Only about 6% of participants reported no health changes at all. Complete absence of health changes would be statistically improbable, and we are reassured that the small number of persons in the "no health change category" reflects the honesty and completeness of reporting by our participants. There was great variation in the frequency of reported health changes, ranging from 0.2 per year to 12.3 changes per year. Section V, Predictive Findings, contains a discussion of psychological and social factors associated with varying level frequencies of health change, overall and for various physiologic systems considered separately.

Serious Illnesses And Injuries During The Study: A Summary of Level 3 Physical Health Changes

The most severe of all health changes were labeled Level 3. This diagnostic level required either a sustained period of disability (usually 15 days or more) or the emergence of a disease whose consequences might threaten future ability to work or shorten life expectancy. Although hospitalization was involved in many of the Level 3 illnesses, hospitalization itself without at least seven days of disability did not warrant a Level 3 rating of severity unless an illness of life-shortening potential was involved. Certain illnesses such as hypertension and angina pectoris were always considered Level 3 on the basis of risk of end organ damage, and others, such as ulcers of the stomach and duodenum, gastritis, diabetes mellitus and liver inflammation were rated Level 3 if they involved hospitalization.

All Level 3 health changes were categorized into four types: hypertension, injuries, psychiatric disability (which was graded Level 3 only if it involved disability and hospitalization), and all other physical illnesses considered collectively. Discussions of the incidence of hypertension and psychiatric disorder are presented in other sections. Level 3 injuries and other physical illnesses are described below.

Level 3 Injuries

During the course of the study, 21 severe injuries were incurred by 18 men, three of them having two severe injuries each. In order to qualify for Level 3 severity, an injury had to result in greater than one month of limited activity with much of this usually involving time lost from work. The complete list of severe injury diagnoses and the number of times each occurred is listed in Table 108.

Ten of the severe injuries resulted in hospitalization, and the remaining 11 were treated on an outpatient basis. These injuries included 9 fractures, 3 dislocations, 6 sprains, 2 cases of internal derangement of joint, and 1 non-specific injury.

Level 3 Physical Illnesses (Excluding Hypertension)

Hypertension, newly diagnosed in 36 men after the study began, was the most commonly occurring serious physical illness. It is discussed separately in a section following. Diagnoses were rendered for 19 other specific conditions affecting ATCs with sufficient severity to be rated Level 3. These 19 conditions affected 23 men at the severe level. As 2 of these men had 2 such diagnoses each, there was a total of 25 diagnoses of Level 3 physical illnesses, excluding those for hypertension.

Twenty of these illnesses required hospitalization and five were treated on an outpatient basis. Examination of the table may raise the question why certain illness of seemingly moderate severity such as cellulitis of the leg, shoulder bursitis, and esophagospasm were considered severe enough to be rated Level 3. In each of these instances, however, the particular illness episode was sufficiently severe or complicated as to warrant hospitalization and, in the case of the shoulder bursitis and the one severe instance of cellulitis, disability for greater than 30 days was also involved. Three of the 13 cases of peptic ulcer disease diagnosed during the course of the study were considered to be severe.

The remaining Level 3 physical illnesses covered a broad gamut which can be considered fairly typical of men aged 25-45. The complete series of diagnoses together with level of care received and days of disability are presented in Table 109.

TABLE 108

Level 3 Injuries Among Air Traffic Controllers During 3 Years Of ATC HCS

# Cases[1]	ICDA	Diagnosis	Level of Care[2]	Days Restricted
1	724.0	Internal Derangement of Joint[3]	3	
1	724.5	Internal Derangement of Joint[3]	3	
1	809.0	Multiple Closed Fractures, Trunk	3	
1	812.0	Closed Fracture, Upper Humerus	2	
1	814.0	Carpal Fracture	2	
1	816.0	Finger Fracture[3]	3	
2	822.0	Patella Fracture	3	
1	824.0	Ankle Fracture[3]	2	All required
1	825.0	Tarsal/Metatarsal Fracture	2	31+
1	825.0	Tarsal/Metatarsal Fracture	3	days down
1	836.0	Knee Dislocation	2	
2	836.0	Knee Dislocation	3	
2	844.0	Sprain Knee/Leg3	2	
1	845.0	Sprain Ankle[3]	2	
1	846.0	Sprain Sacroiliac[3]	2	
2	847.0	Sprain Neck[3]	2	
1	996.8	Injury, other[3]	3	

[1] These 21 L3 injuries were incurred by 18 men, 3 of whom had 2 L3 injuries each.

[2] Level of Care categories are: 2 = medical amulatory care and 3 = hospitalization.

[3] Only a minority of the injuries in this diagnostic category were of sufficient severity to be rated Level 3.

TABLE 109

Level 3 Physical Illnesses Diagnosed Among Air Traffic Controllers During 3 Years of ATC HCS[1]

# Cases[2]	ICDA	Diagnosis	Level of Care[3]	Days of Restricted Activity
1	012.2	Pleurisy with Effusion	3	15-30
1	162.1	Malignant Neoplasm	3	31+
2	410.0	Ischemic Heart Disease w/Hypertension	3	31+
1	410.9	Acute Myocardial Infarction	3	31+
1	413.9	Angina Pectoris	2	31+
1	503.9	Chronic Sinusitis[4]	3	15-30
1	530.9	Esophagospasm[4]	3	7-14
1	532.0	Duodenal Ulcer-Bleeding	2	15-30
1	532.0	Duodenal Ulcer-Bleeding	3	31+
1	533.9	Peptic Ulcer[4]	2	7-14
2	535.0	Gastritis/Duodenitis[4]	3	7-14
1	540.0	Appendicitis	3	15-30
1	540.0	Appendicitis	3	31+
1	541.0	Appendicitis	3	31+
1	569.9	Intestinal Disease, other	3	15-30
1	574.0	Acute Cholecystitis/Lithiasis	3	15-30
1	574.0	Acute Cholecystitis/Lithiasis	3	31+
1	575.0	Cholecystitis/Cholangitis	3	31+
1	592.0	Calculus-Kidney/Ureter[4]	2	7-14
1	592.0	Calculus-Kidney/Ureter[4]	3	7-14
1	682.4	Cellulitis of Leg[4]	3	31+
1	725.1	Herniated Disc[4]	2	31+
1	731.1	Shoulder Bursitis[4]	3	31+
36	401.0	Hypertension	3	0

1 Level 3 Psychiatric illnesses, all of which involve hospitalization, are reported under Psychological Health Changes.

2 There were 25 L3 diagnoses incurred by 23 men, two of whom had two such diagnoses, among illnesses other than hypertension.

3 Level of Care categories are: 2 = medical ambulatory care and 3 = hospitalization.

4 Only a minority of cases in diagnostic categories marked thus were of sufficient severity to be rated Level 3.

SUMMARY

Upper Respiratory Infection (URI) and Influenza (Flu)

URI was diagnosed on the basis of self-reports, via the Monthly Health Review, of clusters of symptoms including, as a minimum, sore throat and sneezing, stuffy or runny nose. Other symptoms adding certainty to the diagnosis were: temperature of 100° or more, dry cough or coughing up substances other than thin phlegm.

Flu was diagnosed upon self-report of either of two minimal clusters of symptoms: certain pains and cough and fever, or the pains and cough and nausea/vomiting. Additional symptoms, such as painful or burning eyes, or excess fatigue, added certainty.

Except for the spring months of 1974, the monthly incidence of URI among the ATCs over the three-year study ranged from 1% to 10% for "possible" URI and also for "probable" URI. Monthly incidence rates ranged from 0 to 3% for "possible" and also "probable" flu.

Altogether 866 episodes of URI affecting 296 men, and 207 episodes of flu affecting 138 men were diagnosed. URI and flu together accounted for 25% of all diagnoses during the three year study.

b. Upper Respiratory Infection and Influenza

Respiratory symptoms are commonly reported as the most frequent of all physical complaints. One study of thousands of persons in a prepaid medical care plan found that respiratory diseases among men aged 25-44 years accounted for 28% of all diagnoses, both acute and chronic, a figure more than twice that of the next most common category of diagnoses (Avnet, 1967). National data gathered by continuous sampling surveys of the National Center for Health Statistics provided an estimate that respiratory conditions accounted for 54% of all acute conditions for the total population during the year July, 1975 through June, 1976 (DHEW Publication No. (PHS) 78-1548, January, 1978).

Procedures and Definitions

The data indicating the large contribution of respiratory ailments to any total count of illnesses caused us to pay particular attention in this study both to establishing rigorous criteria for assignment of diagnoses and to the certainty with which the diagnostic label could be applied.

As described in the section on methods of measurement (IIIC), the Monthly Health Review (MHR) mailed each month to the ATCs in the study was the main source of information concerning transient or acute illnesses.

Information on upper respiratory infection (URI) was elicited through five items listed in the MHR: temperature of 100^{o} or more; sore throat; sneezing, stuffy or runny nose; dry cough, more than occasional; coughing up substances other than saliva or thin phlegm (see copy of MHR, Exhibits).

Information on influenza (flu) was elicited through eight items (either alone or in combination with various other symptoms) listed in the MHR: frequent dry cough; productive cough; fever; nausea or vomiting; painful or burning eyes; headache; stomach or abdominal pain; general myalgia ("aches or pains in muscles or joints other than the back").

ATCs were instructed to check whether their symptoms appeared as illness episodes, isolated events or continuing problems.

The symptoms checked were then assigned a diagnosis of URI or influenza on the basis of patterns or clusters conforming to definitions established for these illness as follows:

	MHR ITEM REPORTED (see code below)
Possible URI	7 and 8
Probable URI (3-point)	7 and 8 plus one of 1, 9, 10
Probable URI (4-point)	7 and 8 plus two or more of 1, 9, 10

1 = Temperature of 100° or more
7 = Sore throat
8 = Sneezing, stuffy or runny nose
9 = Dry cough (more than occasional)
10 = Coughing up substances other than thin phlegm.

A definition of influenza was derived after considerable discussion within our medical team and informal consultation with Dr. H. Bruce Dull of the USPHS Center for Disease Control and Dr. Neil Blacklow, then of the Adult Infectious Disease Section, Department of Preventive Medicine and Epidemiology, Boston University School of Medicine. Although positive diagnosis of influenza can only be made by laboratory studies, this procedure is infrequently followed in clinical practice. In lieu of laboratory diagnosis, therefore, the following definitions of influenza were accepted:

Possible flu

a) Pain (any of the following: headache, stomach pain/abdominal cramps, aches and pains in muscles and joints), AND cough (either of two: dry cough - more than occasional, or any productive cough), AND fever.

- or -

b) Pain (as defined above) AND cough (as defined above), AND nausea/vomiting.

Probable flu

a) Either triad above plus either the remaining major symptoms (fever, nausea/vomiting), OR painful or burning eyes, OR excess fatigue.

Accuracy of Reporting and Validity of Definitions

The reliability of using the MHR to obtain health change information was demonstrated by Dr. Bernard Kreger's telephone interviews which checked the accuracy of the reporting (see IIIC). In addition, a further analysis was made of three months' MHRs (February, March, and April, 1975) to determine whether or not the patterns of symptoms reported were reasonable reflections of the clinical observations of these illnesses and whether or not the procedure and definitions used produced overreporting or underreporting of illness.

The analysis showed a strong pattern of clustering among the five URI symptoms listed, and indicated that the combination of nasal symptoms with sore throat in effect defined "possible" URI. The additional presence of fever or frequent (dry or productive) cough gave the diagnosis greater certainty ("probable URI").

The three-month analysis also helped to resolve the question whether the symptoms selected as defining the flu really had a greater-than-chance association with one another.* Statistical tests of the probabilities of defining "flu pattern" by chance in the ATCs' MHRs revealed that the triads of symptoms occurred more frequently than would be expected by chance. Other internal analyses of the data were performed. All results suggested that the definitions established for a flu diagnosis appeared to be valid indicators of genuine illness among the study subjects.

