MEASURES OF EXERCISE TOLERANCE

by

Craig Taylor
Raymond Fransen

A report on researches conducted at Stanford University and at New York University, by means of a grant-in-aid from the Committee on Selection and Training of Aircraft Pilots of the National Research Council, from funds provided by the Civil Aeronautics Administration.

February 1946

CIVIL AERONAUTICS ADMINISTRATION
Division of Research
Report No. 57
Washington, D. C.

National Research Council

Committee on Selection and Training of Aircraft Pilots

Executive Subcommittee

M. S. Viteles, Chairman

E. C. Andrus

C. W. Bray

D. R. Brimhall

J. W. Dunlap

J. C. Flanagan

W. R. Miles

P. J. Rulon

2101 Constitution Avenue Washington, D. C.
Division of Anthropology and Psychology

Committee on Selection and Training of Aircraft Pilots

February 21, 1946

Dr. Dean R. Brimhall
Asst. to Administrator for Research
Civil Aeronautics Administration
Room 3895, Commerce Building
Washington 25, D. C.

Dear Dr. Brimhall:

Attached is a report entitled Measures of Exercise Tolerance, by Craig Taylor and Raymond Franzen. This report is submitted by the Committee on Selection and Training of Aircraft Pilots with the recommendation that it be included in the series of technical reports issued by the Division of Research, Civil Aeronautics Administration.

The report describes a series of investigations undertaken to provide acceptable measures of exercise tolerance. An outstanding feature of the investigation is the use of refined statistical techniques to determine the applicability of a battery of submaximal cardiovascular-respiratory measures in predicting maximal tolerances for specific types of physical activity. Because of the number of cases and of other factors the investigations are necessarily of an exploratory character. Nevertheless, they have produced outcomes in the way of a methodological approach to the construction of physiological tests which is of great interest in terms of current problems and of methods in the examination of physical fitness.

Cordially yours,

M. S. Viteles

Morris S. Viteles, Chairman
Committee on Selection and
Training of Aircraft Pilots
National Research Council

MSV:pd

FOREWORD

This report presents another of the early investigations undertaken under the auspices of the Committee on Selection and Training of Aircraft Pilots in the general area of problems concerned with the determination of the relationships between physiological response to exercise or stress situations and physical fitness.

Previous studies published in this series of reports[1] have been concerned not only with the development of a maximized battery of physiological measurements for the prediction of particular work or stress situations, but also with the analysis of the predictive significance of these variables in terms of ability to learn to fly. By and large these studies have shown (1) that the cardiovascular-respiratory responses to short exercise tests are not significantly related to flight proficiency, (2) that longer periods of measurement are necessary if a particular physiological function is to be reliably represented by the measurements recorded, (3) that within the range of "average physical efficiency" these measures are not discriminative, (4) that in order to get a valid measurement of physical response to exercise it would be necessary to standardize not only the physiological measures themselves, but also to develop a criterion situation which would permit discrimination among individuals with respect to exercise tolerance.

The studies reported in this report deal with (1) the development of a criterion test for the evaluation of exercise tolerance, (2) an analysis of existing standard fitness tests in terms of this criterion, (3) an analysis of the relationship between the cardiovascular-respiratory measures and this criterion, and (4) determination of the interrelationships among the physiological measures themselves.

---

[1] Larson, L. A. A factor analysis of some cardiovascular-respiratory variables with particular reference to the Schneider and the McCurdy-Larson tests. Washington, D. C.: CAA Division of Research, Report No. 16, June 1943.

Franzen, R., and Blaine, L. Evaluation of respiratory measures for use in pilot selection. Washington, D. C.: CAA Division of Research, Report No. 25, January 1944.

Foley, J. P., Jr., Hunt, T., Kelly, E. L., Johnson, A. P., and Lepley, W. M. Studies of predictors of achievement in learning to fly. Washington, D. C.: CAA Division of Research, Report No. 27, March 1944. (See Part III.)