Additional information regarding the three-month analysis of respiratory symptoms reported in the MHR can be found in a

---

*If the definitions used adequately mirrored the natural clustering of symptoms of these illnesses, one would expect the clusters to occur more frequently than they would by chance. The probability of the occurrence of clusters by chance are calculated as the product of the occurrence of the individual symptoms in the population. Thus, if headaches are reported by 20% and athlete's foot by 10% of the population, the headache-athlete's foot combination could be expected to occur in 2% of the population by chance alone. This type of analysis was applied to various triads of symptoms defined as flu. All of the triads tested occurred from 2.5 to 3.5 times as often as would be expected by chance. These ratios were usually, but not always, statistically significant, the expected number of events being sometimes as small as one and the observed number around five (thus creating violations of the statistical assumptions for most tests of significance).

paper prepared for publication, "Use of a Monthly Health Review to Ascertain Illness and Injuries in the Air Traffic Controller Health Change Study," included in Appendix II.

Incidence of URI

A monthly incidence of URI was calculated for the period from March, 1974 (at which time 60 men had been entered into the study) through February, 1977 (at which time 269 men returned MHRs). This period included the full recruitment phase when a peak of 372 participants (January, 1975) returned MHRs.

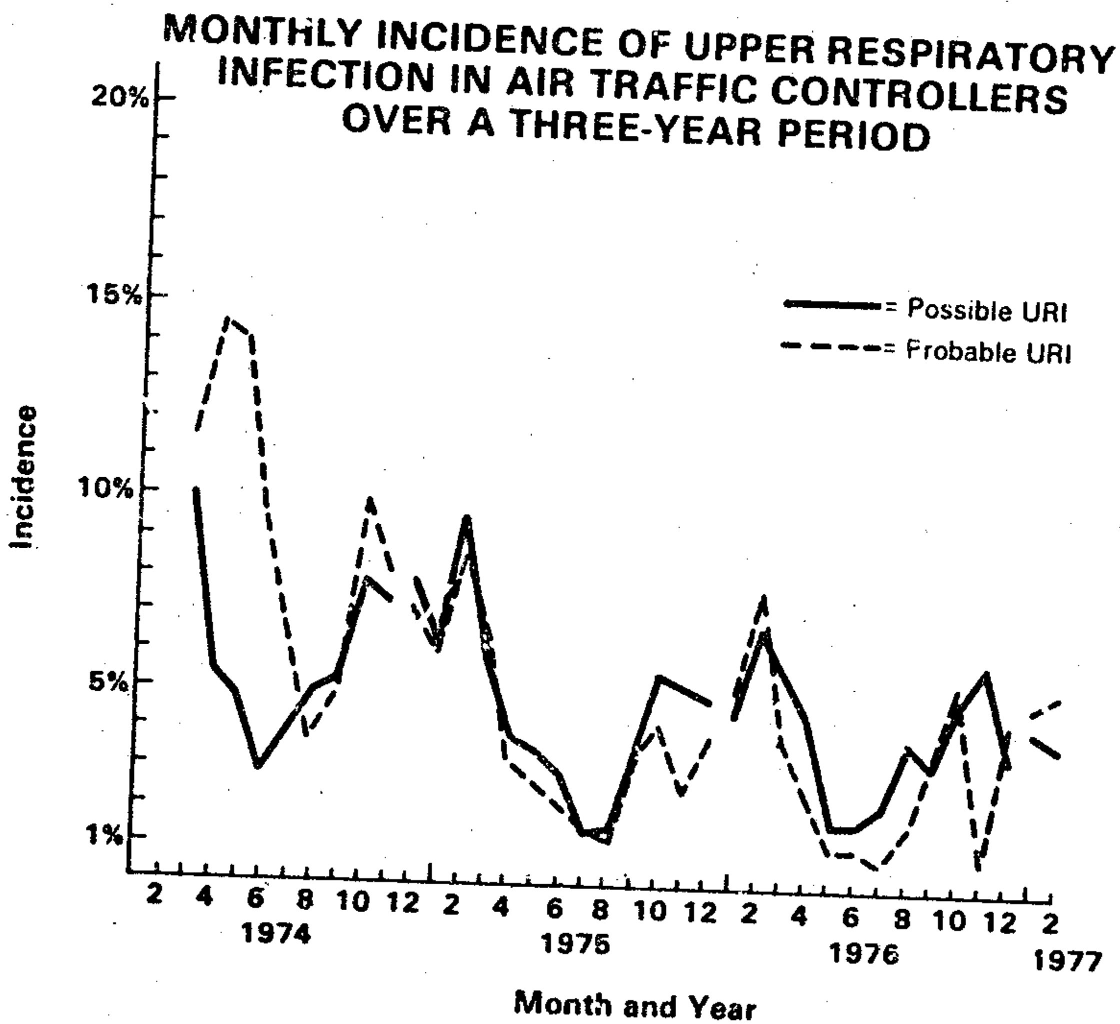
Over this 36-month period, a total of 10,549 MHRs were returned, or an average of 293 per month. Monthly incidence was calculated on the basis of the specific number of returns each month.

Figure V displays the monthly frequency of "possible" and "probable" URI. Except for the spring months of 1974, the incidences of "possible" and "probable" URI each ranged between 1% and 10%.

The three-month data on days of restricted activity and frequency of medical care, gathered for our earlier analysis, suggest that the definitions of "possible" and "probable" URI relate not only to degree of certainty of the diagnosis, but also to degree of severity of the illness.

Figure V shows that the relative frequencies of "possible" and "probable" URI follow one another quite closely with peaks in the fall and winter. The spring and early summer of 1974 present certain exceptions to these generalizations. First of all, the rates of illness are considerably higher than during the rest of the three-year period. Secondly, the relative frequency of "probable" URI is much greater than that of "possible" URI; and finally, a seasonal shift to high frequencies in late spring is noted. There are four possible explanations for this variant pattern. The first is that inasmuch as these data were gathered at the beginning of the study, the respondents may have had a tendency to overreport their symptoms. If overreporting were responsible for this apparent peak, one would expect to find overreporting of symptoms in other physiological systems as well. A number of analyses were done separately for a variety of diagnoses and for a variety of symptoms on the Monthly Health Review. As overreporting of other symptoms tended to be minimal, we concluded that the high rate of reported URI symptoms was primarily attributable to other factors. A second possible explanation was that the MHR used during the first eleven months of 1974 was different from the re-

FIGURE V


MONTHLY INCIDENCE OF UPPER RESPIRATORY
INFECTION IN AIR TRAFFIC CONTROLLERS
OVER A THREE-YEAR PERIOD
20%
15%
10%
5%
1%
Incidence
= Possible URI
= Frobable URI
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2
1974
1975
1976
1977
Month and Year

345

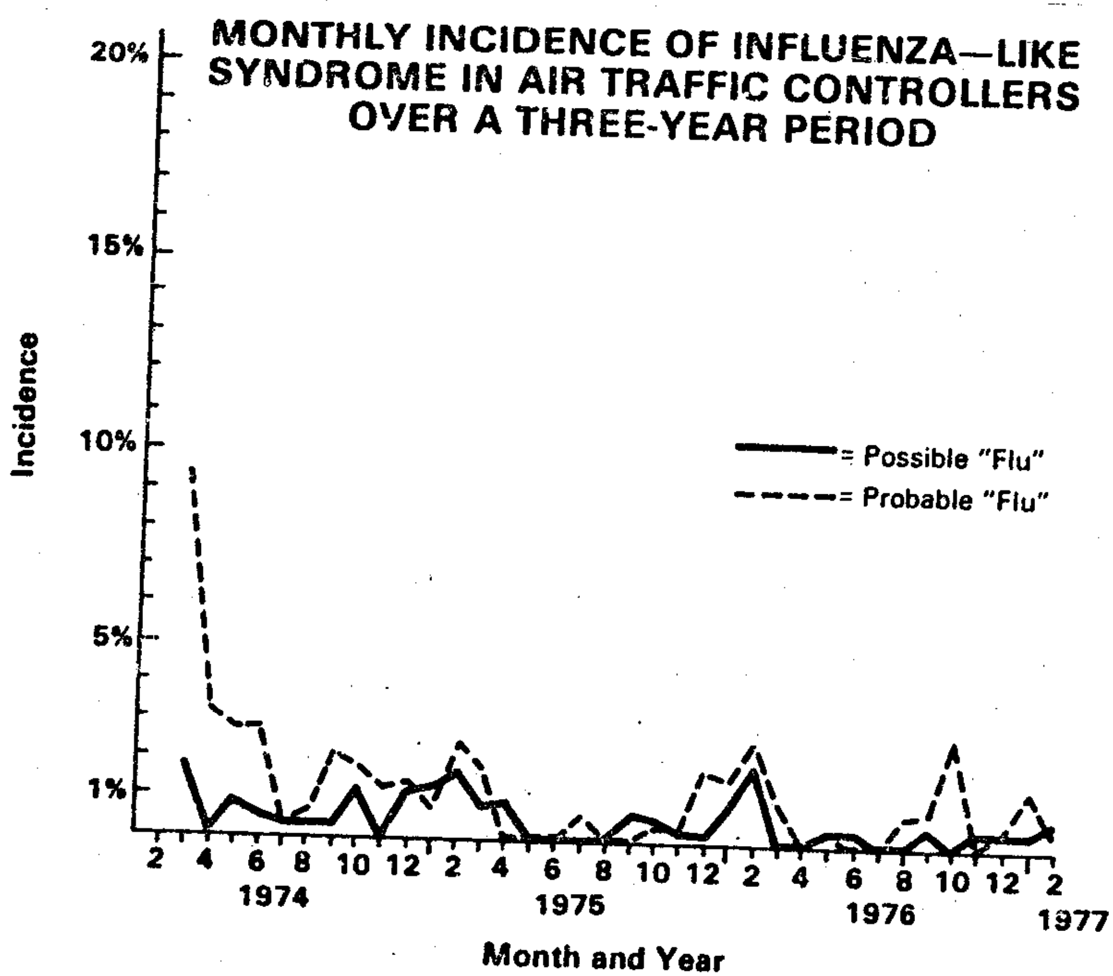
vised form used during the remainder of the study and presented in the Exhibits appended. Alhough the same symptoms were included in both forms of the MHR, the format was different in the 1974 early version. This may have influenced the frequency of URI, and as we shall note later, of influenza, during the spring. A third possibility is that the few men who entered the study in the first two months may either have been more prone to respiratory problems than other ATCs, or that the small numbers provided an extremely unstable basis for multiplying percentage rates and hence a very few URI events would project into a high rate. A fourth possible influence on the high rates was what appears from clinical anecdotes to have been a genuine epidemic of URI and influenza during the late winter and spring of 1974 in the Northeastern U.S.

Incidence of Influenza

The frequency of flu, relatively small in this healthy male population, over these 36 months is displayed in Figure W. "Possible" and "probable" flu each had monthly incidence rates ranging from 0 in the summers of 1975 and 1976 to 3% in the course of each winter. Again, the spring of 1974 appeared to be an exception, with monthly rates of 9.5% for "probable" flu and 2% for "possible" flu. The four possible contributing explanations discussed above for URI also apply to these data. The format of the original MHR form, the small sample participating in the first few months of the study, and a possible genuine excess of influenza during the early spring of 1974 appear to have contributed in unknown proportions to the high rate observed. From the summer of 1974 onward, however, the graph (Figure W) shows a consistent rhythm with "possible" and "probable" flu occurring in similar frequencies, and higher incidences occurring in winters with virtually no disease occurring in summers.

Altogether 866 episodes of URI affecting 296 men, and 207 episodes of flu affecting 138 men, were diagnosed during the three-year study. URI alone accounted for 20% of all diagnoses rendered after intake, and together with flu, accounted for 25% of all diagnoses during the three years of surveillance.

FIGURE W


MONTHLY INCIDENCE OF INFLUENZA—LIKE SYNDROME IN AIR TRAFFIC CONTROLLERS OVER A THREE-YEAR PERIOD
Incidence
20%
15%
10%
5%
1%
= Possible "Flu"
= Probable "Flu"
2 4 6 8 10 12 2 4 6 8 10 12 2 4 6 8 10 12 2
1974
1975
1976
1977
Month and Year

SUMMARY

Injuries

Information regarding injuries was gathered from the monthly self-reports (Monthly Health Review), the examinations at Boston University Medical Center and outside medical reports.

A total of 325 injuries, mostly mild or moderate, were diagnosed over the three years, producing an incidence of 34.4 injuries per 100 man-years and accounting for 11.4% of all mild, moderate and severe conditions diagnosed after intake. Injuries contributed disproportionately to the most severe conditions, comprising 25.3% of all Level 3 diagnoses.

Sprains were the most frequent type of injury, followed by contusions and superficial injuries, fractures and dislocations, other injuries including burns and concussion, and, finally, wounds and lacerations. There were only 21 injuries designated severe, and the majority of these were fractures or dislocations.

Among the 416 men, 30 (7%) had three or more injuries each, 53 had two injuries each, and 89 had one injury each.

Generally, the ATCs had a slightly lower rate of injury than males in the general U.S. population, and a somewhat longer recovery period.

c. Injuries

Along with our efforts to determine the extent and severity of illnesses, we also monitored injuries that occurred during the three years of the study. We were interested in learning whether the ATCs incurred more or less injury than others in the general population, what types of injuries they suffered, and whether or not proneness to injury was associated with, or predictable by, any of the psychosocial, biographical or physical factors that we were investigating.

Data Collection and Recording

Information concerning injuries was gathered primarily from self-reports of the controllers on the Monthly Health Review. Additional information was obtained at the Periodic Examinations (approximately 9 months apart) at Boston University Medical Center and through outside medical reports.

Injury diagnoses were recorded by International Classification of Diseases, Adapted N code, and by level of severity (see Section IIIC for explanation of the development of severity levels).

Incidence

A total of 325 injuries, Levels 1, 2 and 3 (mild, moderate and severe), were sustained during the 36-month surveillance period, producing an incidence of 34.4 injuries per 100 man-years. As with the analyses of illnesses reported in the previous sections, the trivial, Level 0 injuries were excluded from all analyses.

The 325 injuries accounted for 11.4% of all Level 1, 2 and 3 conditions diagnosed after intake into the study. However, injuries contributed disproportionately to the most severe conditions, accounting for 25.3% of all Level 3 diagnoses. The number of injuries and the number per 100 man-years in each category of injury is displayed in Table 110.

Although exact comparisons with injury data for the general population are not possible because of differences in age grouping and criteria for inclusion in the injury count, it is apparent that the incidence of injuries among the ATCs in our study was somewhat lower than the estimated 39.4 injuries (not age specific) per 100 males per year in the U.S. When the ATC rate was adjusted to reflect the same inclusion criteria as that of the government data, the adjusted ATC rate of 37.5 injuries per

... Among Air Traffic Controllers
by Type and Severity

Type of Injury	36-Month Total Number of Injury Events	Number Per 100 Man-Years[1]	Level of Severity (No. of Injuries) 1	2	3	Percent of Total Injuries
Fractures, Dislocations (ICD 802-838)	41	4.3	15	14	12	12.6%
Sprains (ICD 840-848)	131	13.9	70	55	6	40.3%
Lacerations and Wounds (ICD 870-907)	20	2.1	16	4	0	6.2%
Contusions and Superficial Injuries (ICD 910-929)	102	10.8	89	13	0	31.4%
Other (ICD 374, 724, 724.5, 738.3, 850, 944-949, 996-996.8)	31	3.3	19	9	3	9.5%
Total Injuries	325	34.4	209	96	21	100.0%

1 Man-years = average number of Monthly Health Reviews received over 36-month surveillance, or 11,336 ÷ 12 = 944.67

100 man-years approached the general U.S. male population rate even more closely.*

Employed males in the U.S. sustained a higher rate of injury (44.5 injuries per 100 employed males) reflecting the fact that work-related accidents are an important source of injuries among the general male population. The air traffic control environment, however, is not particularly conducive to accidental injury, and one would therefore expect a somewhat lower incidence among this occupational group.

Type and Severity of Injuries

As shown in Table 110, the most frequent type of injury was sprains, accounting for 40.3% of the total. Contusions and superficial injuries were next most frequent (31.4%), and the remaining types were fractures and dislocations (12.6%), other injuries, including burns and concussion, etc. (9.5%) and wounds and lacerations (6.2%).

Predictably, the largest number of injuries were relatively mild, with only 21 (6.5%) designated Level 3 (severe). A majority (57.1%) of these were fractures or dislocations, and the largest number of the remaining severe injuries were sprains. Sprains accounted for 57.3% of the moderate, and 33.5% of the mild injuries, which were mostly contusions and superficial injuries.

Injury Repeaters

The number of ATCs reporting repeated injury, and the number of injuries reported per man is shown in Table 111. Of the 416 men in the study, 89 sustained one injury each, 53 had two injuries each, and 30 men (7%) sustained 3 or more injuries each.

Factors predictive of injuries are discussed in Section V.

Days of Restricted Activity

The total days of restricted activity for all injuries was found to be 3,198, or 338.4 days per 100 man-years. In comparison, the estimated number of days of restricted activity caused by injuries for males aged 17-44 years in the general U.S. population (July 1974 - June 1975) was 298.7 per 100.

* (DHEW Publication No. (HRA) 77-1541, February 1977, pp. 9 and 17.

TABLE 111

Number Of Men Incurring Repetitive Injuries And Proportion Participating In ATC HCS Until Completion

Number of Injuries Per man	No. Men Repeating	No. Men Repeating and Participating in Study Until End	Percent Participating Until End
0	244	131	54%
1	89	62	70%
2	53	39	74%
3	14	12	86%
4	7	5	71%
5	2	1	50%
6	3	2	67%
7	1	1	100%
8	2	2	100%
9	1	0	-

Since the rate of injury for our ATC group was somewhat lower than that estimated for U.S. males of reasonably comparable age, the higher rate of "days down" for the ATC group could indicate more serious injuries or slower recovery from injury. The fact that a larger proportion of the injuries among the ATCs involved the more serious categories of fractures and sprains also helps to explain the greater number of days down.

Table 112 shows the level of care involved in the injuries. Slightly more than a quarter (27.7%) received no medical attention, approximately two-thirds (67.7%) received doctor's care, and 4.6% required hospitalization.

TABLE 112

ATC Injuries by Level of Care and Type of Injury

Level of Care	Number of Injuries					
	Fractures and Dislocations	Sprains	Lacerations and Wounds	Contusions and Superficial Injuries	Other	Total
Self	3	49	3	27	8	90
Medical Ambulatory	31	81	15	75	18	220
Hospital Inpatient	7	1	2	0	5	15
All Levels	41	131	20	102	31	325

SUMMARY

Hypertension

After examination of criteria for determination of hypertension used in several other studies, it was decided to adopt criteria more suitable for this study. Prevalence cases of hypertension were those whose systolic blood pressure was equal to or exceeded 140 mm Hg or whose diastolic pressure was equal to or exceeded 90 mm Hg at both the intake and the next examination, normally 9 months apart. At each examination, two readings of the blood pressure were averaged to produce the recorded pressure for that examination. Incidence cases were new cases developing from among the pool of those free of symptoms at intake, using the same requirement that the elevated blood pressure be shown at two successive examinations.

By applying the criteria of other studies to the ATC HCS data, comparisons could be made of prevalence and incidence in the differing population groups.

Using the ATC HCS criteria, 135 men, or 32.5% were found to have hypertension at intake. Thirty-six additional men (5.7% annualized rate) developed hypertension over the course of the study.

Applying the criteria for determining the condition used in each comparative study to ATC data, the following comparisons of prevalence and incidence emerged:

1) By HANES (U.S. Health and Nutrition Examination Survey, 1971-74) criteria, 55% of the ATCs were hypertensive (both borderline and definite) at intake, a rate 1.5 times that of the white U.S. males of the HANES population.
2) By Framingham Heart Study criteria, 64% of the ATCs were hypertensive at intake; for the category of definite hypertension alone, this was a rate 1.5 times that of males of comparable age in Framingham, Massachusetts. Under the same criteria, 64% became hypertensive over the 3 years of the study (all categories of hypertension) producing a rate 4 times that of the Framingham males.
3) By FAA criteria, which allowed higher blood pressures with increasing age and measured the blood pressure with the person in a supine position after rest, only 11% of the ATCs were hypertensive at intake.
4) Comparisons with the Western Collaborative Group Study, 1964, applying the Framingham criteria for definite hypertension, showed that the ATCs prevalence of hypertension may be 2 to 4 times that of the Western Collaborative population of middle class men working in above average jobs.

The various comparisons led to the conclusion that air traffic controllers have a substantially elevated risk of hypertension.

d. Hypertension

Hypertension is known to be a condition having the potential for future very serious end-organ damage and even the shortening of life expectancy. Its prevalence* was previously reported to be considerably higher among air traffic controllers than among second-class airmen (Cobb and Rose, 1973). We therefore selected our methodology carefully, utilizing the experience of our project physician who had been, and is still, a medical examiner for the Framingham Heart Study. We selected several criteria for the diagnosis of hypertension in order to enable us to compare our findings with those of two major epidemiological investigations of blood pressure levels and hypertension: the 1971 - 1974 Health and Nutrition Examination Survey (HANES) conducted by the U.S. Public Health Service using a nationwide probability sample of the population (DHEW Publication No. (HRA) 78-1648, Sept. 1977, pp. 2-4, 18) and the Framingham Heart Study (Kannel and Gordon, 1974, Dec. 30, pp. 31-32), a longitudinal surveillance of a sample of the adult residents of Framingham, Massachusetts. We also utilized, for comparison purposes, the diagnostic criteria for hypertension established by the FAA for recertification of ATCs at their annual medical examination. In addition to these three sets of diagnostic criteria from other groups, we developed our own diagnostic criteria for the ATC HCS. As described later, these criteria gave us a more stable diagnostic decision based on brief and well specified examination procedures which are equivalent at all ages. The methods of examination and rules for making of diagnoses according to these four different criteria are detailed in Table 113.

Some comments on the strengths and weaknesses of the various diagnostic criteria are in order. The HANES criteria relies on a single reading of blood pressure at a single examination for the determination of hypertensive status. It therefore would be expected to generate the highest percentage of prevalence cases if, as is usually assumed, first reading of blood pressure tends to be higher than subsequent readings. On the other hand, when more than one reading is used and the elevation of any reading can result in the diagnosis of hypertensive status, such as obtains for the Framingham Heart Study criteria, lability of blood pressure would produce a large enough number of borderline prevalence cases.

---

* Any study participant who had experienced the conditions listed in Table 102 at any time before entry into the study, or for whom the condition was diagnosed at entry, was considered a prevalence case.