McFarland, R. A., and Franzen, R. The Pensacola study of naval aviators. Final summary report. Washington, D. C.: CAA Division of Research, Report No. 38, November 1944.

McFarland, R. A., and Franzen, R. Detailed statistical analysis of data obtained in the Pensacola study of naval aviators. Washington, D. C.: CAA Division of Research, Report No. 41, January 1945.

The experimental work on these problems was begun in March of 1941 at Stanford University with funds made available from a grant-in-aid from the NRC Committee on Selection and Training of Aircraft Pilots, originally made to Dr. J. K. Lewis. The first series of studies was completed in November of 1941 and a report, Studies in Physical Fitness I, was submitted to the Committee. By the end of June, 1942, three additional series of investigations were completed and the results analyzed statistically in cooperation with Dr. Raymond Franzen. Two additional reports were submitted presenting these materials.

On April 1, 1942, financial support for continuation of these studies was transferred to the Committee on Aviation Medicine of the OSRD under contract OEM cmr 122. Statistical analyses of the results of these studies continued to be carried out by Dr. Franzen for the Committee on Selection and Training of Aircraft Pilots.

In addition to the reports of research in this area submitted to the NRC Committee on Selection and Training of Aircraft Pilots, other studies[2] were submitted to the Committee on Medical Research, Office of Scientific Research and Development. The last two reports cited below were prepared jointly by Dr. Taylor and Dr. Franzen and were also submitted in summary to the Committee on Selection and Training of Aircraft Pilots.

In reading the following report it will be noted that these studies can be considered as exploratory only and that much experimental work remains to be completed before definitive conclusions can be reached. Due to the nature of the sample used (a homogeneous group of physically "normal"

---

[2]Taylor, C., Allen, S. C., and Hall, V. E. Orthostatic insufficiency by the tiltboard method. Committee on Aviation Medicine, OSRD. June 1943.

Taylor, C., Brown, G. E., and Morris, R. Muscle endurance tests of exercise tolerance. CMR, OSRD. June 1943.

Taylor, C. A maximal pack test of exercise tolerance. CMR, OSRD. June 1943.

Taylor, C., and Brown, G. E. Some observations on 325 Schneider tests on healthy young men. CMR, OSRD. June 1943.

Crescitelli, F., and Taylor, C. The lactate response to exercise and its relationship to physical fitness. CMR, OSRD. June 1943.

Taylor, C., and Morris, R. Studies of the effect of training on scores in a submaximal test of exercise tolerance. CMR, OSRD. June 1943.

Taylor, C., and Franzen, R. Cardiac patterns related to exercise tolerance. CMR, OSRD. July 1943.

Taylor, C., and Franzen, R. Submaximal tests of exercise tolerance. CMR, OSRD. July 1943.

subjects within a very restricted age range) and the relatively small number of subjects involved as compared with the large number of variables under investigation, the results cannot be generalized and must serve only to indicate the directions which further investigations should take. Such investigations will, of course, involve a larger and more heterogeneous population of subjects and will consider the validity of the criterion test in terms of flying and possibly of other work situations.

The studies described in this report have served to indicate inadequacies of existing tests of physical fitness and the unreliability attendant in the physiological measures being customarily recorded as predictors of physical fitness. They have also suggested that a simple, easily administered, "submaximal" test which is under control of the examiner represents a practical means of assessing physical fitness.

CONTENTS

MEASURES OF EXERCISE TOLERANCE

PREFACE

This report embodies the results of a series of preliminary investigations undertaken with a view to the development of a standard test of exercise tolerance and the construction of a battery of cardiovascular-respiratory measures which might serve as an index of the physical fitness of the individual being measured.

One of the first problems encountered in these investigations was to secure a valid criterion measure of the physical fitness or exercise tolerance of the individual, i.e., some method which would measure the actual amount of work the individual was capable of performing. Since there is little agreement as to the meaning of the term "fitness" and since any meaning which might be attached to the term cannot be separated from the particular work or exercise situation in which the individual finds himself, it was decided to construct a standard measure of exercise tolerance which would reliably assess the subject's <u>maximal capacity for a particular type of work</u>.