TABLE 113

Four Sets Of Criteria For Determination Of Hypertension

	Readings (Syst. & Diast.)	Definite HT	Borderline HT	Normotensive
Framingham[1]	2 Seated, left arm, or 1 if second reading not available, at one exam.	Either syst. ≥ 160mm Hg or diast. ≥95mm Hg on both readings	Any combination of two readings other than those meeting criteria for definite hypertensive or normotensive	Both syst. < 140mm Hg and diast. < 90mm Hg on both readings
HANES[2]	1 seated, right arm, at one exam	Either syst. ≥160mm Hg or diast. ≥ 95mm Hg	syst. < 160mm Hg and diast. < 95mm Hg but not both < 140mm Hg syst. and 90mm Hg diastolic	Both syst. < 140mm Hg and diast. < 90mm Hg
FAA[3]	# Not specified, reclining BP used	Must require no medication for control of blood pressure and must demonstrate BP levels no greater than the following maximum reclining values: Systolic Diastolic Age 20-29 140mm Hg 90mm Hg 30-39 150mm Hg 90mm Hg 40-49 150mm Hg 100mm Hg 50-over 160mm Hg 100mm Hg		
ATC HCS	Average of 2 seated right arm readings at each of 2 successive examinations	Either systolic ≥140mm Hg or diastolic ≥ 90mm Hg on two successive visits approximately 9 months apart		Any ATC who does not meet the criteria for diagnosis of Hypertension.

1 Kannel and Gordon, Feb., 1974, Section 30, pp. 31-32.

2 DHEW Publication No. (HRA) 78-1648 Sept. 1977, pp. 2,3,4,18.

3 Air Traffic Control Specialist Health Program, Dept. of Transportation, FAA, Order 3930.3, May 3, 1977, Appendix 1, p.5-6.

Blood pressures taken of persons lying down are systematically lower than those taken in the sitting position. Extending the time of recumbent rest before taking the blood pressure would also lower the readings further. For these reasons, the FAA criteria would tend to generate fewer cases. In addition, the usually observed increase in hypertension prevalence by age would be negated, if not reversed, by the different blood pressure elevations required at different ages for the diagnosis of hypertension by FAA criteria. In general, FAA criteria would designate fewer people as hypertensive than either of the other criteria described above.

The ATC HCS criteria developed by the investigators of this study calls for the averaging of two readings of the blood pressure, thus reducing the likelihood that novelty effect or lability will falsely lead to the diagnosis of hypertension. The observation in this study and elsewhere that blood pressure shows variability from day-to-day also led us to set the requirement that blood pressure must be elevated at two successive examinations in order for a diagnosis of hypertension to be made. All these provisions of the ATC HCS diagnostic criteria tend to reduce the frequency with which hypertension is diagnosed and hence, for the same group of blood pressure readings, the rates of prevalence and incidence of this condition would be lower than obtained from either the HANES or the Framingham criteria.

For all of these diagnostic criteria, the incidence rates for hypertension would be expected to be influenced by the prevalence rate in that the number of normotensives remaining at risk for the determination of incidence would be reduced by the number of prevalence cases designated at intake. In addition, criteria like the HANES, which tend to designate persons with normal but higher and more labile blood pressures as "borderline hypertensives" remove these high risk people from the group to be followed and hence, may tend artificially to reduce future incidence rates.

The comparative findings regarding the prevalence of hypertension in the ATC HCS using the four diagnostic criteria described are displayed in Table 114. The high prevalence rates of hypertension as judged by the HANES and the Framingham criteria were very striking. By the HANES criteria, 27% of the men were borderline hypertensive and an additional 28% are definite hypertensives for a total prevalence of 55%. The Framingham criteria developed an even higher prevalence of hypertension but two-thirds of it was in the borderline category. Forty-four percent of men received this diagnosis and nearly 20% additional were considered definite hypertensives for a total of nearly 64% prevalence by this system. Neither the FAA criteria nor the ATC HCS criteria distinguish between borderline and definite hypertension (see

TABLE 114

Comparative Prevalence of Hypertension Among Air Traffic Controller Sample at Intake to ATC HCS by Four Alternative Sets of Criteria

Criteria Applied to ATC HCS blood pressure data	Prevalence among ATCs in Sample					
	Borderline Hypertensive		Definite Hypertensive		Total Hypertensive at Intake	
	#Men	%	#Men	%	# Men	%
HANES	113	27.2%	117	28.1%	230	55.3%
FRAMINGHAM HEART STUDY	184	44.3%	81	19.5%	265	63.8%
FAA	N/A		N/A		47	11.3%
ATC HCS	N/A		N/A		135	32.5%

Table 115). By the FAA criteria, only 11% of men were hypertensive, but by the ATC HCS criteria 32% received this diagnosis.

Further studies of blood pressures and of different criteria for diagnosing hypertension were done using blood pressure determinations from the second through fifth medical examinations conducted at BUSM. This provided the means for checking the stability of the diagnoses made by different criteria. Application of the HANES and Framingham criteria to our data demonstrated the instability of prevalence and incidence rates for borderline hypertension generated by the use of only one blood pressure reading or the higher of two such readings at a single examination. Use of the first reading alone produced many "hypertensives" who were responding to the novelty of the examination procedure with an elevation of pressure and whose blood pressures at ensuing examinations were normal. Two readings of blood pressure at a single exam and application of the Framingham criteria, produced many cases of "borderline hypertension" on the basis that one of the readings was above the normal cut-off, even though the other was clearly in normotensive range. Examinations on later dates showed a consistent return to normal pressures in most of these men. Definite hypertension diagnosed on the basis of systolic pressures 160 mm Hg or greater or diatolic pressure of 95 mm Hg or greater was identified a great deal more consistently under any of the diagnostic criteria.

Because we intended to perform additional predictive studies of those who became hypertensive during the course of the ATC HCS, we sought a more stable and conservative indicator of hypertensive status. Hence, we adopted the ATC HCS criterion which specified that hypertension would be diagnosed only upon elevated blood pressure (the average of two readings while sitting: sbp equal to 140 mm Hg or greater or dbp equal to 90 mm Hg or greater) at two consecutive examinations, which normally occurred approximately nine months apart. Additional grounds for the diagnosis of hypertension <u>at intake</u> were the determination by our study physician that the participant (1) had a history of being diagnosed as hypertensive by a physician <u>and</u> at intake had elevated blood pressure (140/90 as defined above); or (2) had a history of being diagnosed as hypertensive by a physician <u>and</u> was currently adhering to a regimen of anti-hypertensive medication. Our ATC HCS criteria for diagnosis of hypertension were therefore considerably more stringent than either the HANES or the Framingham Study. In addition, they differed qualitatively from the criteria designated by FAA regulations.

We decided not to adopt, for our research purposes, the FAA criteria used for recertification of controllers upon annual examination. Our decision was based partially on the fact that the FAA acceptance of elevated blood pressures as normotensive among older controllers was not appropriate to our study goal, which was to identify those controllers whose blood pressure levels became elevated during the course of the study and to discover the antecedents of this process. Another reason for not using this criteria was that the use of supine blood pressures and the lowest of an unspecified number of readings made comparisons with other studies difficult and the resemblance to daily blood pressures at work too remote.

The incidence of new cases of hypertension developing among men judged normotensive at intake into the ATC HCS is displayed in Table 115. In view of the very high prevalence rate of hypertension judged to be present when the Framingham criteria were applied, only 150 of the 415 men for whom complete diagnostic data at intake was available were considered normotensive and at risk for new disease. Of these, 14 developed definite hypertension and 83 developed borderline hypertension over the course of the next four examinations utilizing the Framingham criteria. This represented a combined incidence of 65% over a three-year risk period. This seems definitely out of keeping with clinical experience even with presumably high risk groups. It must be remembered, however, that "borderline hypertension" would be diagnosed by these criteria if either of two readings at any of the four subsequent examinations ventured above 140 or 90 respectively. But this diagnostic criterion does not seem appropriate for clinical or administrative guidance regarding what is clearly a serious problem. Still, these data do provide the basis for comparison with other groups of men who were examined repeatedly using the same methodology and decision rules.

Use of the ATC HCS criteria yielded 36 new cases of hypertension over the ensuing four examinations. When one considers that a man must have been normotensive at the intake examination and hypertensive at two successive examinations to qualify under this criterion, it becomes apparent that the actual period at risk was not three years, but rather about 2.25 years. This is true because hypertension would have to be first diagnosed at exam 4 in order to be confirmed by the time of the last examination of the study. This generates an annualized incidence rate of 5.7% per year which is still quite substantial considering the stringency of the diagnostic criteria.

Comparisons of Prevalence of Hypertension with Other Populations

Although many studies have been published with data on prevalence of hypertension among different community and industrial groups, the differences in diagnostic criteria and the age, sex and race distributions of the study population (or lack of speci-

TABLE 115

Comparison Of Incidence Of Hypertension
ATC HCS Study Vs Men Of Comparable Age
Framingham Study

	Framingham Study	ATC Study	
Age Range in This Analysis	30-39 years	Mean age: 36 years	
Total Men Examined	N/A	415	
Hypertensive at Intake			
Framingham Criteria			Prevalence
"Definite"		81	19.5%
"Borderline"		184	44.3%
Normotensive at Intake	362	150	
	Status after 6 years Biennial Exams	Status after 3 years Nine monthly exams	
Normotensive	253	53	
Borderline hypertensive	100	83	
Definite Hypertensive	9	14	
Annualized Incidence Rates/100			
Borderline Hypertension	4.6	18.4	
Definite Hypertension	.4	3.1	
Total Hypertension	5.0	21.5	

fication thereof) make it very difficult to derive comparisons that would be even partially satisfactory. The most adequate sources for comparison remained the DHEW's Health and Nutrition Examination Survey (HANES) of 1971 to 1974 and the Framingham Heart Study. In view of the fact that we had recorded separately each of the several blood pressure readings at each examination, we were able to make rather precise comparisons of our findings using diagnostic criteria to fit the comparison study. Inasmuch as our sample consisted totally of males, 99% of whom were white, we could make comparisons with sex-race specific data from these epidemiologic studies. Comparisons of hypertension prevalence in the ATC HCS with that of white males in other populations by different diagnostic criteria are presented in Table 116. These data reveal a considerably greater prevalence of hypertension among our ATC sample whenever similar criteria are used to determine cases. This is true no matter which criteria are applied. The ATC prevalence of hypertension (both definite and borderline) was 1.5 times that of white males in the U.S. (HANES data). The prevalence of definite hypertension, the more severe segment of the total hypertensive group, by HANES criteria was 1.6 times higher among ATCs than men in the general population. Use of the Framingham version of "definite hypertension" yielded a 1.5 times excess among ATCs as compared with the population of that Massachusetts town.

The Western Collaborative Group Study (Rosenman et al., 1964) also measured blood pressure levels in 3,524 men employed in white collar, engineering and managerial positions in ten large industries in California. The WCGS population was similar in occupational prestige, but somewhat higher on the average in education as compared to the ATC study group. The WCGS men, however, were considerably older than these ATCs. Using the Framingham criteria for definite hypertension, WCGS men aged 50 - 59 years at intake (average age 53.7 years) had a hypertensive prevalence of 13.9% and men ages 40 - 49 years at intake (average age 43.2 years) had a hypertensive prevalence of 8.4%. If one then makes a linear extrapolation to the average age of 36.2 years (which was the average for our ATC group) one can estimate by graphic extrapolation what the prevalence rate of hypertension would have been in the WCGS if their younger group had averaged 36.2 years rather than 43.2 years. The rate derived by this estimation is a prevalence of 4.8%. The observed prevalence in the ATC study using comparable diagnostic criteria, was 19.5% or 4.0 times the age specific estimate derived from the very large WCGS study. Even if linear extrapolation is not an accurate procedure, it should be noted that prevalence of definite hypertension in the ATC study is still more than twice the rate of that observed for the WCGS men ages 40 - 49 years even though the latter group averages 7 years older than the ATC group. This set of comparisons suggests

TABLE 116

Comparison Of Prevalence Of Hypertension Among ATC HCS Sample At Entry To Study, Male Residents Of Framingham And White Males In The United States

	Age	% with Definite by Framingham Criteria (Syst. ≥ 160mm Hg or diast. ≥ 95mm Hg)	% with Definite HT by HANES criteria (Syst. ≥ 160mm Hg) or diast. ≥ 95mm Hg)	% with Hypertension by HANES criteria (Syst. ≥ 140mm Hg or diast. ≥ 90mm Hg	% with Hypertension by ATC HCS criteria (Syst. ≥ 140mm Hg or diast. ≥ 90mm Hg)
Framingham Males (1948)	35-39	13.3%[1]			
HANES white males (1971-74)	35-44		17.3%[2]	35.9%[3]	
ATC HCS Males	25-48 (av. 36)	19.5%	28.1%	55.4%	32.5%
Relative Risk (ATC/other)		1.47	1.62	1.54	

1
Kannel, William B., M.D. and Gordon, Tavia, editors. The Framingham Study, An Epidemiological Investigation of Cardiovascular Disease, Section 29. Means at each examination and interexamination variation of specified characteristics: Framingham Study, Exam 1 to Exam 10. Tavia Gordon and Dewey Shurtleff. December, 1973, Table A-22. (DHEW Publ. No. (NIH) 74-478).