Accordingly, the maximal treadmill test was developed for purposes of these initial studies. Since such a test demands that the subject be exercised to the point of physical collapse or exhaustion, it is not practical for use in the field. The ultimate goal is therefore to construct a shorter test from which the exhaustion endpoint of the individual can be reliably predicted.

A second major problem encountered in these studies was the reliable measurement of the physiological responses to exercise. Accordingly, one of the major objectives of these investigations has been the development of reliable measurements of the cardiovascular and respiratory responses to the tests of exercise tolerance.

In such studies as these there are, of course, many sources of uncontrolled variance such as motivation of the subject, diet, previous practice, anthropometric differences, and so forth. In so far as possible, these major sources of varience have been taken into account. It will be necessary, however, in final studies to be conducted in this area to control rigidly every major source of variance if a valid measure of exercise tolerance is to be developed.

The studies discussed in the following sections of the report are admittedly exploratory only. They are confined to the study of a relatively homogeneous population of subjects with respect to such factors as age, general state of health, sex, and so forth. Such studies as these do, however, point the way for further research in this area and go a long way toward the development of a test of exercise tolerance and of measures of the physiological reactions to exercise which might prove of practical predictive significance in the evaluation of an individual's suitability (in terms of physical fitness) for occupations which demand a high degree of physical hardiness or the ability to withstand extreme fatigue.

SECTION I

MAXIMAL TESTS OF EXERCISE TOLERANCE

INTRODUCTION

In order to construct a valid battery of physiological measures which will predict the ultimate "exercise tolerance" of the individual, it is necessary that the measures be recorded for each individual throughout an exercise or work period which culminates in complete physical exhaustion.

In this section are presented the results of a series of exploratory experiments which involve the analysis of cardiovascular and respiratory measures of subjects who were subjected to a maximal test of exercise tolerance resulting in final and complete exhaustion of the subject.[1]

As was indicated earlier in this report, such a severe test of exercise tolerance (continuing until exhaustion) is impractical for field use. Initial studies such as these are necessary, not only for purposes of maximizing a battery of predictive physiological measures, but also for the construction of a criterion measure of exercise tolerance against which shorter more practical "fitness" tests may be validated in subsequent studies.

THE (MAXIMAL) TREADMILL TEST

Apparatus. Briefly, the treadmill test consists of an endless belt running over two steel pulleys -- one motor-driven, the other an idler. Between the pulleys the running surface of the belt is supported by a series of ball-bearing rollers which support the weight of the subject and furnish a satisfactory running surface. The front end of the treadmill can be raised or lowered by a vertically placed screw. This permits different inclinations (pitches) of the running surface to be used in order to increase the physical work involved in the exercise. Differences in pitches are expressed as per cent grade according to the conventional record. Differences in running rate may be brought about by a speed-adjusting screw which controls the setting of a Toledo Timer[2] interposed in the V-belt drive system. Thus, the treadmill test offers all of the variations in exercise stress which can be brought about through changing speed or pitch of the running surface.

Procedure (Maximal Tests). The complete test procedure is divided into three parts, the warmup period, the rest period, and the test itself,

---

[1]The predictive value of individual cardiovascular and respiratory variables was investigated in preliminary studies involving 100 subjects. The major results of these studies are presented in Appendix A.

[2]Toledo Timer Co., Toledo, Ohio.

as follows:

1. The Warmup Period. The belt speed of the apparatus was set at 108 meters per minute with an inclination (pitch) of 5 per cent. Each subject was required to walk at this rate for four minutes. The primary purpose of this portion of the procedure is to accustom the subject to the motion of the treadmill test and to provide him with a "limbering up" period.