2
Blood Pressure Levels of Persons 6-74 years, United States, 1971-1974, U.S. Dept. Of Health, Education, and Welfare, Public Health Service, Hyattsville, Md., September 1977, Table 40, p. 72 (DHEW Publ. No. (HRA) 78-1648).

3
DHEW Publ. No. (HRA) 78-1648, Tables 40 and 42, pp 72,74.

that the prevalence of hypertension in the ATC may be 2 to 4 times that of middle class men working in above average jobs.

Comparisons of Hypertension Incidence Data

There is very little information on incidence of hypertension in the U.S. Repeated measures of the same population at annual intervals are a prerequisite for such data. Rough estimates of the incidence of hypertension, however, have been offered by experts in the field on the basis of their reviews of published epidemiologic data (Stamler, 1968, p. 152; Julius and Schork, 1971). Incidence was estimated to be within the range of approximately 1% to 4% per year among persons approximately 30 to 60 years of age.

In order to obtain a more precise estimate of incidence of hypertension in a comparable population, the ATC HCS investigators obtained access to printouts of hypertension data from the Framingham Heart Study. We are indebted to Dr. H. Emerson Thomas, Co-principal Investigator; Mrs. Pat McNamara, Research Associate; and to Dr. Bernard E. Kreger, Examining Physician, for making these computer printouts available to us for this specific purpose. The printouts contained a listing of all normotensive entrants into the Framingham Heart Study by secret I.D. number, age, sex, hypertension status at intake and hypertension status at each of the next several biennial examinations. By re-checked hand counting we determined that there were 362 normotensive males, ages 30 - 39 years entering the Framingham Study. After 3 biennial examinations (six years of follow up) 253 remained normotensive throughout the period, 100 showed borderline hypertension on one or more occasions (but not definite hypertension), and 9 showed definite hypertension. This yielded annualized incidence rates per 100 men of 4.6 for borderline, 0.4 for definite, and 5.0 for total hypertension. These data are presented in Table 115. By comparison, applying the same Framingham criteria to our ATC men and making appropriate adjustments for the differential time at risk, the annualized incidence rates per 100 men in the ATC HCS were 18.4 for borderline hypertension, 3.1 for definite hypertension, and 21.5 for total hypertension. For all categories of hypertension, ATC men showed at least a four-fold increased risk as compared to Framingham men.

Even when the more stringent diagnostic criterion developed in our own study was used for the diagnosis of hypertension incidence, an annualized rate of 5.7 per 100 per year is generated. Despite the stringency of the criterion, this is approximately 3 times the incidence estimated for the general population by reviewers of the epidemiologic literature.

Our general conclusion, then, is that air traffic controllers have a substantially elevated risk of hypertension and this is observed both when a cross-sectional measure is taken at intake or when incidence figures are calculated in longitudinal study. We estimate that the prevalence of hypertension among ATCs is between 1.5 to 2 times that of the general population, but that incidence rates are from 3 to 4 times that of comparable men in other lines of work or in an average community. Hypertension is thus the most serious health problem among air traffic controllers, and high priority should be placed on discovering the circumstances responsible for it and types of persons most susceptible to these circumstances.

SUMMARY

Hearing Impairment

Air conduction audiograms and speech discrimination tests performed by a professional audiologist at Boston University Medical Center revealed the following:

1) ATC mean thresholds at 4,000 and 6,000 Hz were 7.2 dB poorer at 25-29 years of age than non-noise exposed males and the differences increased with age;
2) High frequency hearing loss observed in the ATCs is consistent with the audiometric pattern associated with noise exposure;
3) The hearing loss was associated with age, but not with years of experience as a controller;
4) High frequency loss was particularly marked in men over 40 years of age;
5) 98.6% of the ATCs passed word tests in which they were required to repeat pairs of two syllable words, one presented in each ear, in a partially overlapped manner;
6) No psychosocial predictor variables were found to be associated with hearing outcomes;
7) No association was found between hearing impairment in the "good" ear and measures of job-related stress;
8) Men who were hypertensive at intake were significantly more often among those found to have severe hearing losses at Round 3.

e. Hearing Impairment

Air conduction audiograms were done at Daniels Hearing Center of University Hospital, Boston, Massachusetts on 351 subjects enrolled at Round 3 of the ATC study. Bone conduction thresholds were obtained when air conduction was abnormal. Speech discrimination ability was tested at 40 dB above the pure tone average using Northwestern University Auditory Test No. 6. All testing was administered with 50 dB of speech noise in the non-test ear. The Staggered Spondaic Word Test was administered at 50 dB HTL or 40 dB above the pure tone average if thresholds were greater than 25 dB.

The air traffic controllers' mean thresholds at 4,000 and 6,000 Hz were 7.2 dB poorer at 25-29 years of age than non-noise exposed males. The difference between controllers' thresholds and the NIOSH data increases with each age group, falling 16.5 dB below the norms at age 45-49. Many of the controllers have a history of military service and employment in various aspects of aviation. Our data has not yet been analyzed to compare the hearing thresholds of controllers with other occupation or recreation-related noise exposure to those with no history of exposure prior to work experience in this field. However, our test results indicate that the air traffic controllers had significantly poorer thresholds at 4,000 and 6,000 Hz than non-noise exposed males of the same age. The configuration of controllers' composite audiograms shows the greatest loss at 6,000 Hz for all age groups with a mean recovery of 4 dB at 8,000 Hz. A high frequency hearing loss of this type is consistent with the audiometric pattern associated with noise-induced hearing loss.

The mean air conduction threshold for all 702 ears was compared with NIOSH median hearing threshold for non-noise exposed males by frequency. These data are presented in Table 117.

Hearing loss at each frequency was studied in relation to the ATCs' ages and years of experience in the FAA. These two variables are, of course, highly correlated. When years of experience in the FAA was controlled for age, it was not significantly associated with hearing loss. However, when age was studied, with experience controlled, age remained strongly associated with hearing loss. High frequency loss was particularly marked in men over 40 years of age, more so than one would have predicted by extrapolating from younger ATCs. There is reason to suspect that many of this older cohort had exposure to intense noise early in their working

TABLE 117

Mean Air Conduction Threshold Of
351 ATCs And A Group Of Non-Noise Exposed Males

FREQUENCY Hz		500	1000	2000	3000	4000	6000
Age Group							
25-29	NIOSH Data*	11.0	5.4	3.8	5.6	7.4	10.8
	ATC Mean	8.1	6.6	7.0	8.6	14.3	18.2
30-34	NIOSH	11.8	6.0	4.6	7.0	9.8	13.6
	ATC Mean	9.2	7.3	7.6	12.9	18.0	20.8
35-39	NIOSH	12.6	7.0	5.6	9.0	12.4	16.6
	ATC Mean	8.4	7.2	7.9	13.1	21.8	24.2
40-44	NIOSH	13.6	7.6	6.6	11.2	15.6	20.4
	ATC Mean	10.6	10.2	12.2	20.5	29.6	33.0
45-49	NIOSH	14.4	8.4	8.0	13.8	19.4	24.4
	ATC Mean	13.3	14.0	14.0	23.3	35.8	41.0

* NIOSH Madian Hearing Levels for non-noise exposed males published in Criteria for a Recommended Standard...Occupational Exposure to Noise, U.S. Dept. HEW,1972.

lives (perhaps before entering the FAA) to a degree not experienced by the younger men.

An audiologic test of particular interest is the Staggered Spondaic Word Test. This test involves verbal repetition of 2 two-syllable words, one presented to each ear in an overlapping manner. This task is particularly relevant to the competing noise and messages encountered daily in rapid sequence by ATCs. It is encouraging to note that 98.6% of our ATC participants passed this test.

The audiometric results for the worse of the two ears for each man were considered to be an index of the full extent of the pathology of hearing loss and was utilized as a health outcome measure. None of the psychosocial predictors from Rounds 1 and 2 were associated with this outcome.

The result for the better ear for each man was considered a predictor. It was hypothesized that a man who had severe hearing loss in his "good ear" would encounter more job-related stress because of this limitation. Statistical analyses of the psychosocial stress measure did not reveal significant associations.

One finding of interest was the relation of hypertension to hearing loss. This may have been influenced in part by age. Men who were hypertensive at intake into the study were significantly more often among the men found to have more severe hearing losses at the Round 3 audiologic exam. This association held both for the better and worse ear for the men. One possible implication of the finding is that if it is administratively unwieldy to provide full audiologic examinations to all ATCs, these should at least be provided for men with higher blood pressures.

The audiology examinations were performed by Nancy J. Miller, M.A., CCC. in Audiology, Director of the Daniels Hearing Center.

SUMMARY

Frequent Complaints: Sleep Problems and Headaches

Headaches were the most common health complaints of the ATCs, and sleep problems were second most common. Thirty-five percent of the men reported that they had sleep problems four or more times a month. When all sleep problems were considered, the two most frequent explanations offered were: "Too many ideas spinning through my mind" (given by 70%) and "changes in sleeping schedule such as changes in work shifts" (given by 69%).

Responses to an inquiry as to their best and their worst time of day indicated that the ATCs varied greatly in the time of day when they considered themselves to function best.

In response to a questionnaire, 48% of the men reported headaches during one month lasting more than 1 hour, and 33% of these had headaches so severe as to stop their normal activity. However, only 6% had the more severe headaches on three or more days per month. Although headaches were a common problem among the ATCs, they were, for the most part, not disabling.

f. Frequent Complaints: Sleep Problems and Headaches

Sleep Problems

Trouble getting to sleep, or staying asleep until the desired time for awakening, was the second most common health complaint reported monthly by participants in the study, exceeded only by the frequency of prolonged headaches. Sleep problems were reported as occasional difficulties by 15% of the men each month and as a continuing problem occurring more than half the time by another 10% of the participants. No systematic seasonal changes in the prevalence of sleep problems were noted over a two-year period. The substantial prevalence of sleep problems and their implications for health and effective functioning as an air traffic controller led us to a more detailed study of this problem in February, 1977.

The Sleep Questionnaire

A one-page questionnaire regarding sleep problems and their presumed causes was added to the monthly mailings sent to 315 ATCs active in the study at that date. Two hundred sixty questionnaires were returned with sufficiently complete data for analysis. This represented an 83% completion rate.

Respondents were asked to indicate how often during the past month they experienced each of four kinds of sleep problems:

a. trouble falling asleep,

b. trouble staying asleep (awaking far too soon and not being able to get back to sleep),

c. waking up several times per night,

d. waking up after their usual amount of sleep feeling tired and worn out.

Findings

The response distributions to each of these four questions were remarkably similar. About one-third of the men reported that they had not encountered the problem at all. Another third indicated they had had sleep problems only from one to three times during the month; about one-fourth reported that they had these troubles on four to fourteen occasions; and about 10% reported that they had them fifteen or more days during the month. These data are displayed in Table 118.

TABLE 118

Frequency Of Reported Sleep Problems

Among 260 Air Traffic Controllers

ype of Problem	How often occurring this month? Not at all	1-3 days	4-14 days	15+ days	Total
rouble falling asleep	32.4%	31.3	26.2	10.2	100%
rouble staying asleep ncluding "early waking"	37.0	26.3	26.7	10.0	100%
ke up several times per ght	34.1	31.4	22.4	12.2	100%
ke up feeling tired and rn out	27.8	32.1	31.4	8.8	100%

Concurrence of Different Sleep Problems

Analysis was undertaken to see whether the four different types of sleep problems occurred independently of each other or in combination in the same individuals. The number of nights a participant was troubled by each problem in the reported month was stratified into four levels: not at all, 1-3 days, 4-14 days, and 15 or more days. Cross-tabulations of the frequencies of sleep problems by each other revealed a strong degree of association. The chi squares generated amongst the six possible pairs of these four variables ranged from 38 to 200 (with nine degrees of freedom)—all highly significant. The gamma statistic, a measure of association, ranged from +.38 to +.77. These data are presented in Table 119.

The association is illustrated, for example, by the 79 ATCs who had trouble falling asleep only 1-3 days in the past month, and of whom only two (2.5%) reported trouble staying asleep throughout the night on 15 or more occasions. In contrast, of the 27 men who had trouble falling asleep on 15 or more nights of the month, 13 (30%) also had trouble staying asleep during the same number of nights. As one might suspect, the strongest association was between waking far too early and waking up several times during the night. The lowest associations--though still highly significant--were between waking up after the usual amount of time feeling tired and worn out and the three other forms of insomnia.

ATC Explanations of Their Sleep Problems

Considering all sleep difficulties collectively, the most frequent cause stated by respondents to the questionnaire was: "too many ideas spinning through my mind." Seventy percent of men offered this explanation. "Changes in sleeping schedule such as changes in work shifts," was offered as an explanation by 69%. Fifty-five percent thought worries or problems contributed to sleep difficulties, and the remaining suggested causes were all endorsed much less frequently.

It would seem reasonable to expect that different manifestations of sleep problems would be associated with different beliefs as to their causes. Similarly, it seemed quite likely that presumption of cause might differ systematically between people who rarely had a sleep problem and those who had one very often. To answer these questions, cross-tabulations were made between the monthly frequency of the problem and the reported causes, separately for each type of sleep problem.

Respondents who had trouble falling asleep four or more days

TABLE 119

Intercorrelations of Four Types of Sleep Problems

in a Specified Month for 256 Air Traffic Controllers*

		(2)	(3)	(4)
(1) Trouble Falling Asleep		.69	.50	.44
	X^2	146.5	73.5	57.2
(2) Trouble Staying Asleep (including waking too early)			.77	.43
			200.7	50.3
(3) Wake up Several Times per Night				.38
				38.0
(4) Wake up After Usual Amount Feeling Tired and Worn Out.				

*Of 260 men, only the 256 who gave some answer (ranging from "not at all" to "22-31 days") to each of the four questions involving sleep problems were included in this analysis.

during the month were more likely than men with lesser frequencies of this problem to attribute it to changes in schedule or worries or problems. Men who reported trouble falling asleep 4-14 days in the past month more often attributed the problem to "things eaten or drunk" before retiring than did men who had this problem only rarely or those who had it 15 or more days per month. Persons who reported waking up feeling particularly tired and worn out four or more days in the previous month more often attributed their excess fatigue to worries or problems than did persons who experienced this problem only rarely. It should be noted that persons were permitted to endorse as many possible causes for their sleep problems as they felt might apply and that only a minority of those having such problems reported that they had no ideas as to the causes. Nevertheless, there was an association between high frequencies of problems and perplexity as to their cause. Respondents who reported trouble staying asleep, waking up several times per night, or waking up feeling tired or worn out on 15 or more days in the past month more often had no ideas about the cause of these problems than did persons who reported these problems less frequently (Tables 120 - 123).

The hypothesis that different types of sleep problems would be associated with different explanations of cause was not supported. For these analyses, we considered only those respondents who reported a given problem four or more times per month. For all four sleep problems, the most common causes cited were: changes in schedule, too many ideas spinning through one's mind, and worries or problems. Change in schedule was the most frequently cited reason for trouble falling asleep, trouble staying asleep, and waking up after rest feeling worn out. Ninety-one percent of persons who woke up feeling fatigued 15 or more days in the reported month believed changes in schedule to be the primary reason for their problem. "Too many ideas spinning through my mind" was most frequently cited by those who woke up frequently during the night. "Worries or problems" was, for most groups and most types of sleep difficulty, the third most frequently cited cause.

Best and Worst Time of Day for ATCs

At the same time that sleep problems were inquired about, participants were asked to indicate which time of day was best for them and which time of day they found most difficult. The task was introduced as follows:

"Some people feel far better--more alert, competent, and happy--at one time of day,

TABLE 120

Reported Causes Of Sleep Problem

By Type And Frequency Of Problem

"Trouble Falling Asleep"

	Monthly Frequency of Problem N=78 1-3 days	N=65 4-14 days	N=25 15+ days	signif-icance*
hings eaten or drunk	15/76 .20	27/63 .43	6/22 .27	.01
epeated use of the athroom	13/76 .17	12/61 .20	3/22 .14	---
oo many ideas spinning hrough head	58/78 .74	55/65 .85	21/26 .81	---
orries or problems	36/75 .48	44/62 .71	21.25 .84	.002
hanges in schedule	59/79 .75	57/66 .86	21/25 .84	.004
t allowing enough time	25/76 .33	28/61 .46	8/22 .36	---
ideas about cause	5/61 .07	8/33 .13	7/20 .32	.01

dicates probability by chi square that the group with different fre-
ncies of sleep problem differ in rates of endorsing this "cause."

TABLE 121

Reported Causes Of Sleep Problem

By Type And Frequency Of Problem

"Trouble Staying Asleep"

	Monthly Frequency of Problem N=62 1-3 days	N=60 4-11 days	N=25 15+ days	signif-icance*
Things eaten or drunk	19/62 .31	25/60 .42	5/25 .20	---
Repeated use of the bathroom	18/62 .29	9/58 .16	3/25 .60	---
Too many ideas spinning through head	52/66 .79	54/64 .69	18/25 .72	---
Worries or problems	36/62 .58	44/62 .71	17/24 .71	---
Changes in schedule	46/66 .70	52/64 .81	19/25 .76	---
Not allowing enough time	24/62 .39	25/60 .42	7/23 .30	---
No idea about causes	5/44 .11	7/37 .19	8/20 .40	.03

*Indicates probability by chi square that the group with different frequencies of sleep problem differ in rates of endorsing this "cause."

TABLE 122

Reported Causes Of Sleep Problem

By Type And Frequency Of Problem

"Wake Up Several Times Per Night"

	Monthly Frequency of Problem N=75 1-3 days	N=53 4-14 days	N=29 15+ days	signif-icance*
Things eaten or drunk	29/75 .39	18/53 .34	8/29 .28	---
Repeated use of the bathroom	14/72 .19	15/54 .28	8/30 .27	---
Too many ideas spinning through head	61/79 .77	46/56 .82	23/30 .77	---
Worries or problems	48/76 .63	35/54 .65	18/28 .64	---
Change in schedule	56/80 .70	41/55 .75	20/30 .67	---
Not allowing enough time	28/74 .38	26/54 .48	7/26 .27	---
No idea about cause	3/53 .06	8/37 .22	10/21 .48	.000

* Indicates probability by chi square that the groups with different frequencies of sleep problem differ in rates of endorsing this "cause."

TABLE 123

Reported Causes Of Sleep Problem

By Type And Frequency Of Problem

"Wake up After Usual Amount Feeling Tired"

	Monthly Frequency of Problem N=77 1-3 days	N=74 4-14 days	N=19 15+ days	signif-icance*
Things eaten or drunk	25/77 .32	25/74 .34	4/19 .21	---
Repeated use of the bathroom	17/75 .23	16/74 .22	7/19 .37	---
Too many ideas spinning through head	54/78 .69	60/78 .77	18/22 .82	---
Worries or problems	36/74 .49	55/74 .74	15/22 .68	.005
Change in schedule	55/79 .70	58/77 .75	20/22 .91	---
Not allowing enough time	32/76 .42	29/73 .40	13/20 .65	---
No idea about causes	3/51 .06	10/52 .19	4/15 .27	.05

*
Indicates probability by chi square that the group with different frequencies of sleep problem differ in rates of endorsing this "cause."

> but noticeably below that level at other times. You hear someone saying; 'I'm a morning person' and another: 'I'm a night owl.'
> For each description, please check the one set of hours which best fits."

The following choices of hours were provided to be checked as the best time of day or the most difficult time of day: 7 a.m.-12 noon, 12 noon-5 p.m., 5 p.m.-10 p.m., 10 p.m.-7 a.m., and "all times are the same."

There was a surprising spread of nominations for the best time of day and the most difficult time of day. Morning hours were most commonly picked in both of these categories (see Table 124). Thirty-six percent of the men felt morning was best for them, about one quarter of respondents picked afternoon, and another quarter picked evening hours, while only 5% perceived the midnight hours as their best time.

With respect to the most difficult time of day, nearly 30% picked morning, 25% picked the midnight hours, about 19% picked afternoon and the same percent picked evening. About 10% of the men said that all times of day were about the same for them. Editing of these data suggested that the questionnaires were answered thoughtfully and meaningfully. None of the 244 respondents to this series of questions checked the same time of day as being both best and worst. Cross-tabulations of answers to these two questions revealed that men for whom the morning hours were best tended more often than the total respondent sample to pick afternoons and evenings as their worst time of day. People for whom evenings (5 p.m.-10 p.m.) and midnight hours (10 p.m.-7 a.m.) were best times of day tended rather strikingly to say that morning was the most difficult and clumsy time for them.

Conclusions

These findings suggest that not all people are equally comfortable or equally capable of working the several shifts which ATC work requires. It would be worth investigating the benefits to be gained from assigning persons wherever possible to work at the times of day when they are more energetic, alert, and capable and not assigning them to the times when they feel tired, irritable, or inept. This suggestion is further supported by the analysis of the sleep problems data earlier in this section. Here it was found that changes in sleeping schedule, such as caused by changes in work shifts, were among the most common reasons given for frequent problems in falling

TABLE 124

Preferred And Most Difficult Time Of Day

For 244 Air Traffic Controllers in February, 1977

	All the same	7am-12pm	12pm-5pm	5pm-10pm	10pm-7am
Best time of day	8.6%	36.5	27.5	27.5	4.9
Worst time of day	9.4%	28.7	18.9	18.9	24.2

If the Best Time was:	The most frequently selected Worst Time was:
7am-12pm	12pm-5pm and 5pm-10pm
12pm-5pm	10pm-7am
5pm-10pm	7am-12pm
10pm-7am	7am-12pm

asleep, staying asleep, and waking up refreshed. It would be important to determine what relationship, if any, these verbal reports may have to differences in the quality of work and the percent time actually spent on the boards during morning, afternoon, and night shifts by men who are working preferred vs. disliked time schedules.

The presence of sleep problems was cross-tabulated by most preferred and least preferred time of day. No significant relationships were found. A few trends were noted which further supported the internal consistency of the data but did not provide a basis for further programmatic suggestions.

Headaches

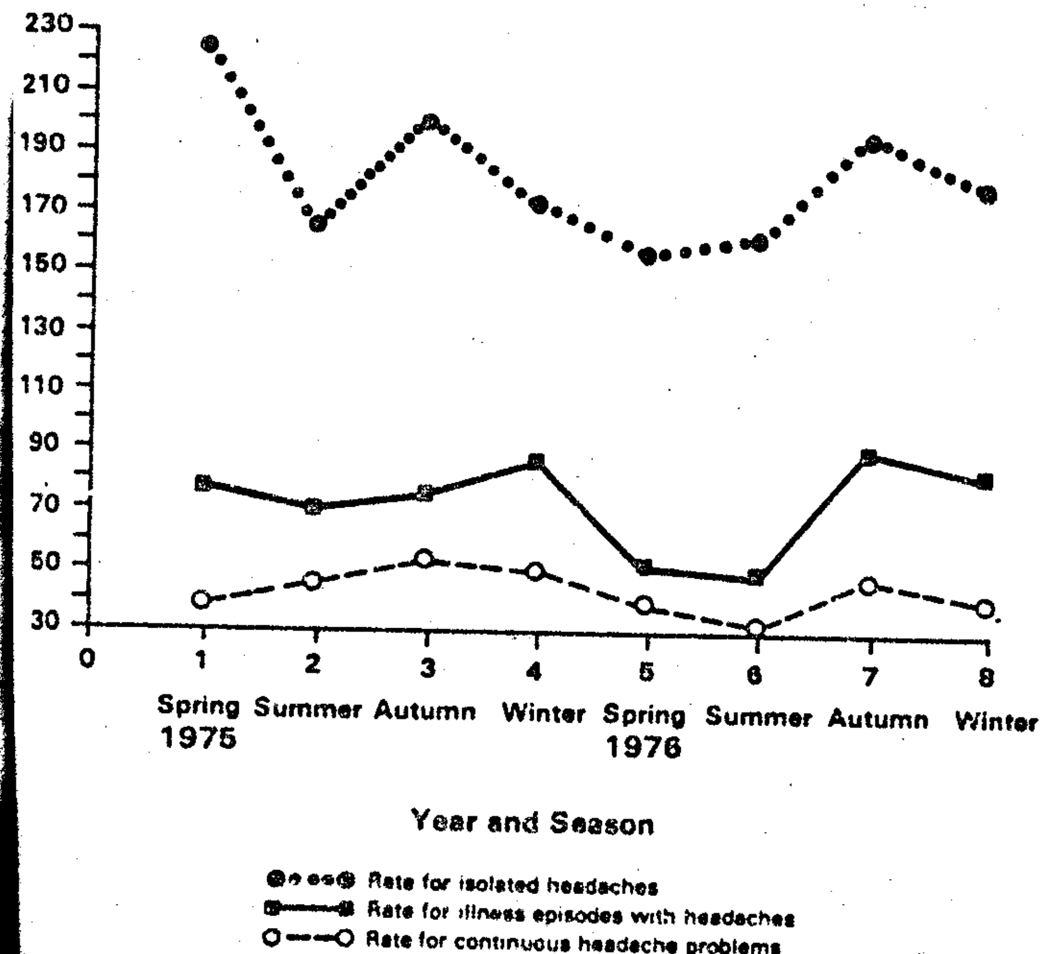
Headaches were the physical health problem most commonly reported on the Monthly Health Reviews returned throughout the study by our ATC participants. In the average month, about 18% of the men reported headaches that lasted more than one hour and were not part of illness episodes. In addition, 4% of the men reported such headaches for at least half the days of the month. There was no regular cycling of headache prevalence by the seasons although there was a tendency for fewer reports of headaches in summers than in autumns. The seasonal prevalence of headaches for 1975 and 1976 is shown in Figure X. These data suggested that headache was quite a problem among the ATCs, but comparison with other occupational groups was impossible because none have been as thoroughly or repeatedly monitored for this symptom as the participants in the ATC HCS.

Because of the high prevalence of headaches, it was deemed advisable to obtain a more detailed picture of the frequency, severity, duration and other physical properties of this common problem. A headache questionnaire was developed after a review of the clinical and epidemiological studies of headaches which have been published in the medical literature. This questionnaire was pilot-tested, revised and then circulated to the entire study group in February, 1977, accompanying the mailing of the Monthly Health Review. The data which follow were derived from the 264 completed questionnaires that were returned. They represent a detailed picture of the frequency, severity and physical properties of headaches experienced during February, 1977.

One or more headaches were reported during the study month by 69% of the 264 respondents. However, only 48% of the men had headaches lasting for more than one hour and only 33% reported headaches so severe that "they made you stop what you were doing." Forty percent of the men reported headaches on three to ten days of the month and 4% reported such symptoms more than 10 days of the month. Prolonged or severe headaches were much less frequent

FIGURE X

SEASONAL PREVALENCE OF HEADACHE (LASTING MORE THAN 1 HOUR)



with only 23% of the men reporting headaches extending more than one hour on three or more days of the month, and only 6% reporting severity sufficient to force a cessation of activities on three or more days per month.

We were interested in determining which among these headaches could be considered migraine, muscle tension, or other types. Considerable work on computer diagnosis of headache type from self-reported symptoms has been performed at a number of medical centers. We were particularly guided by the work of Stead and colleagues (Stead, et al., 1972), and aided by the epidemiologic studies of A.M. Ostfeld (1962) and by another study of computer diagnosis which used criteria similar to those derived by Ostfeld (Freemon, 1968). The diagnostic criteria for various types of headaches used by Stead and his group were based on criteria suggested by the Public Health Service Committee on Classification of Headache. Using these criteria and the scoring cut-offs presented in the paper by Stead, et al., it was determined that 27 of the 264 respondents (10%) were susceptible to migraine headaches, 10 men (4%) were subject to classical muscle tension headaches, 15 men (6%) were susceptible to cluster headaches.

The symptom series and computer logic of the Freemon System for Diagnosing Headaches could not be duplicated exactly. Nevertheless, there was fairly close agreement between the Freemon and the Stead systems with 24 of the 27 migraine cases diagnosed by the Stead system also so classified according to Freemon. Many of the headaches reported by ATCs were not classifiable into strictly defined diagnostic groups. This is probably also the case for headaches presented for diagnosis at Neurology Clinics. The following description of the characteristics of headaches and their circumstances of occurrence is based on all respondents who either had headaches lasting for more than one hour or had such severe ones that they forced cessation of activity. About 112 men gave detailed descriptions of such headache experiences during the month studied.

Just before the headache:

1. 30% of men with at least occasional prolonged or severe headaches reported that the muscles of their head or neck felt tight.

2. 26% said their eyes became sensitive to light, but 27% said that usually they had no prodromal symptoms at all.

Starting location:

1. For 42% with headaches, starting location was around or behind the eyes.

2. For 30%, the headaches usually started at the back of the head or neck.

3. For only 16% was the headache initiated unilaterally.

Usual time of occurrence

1. For over 30% of the men, periods of intense concentration, or periods of worry or emotional tension surrounded the headaches.

Type of pain:

1. Half of the men reported that they usually had a dull steady ache or pain.

2. Thirty percent reported a steady pressure and 27% reported throbbing or pounding.

3. Thirty-one percent reported that bright lights made the headache worse.

Duration of headaches:

1. Since headaches lasting less than one hour were not reported unless severe, the distribution of remaining headaches by duration was truncated. Nevertheless, it was interesting that seven of the men reported that their headaches usually lasted more than eight hours and 20 additional men reported that they occasionally lasted that long. Thus 10% o the total men reporting (with or without headaches) had headaches of eight hours or greater duration, at least occasionally.

Severity

1. As expected, most headaches most of the time did not interfere with ongoing activities. However, 8 men reported that they usually had pain of sufficient severity to force them to stop whatever they were doing, and 37 men reported that they occasionally had such pain. This represented 3% and 14%, respectively, of the total group of respondents.

Relief of Headaches

1. Of the 112 men reporting either prolonged or severe headaches, 58% got adequate relief from aspirin. Twenty-four percent needed stronger non-prescription pain relievers,

while 44% got relief without drugs, through relaxation, massaging the head or neck or other procedures.

It was not possible for us to compare the prevalence and severity of headaches among ATCs with the experience of other groups because published data are derived from patients who presented themselves at general medical or headaches clinics. Although headaches seemed to be quite common among the ATCs, they were, for the most part, of short duration and not disabling. It would be useful to do a more extensive analysis of the 23% of men with prolonged headaches and the 6% of men with very severe headaches, to determine psychosocial and physical characteristics associated with these problems. This type of analysis was not possible in the time available for the present study.

D. Psychological Health Outcomes

While it was as important to assess psychological health outcomes as it was to evaluate physical health outcomes, the methodology for determining psychological health change was neither as standardized as that for diagnosing physical health change, nor was the severity of psychological problems as consistently definable. Nonetheless, we sought a methodology that would provide both objective assessments by psychology professionals and subjective reports by the ATCs themselves of various aspects of psychological functioning both at entry into the study and throughout the 3-year surveillance period. We were interested in determining the prevalence and incidence of psychological problems to gain some understanding of the general level of psychological functioning of men involved in this unique occupation, and we were equally interested in ascertaining whether or not it was possible to identify those who would later develop psychological problems by any predictive measures. We weighed carefully the numerous options available for assessing psychological health - such as clinical interviews, clinical psychological testing, self-report personality inventories, structured interviews, symptom inventories and behavioral indices - and eventually selected an approach that would yield both objective assessments in depth at less frequent intervals and self-reports on symptom check-lists at more frequent intervals. The objective assessments were accomplished by use of the Psychiatric Status Schedule, a standardized structured interview, and the self-reports were transmitted via the Zung Anxiety and Depression Scales accompanying the Monthly Health Review. These instruments and their use are described in the first section below. A description of the criteria for indexing psychological health changes by levels of severity follows in the second section; and the findings of prevalence and incidence of psychological problems are presented next. The findings and discussion concerning early identification of persons prone to psychological problems are presented in Section V, Predictive Findings.

SUMMARY

Methods of Measurement of Psychological Health Outcomes

Psychological health outcomes were assessed in two ways. First, the major psychological health outcomes were evaluated by trained psychologists using the Psychiatric Status Schedule (PSS), a structured clinical interview procedure. The PSS was selected because it had been devised to yield highly reliable assessments of psychopathology and because considerable data from other populations were available for comparative purposes.

Five criterion areas of psychological status were evaluated: subjective distress, impulse control disturbances, work role disturbances, mate role disturbances and alcohol abuse.

Symptomatic cut-off scores were established for each of the five criterion areas of psychological health according to statistical and clinical standards for significance. The cut-off scores for identifying systematic levels of disturbance were set from 1 to 3 standard deviations above the means observed for a normal urban community sample. The cut-off levels were substantially below the means observed for pre-treatment psychiatric outpatients for the subjective distress scale, were approximately equal to pre-treatment outpatients for impulse control and work role scales, and were higher than pre-treatment outpatients for the alcohol abuse and mate role disturbance scales.

The second set of psychological health outcomes were evaluated on a monthly basis using the Zung Self-Rating Depression Scale (SDS) and the Zung Self-Rating Anxiety Scale (SAS). These self-report instruments assessed clinically significant depression and anxiety respectively and had been found reliable and valid for this purpose in several studies.

Over 90% of the men in the study completed the SDS and SAS for each month they were sent a form.

1. Methods of Measurement

Psychological health was monitored along with physical health at the examination visits to Boston University Medical Center, made at approximate 9-month intervals. At these visits the Psychiatric Status Schedule (PSS), a structured psychiatric interview procedure, was used by doctoral-level psychologists to evaluate the ATCs.

The PSS was selected because it allowed assessment of psychopathology in depth, because comprehensive data were available on normative psychiatric populations and because of the instrument's accepted reliability and validity (Spitzer, Endicott and Fleiss, 1967; Spitzer, Endicott, Fleiss and Cohen, 1970; Strupp, 1972; Luborsky, 1972; Weissman, 1975).

In addition to the pre-existing data on other populations and the recommendations of other researchers, the use of the PSS meant that we were assessing psychological health not only systematically, but also through the eyes and experience of a clinician.

The psychologist administering the Psychiatric Status Schedule used the standard press board booklet, 32 pages in length, with a precoded answer sheet for recording the presence or absence of 321 symptoms. The Schedule presented primary questions to elicit information about symptoms, and also provided follow up questions in the event that the first answers were insufficient to judge the presence or absence of a symptom.

Questions were grouped to cover symptoms in the areas of somatic complaints, moods, fears, interpersonal relations, ideation, cognitive processing, suicidal tendencies, drug use, alcohol use, delusions, hallucinations, behavior in the interview situation, and role functioning as a wage earner, housekeeper, student, mate, parent, and patient, as applicable for a given subject. A computerized program then scored the symptoms using item weights of 0, 1, 2 or 3, and converted the raw scores into T-scores (mean = 50, S.D. = 10) based on the results from the standardization study of 770 psychiatric patients.

A total of 50 T-scores were possible. Four scores, called macro-scales because they subsumed 16 symptom scale scores, assessed subjective distress, behavioral disturbances, impulse control disturbances, and reality testing disturbances. An alcohol abuse symptom scale was scored independently. Of seven role scales assessing impairment due to psychiatric disturbances in wage earner, housekeeper, student, mate, parent and patient (denial of illness) roles, and in addition, an overall average of the six scale scores, we used only the wage earner and mate role scales.

We focused primarily on the four macro-scales, the independent symptom scale of alcohol abuse and the three role scales named. The other symptom scales were not utilized because they were subsumed by the macro-scales. The housekeeper role, parent role, and the student role scales were not used because they were not applicable to all subjects. Our findings then led us to concentrate on only five outcomes: subjective distress, impulse control disturbances, work role disturbances, mate role disturbances and alcohol abuse.

Training and Reliability of Interviewers

Three doctoral-level psychologists and one master's-level psychologist were trained to use the PSS. All had previous clinical experience in the assessment of psychopathology in both normal and psychiatrically ill populations. Initial training consisted of familiarization with the instrument and the use of the training manual and tapes available from Spitzer, Endicott, and Cohen (Spitzer, Endicott, and Cohen, 1968; Spitzer, Endicott, and Cohen, 1968b). After reaching the criterion level suggested in the training manual (less than 2 false positives and less than 4 false negatives out of 321 possible answers for each of the standard tapes and four special tapes that we produced), a minimum of 20 subjects were interviewed by pairs of psychologists. In interviews, the interviewers alternated the active interview task, with the inactive member of the pair observing and independently filling out the PSS answer sheet. This procedure was undertaken to standardize the interviewers' interpretation of responses and scoring among themselves.

A computer program for scoring the entire PSS protocol was obtained from Dr. Endicott. Twenty-five test cases were provided to check the accuracy of the transplanted program. Scoring by the computer program was initiated after its absolute accuracy was established.

Pearson product moment correlations were calculated for the scale scores produced by the interviewer pairs for each subject. With the exception of three scales, the interviewers demonstrated a high degree of consistency with reliability coefficients ranging from .83 to 1.00. The behavior disturbance, speech disorganization, and social isolation scales had lower reliability.

The low interviewer reliability for those three particular scales primarily was due to variance and a difference in scoring in only one item in each case. Despite the high inter-rater correlations for the other scales, it was possible that there could have been significant mean differences between the pairs of interviewers. A series of t-tests were performed and the results indicated no significant differences between the interviewers in any pair. The scale means and standard deviations indicated that the high rater reliability primarily was due to a complete agreement on the lack of pathology for most subjects at any given examination. For scales with variability, inter-rater reliability was at least equivalent to that reported by Spitzer et al. (Spitzer, Endicott, Fleiss and Cohen, 1970).

We also examined whether or not the scale scores were affected by the role (observing or interviewing) of the interviewer. No significant difference was found between the scores generated in the observer role and in the interviewer role.

The above procedures were carried out each time a new psychologist was hired for the study. A new interviewer's scores were not accepted as data of record until the inter-rater reliability was above .8 with the current interviewer and until there was no mean difference in the scores obtained by the two individuals. Since we also required the interviewer to surpass the criterion level for training suggested in the manual for the PSS, we were confident that our interview results were highly reliable.

Internal Consistency of PSS Criterion Scales

Spitzer et al. reported the internal consistency of the PSS scales in their standardization study of 770 newly-admitted psychiatric inpatients (Spitzer, Endicott, Fleiss, and Cohen, 1970). Table 125 presents the internal consistency figures derived from the intake interviews with the air traffic controllers, and for comparison, those of the inpatient standardization sample. The table shows that high internal consistency reliability was obtained for less severe and more prevalent areas of psychiatric dysfunction (subjective distress, alcohol abuse, and mate role). The low internal consistency for behavioral disturbances, impulse control disturbances, reality testing disturbances, and several of the role scales seemed reasonable and appropriate since one would not expect such symptoms to be common or consistent in a highly selected non-patient group such as the air traffic controllers. In the standardization sample the internal consistency was much higher, as would be expected with a sample of patients whose symptoms were clustered more at the extremes.

Overall Psychiatric Status Profile at Intake

Table 126 presents the means and standard deviations of scores on the PSS for our sample of air traffic controllers, a large group of psychiatric inpatients, and a group of 130 urban community residents. The absolute zero point for each scale, indicative of no pathology, also is presented in the

TABLE 125

Internal Consistency Of PSS Scales For Air Traffic Controllers[1]
At Intake Compared To Standardization Group Of Patients

Macro-Scales	Number of Items	Air Traffic Controllers	Patient Standardization Sample[2]
Subjective distress	80	.82	.89
Behavioral disturbances	72	.47	.80
Impulse control disturbances	33	.52	.86
Reality testing disturbances	24	.38	.86
Role Scales			
Wage-earner	13	.28	.68
Mate	10	.72	.76
Parent	12	.19	.65

[1] N = 416

[2] Kuder-Richardson Formula 20 internal consistency, $r = \frac{k}{k-1}\left(1 - \frac{\Sigma pq}{Sy^2}\right)$

TABLE 126

Means and Standard Deviations Of Scores On Psychiatric Status Schedule For Air Traffic Controllers At Intake, A Community Sample, And A Psychiatric Inpatient Sample

Macro-scales	Air Traffic Controllers N=416 Mean (s.d.)	Community Sample N=130 Mean (s.d.)	Inpatient Sample[1] N=603 Mean (s.d.)	Score for no pathology (zero point)
Subjective distress	34.1 (3.8)	33.8(4.6)	49.7(10.3)	31
Behavioral disturbance	42.8 (2.9)	42.1(2.6)	50.7(10.4)	41
Impulse control	46.7 (3.0)	45.6(1.3)	50.7(11.0)	45
Reality testing disturbance	44.2 (.8)	44.5(1.4)	50.9(10.9)	44
Role Scales				
Wage earner (416, 93, 142)[2]	41.2 (2.9)	40.7(2.0)	51.4(11.5)	40
Mate (384,82,144)[2]	43.1 (6.3)	42.4(5.5)	50.8(10.0)	39
Parent (356,58,116)[2]	44.0 (2.8)	45.4(4.0)	51.4(12.1)	43
Symptom Scales				
Alcohol abuse	47.2 (3.1)	46.4(2.1)	50.7(10.5)	46

[1] Spitzer, Endicott, Fleiss and Cohen (1970).

[2] Number of subjects for all male, non-patient sample; community sample; and inpatient samples respectively.

last column.

The table indicates that the psychiatric profile of air traffic controllers and of the urban community sample were relatively similar to one another and quite different from the inpatient sample. However, the air traffic controllers scored statistically higher than the urban community sample on the impulse control disturbance ($t = 5.9$, p .001), and wage-earner role ($t = 3.4$, p .001). On the other hand, the controllers scored significantly lower than the urban community sample on the parent role scale ($t = 2.6$, p .01). The air traffic controllers and the community sample were significantly lower than the inpatients on all scales, as one might expect.

Some of the differences between the urban community sample and our sample were statistically significant, but of unknown practical or clinical significance, whereas the large absolute differences between these two groups and the inpatients were both statistically and clinically meaningful. Furthermore, some of the difference between air traffic controllers and the urban community sample may have been due to the difference in sex composition of the groups. Our sample was all male, while the community sample was 58% female. Each of the scales on which there was a significant difference between these two groups may have reflected personality and symptom reporting differences between males and females.

Classification of ificant Psychiatric Symptomatology

The Psychiatric Status Schedule provided scale scores for the various areas of psychiatric symptomatology. However, for the purpose of this study we needed to establish the levels at which we would consider an individual to have significant psychiatric symptomatology in each area. These cut-off scores were established to have statistical as well as clinical meaning.

To achieve bo statistical and clinical meaning, we compared the air traffic controllers with five other groups reported in the literature. Table 127 displays comparative statistics for inpatients, former inpatients, pre-treatment outpatients, an urban community sample, urban community leaders and the air traffic controllers at intake into the health change study.

Each set of three rows in Table 127 gives the mean, standard deviation, and number of subjects in each group who were scored when this number was different from the total number of subjects in a group. It is apparent that the three non-

TABLE 127

Comparative Statistics for PSS Samples[1]

	Number of Subjects	% Male	Means and standard deviations of PSS scales								
			Subjective Distress	Behavioral Disturbance	Impulse Control Disturbance	Reality Testing Disturbance	Alcohol Abuse	Wage-Earner Role	Mate Role	Parent Role	Student Role
Inpatients	603	27.2	49.67 10.29	50.67 10.37	50.70 10.98	50.91 10.85	50.73 10.54	51.43 11.47 (142)	50.78 10.04 (144)	51.37 12.07 (116)	49.78 9.76 (83)
Former Inpatients	97	26.8	41.91 7.22	54.26 10.19	46.05 3.17	47.02 4.96	47.05 3.51	45.00 4.47 (6)	50.20 8.10 (25)	48.53 .2.32 (17)	N/A
Pre-treatment Outpatients	91	62.6	48.82 10.36	46.00 7.99	50.23 8.00	46.53 6.05	47.42 6.00	46.77 9.14 (44)	50.14 9.15 (37)	51.41 10.76 (29)	49.09 -9.14 (45)
Urban community Sample	130	42.3	35.75 4.61	42.12 2.55	45.56 1.31	44.48 1.36	46.39 2.07	40.65 1.98 (93)	42.41 5.52 (82)	45.36 4.09 (58)	39.90 2.02 (10)
Urban community Leaders	26	84.6	33.69 3.65	42.38 4.51	45.85 1.43	44.15 .54	46.85 2.99	40.58 2.16 (26)	43.36 5.22 (22)	44.30 4.11 (10)	42.00 4.24 (2)
Air traffic controllers (ATCs)	416	100.0	34.11 3.77	42.84 2.88	46.69 2.98	44.22 .84	47.16 3.09	41.20 2.86 (416)	43.01 6.30 (384)	44.00 2.82 (356)	40.92 3.08 (113)
Raw Zero Score			31.00	41.00	45.00	44.00	46.00	40.00	39.00	43.00	39.00
Symptomatic Cut-Off			41.00	51.00	49.00	46.00	52.00	45.00	53.00	52.00	50.00
% ATCs above Cut-off			6.0	2.6	12.7	2.6	7.5	4.1	7.6	2.2	1.8
% Urban community Above Cut-off			9.9	1.5	3.1	6.1	2.3	2.1	6.1	8.5	0.0

1. Data for all samples except air traffic controllers were supplied by Jean Endicott, Ph.D. of the Evaluation Section, Biometrics, NY State Dept. of Mental Hygiene

NOTE: The figures in parentheses for the role scale results give the number of subjects for whom the role scale was appropriate and scored.