2. The Rest Period. Immediately following the warmup period, the subject was required to sit in a relaxed position for three minutes. This rest period favors the onset of perspiration, an important feature of heat loss and one of the difficult adjustments of "second wind." By inducing some of these physiological adjustments before the test, part of the stress of adaptation is removed.

3. The Treadmill Test. The initial run was begun with the speed of the belt adjusted to 162.16 meters per minute and the pitch set at 5 per cent grade. At this speed, the subject runs at what might be called a moderately fast jog.

This low rate not only assured sufficient duration of exercise to permit full physiological adjustment, but eliminated the difficulties of maintaining a safe pace on the treadmill and the fear of falling off which occurs in some subjects when faster speeds are used. The uphill nature of the run gave somewhat more stress at lower speeds, and by adding an unusual feature to the performance, minimized the factor of skill in running as a source of variance in the measures. Since the pace of the subject was controlled by the motor-driven belt, the strategy of planning the run, a prominent feature in competitive racing which varies with the experience of the subject, was completely eliminated.

After one minute of running on the 5 per cent grade the pitch of belt was increased 1 per cent. Thereafter, the pitch was increased 1 per cent after each minute's running, forcing the subject to run up a hill which was progressively steeper.[3]

Each subject was instructed to run to the point of complete exhaustion. Several methods were used to urge him to this endpoint. Previously obtained records were pointed out to the subject and he was impressed with the fact that only an experiment terminating in complete exhaustion could be considered valid. For the younger subjects, a challenge of this sort was usually

---

[3]The use of pitch increase instead of an increase in speed as a means of stepping up the intensity of the work was selected after each method had been given a preliminary tryout during the exploratory studies with the apparatus. Because of difficulties with higher speeds which often arose in subjects not used to running on the treadmill, grade increase was selected. From a physiological standpoint, there seems to be little to choose between these two types of work increase.

all that was necessary to get the most out of them. Two operators were in charge of the tests: one to take the heart rates and other measurements; the other to guide the subject.

By a combination of "cheer-leader" tactics in urging the subject on, and by closely watching the physiological state of the subject, the examiner was able to bring the test consistently to a fairly well defined exhaustion endpoint. The onset of excessive respiration, gasping for breath, staggering gait, and facial expression of distress are all signs that most subjects display when they are approaching exhaustion. When these appeared, the experiment could seldom be continued longer than a minute and was most often terminated within a half-minute. A last reminder was given to the subject to run as long as possible.[4] When completely through running, the subject gave the signal to stop the treadmill or simply stumbled off the apparatus.

Scoring the Treadmill Test. The subject's score on the treadmill test was expressed as the total work output in kilogram meters. Since the change in pitch is the basis of change in rate of work, the total work output is proportional to the change in vertical distance climbed per unit time. Thus, the rate of ascent may be calculated by the formula:

Belt speed (meters/minute) x grade (ratio of vertical component to length of inclined running surface) = Rate of Ascent.

During the first minute of running at a pitch of 5 per cent and a speed of 162.16 meters, the vertical ascent would be (162.16 x .05) 8.11 meters. Each additional pitch increase of 1 per cent adds 1.6 meters to the rate of vertical ascent. Now, since the rate of ascent reached at the exhaustion point, the total ascent, and the time of running are all proportional, the total work may be calculated from the formula:

Body Wt. (kilos) x Total Ascent (vertical meters) = Total Work.

The product of this calculation is expressed in vertical kilogram meters (VKgM).

Since body weight is a relatively fixed quantity for each subject, the lighter subject of equal fitness with a heavy man would have to run longer and climb a greater vertical distance in order to accomplish the same amount of work. A priori, this appears to penalize the lighter subject unduly and thus to detract from the accuracy of the measures of work.

Accordingly, a small correction for this effect was made. The regression of work done upon speed reached and body weight was calculated.

---

[4]Editor's note. It should be noted that the motivation of the subject constitutes one of the major sources for uncontrolled variance in the study.

in multiple regression form yielding the following formula:[5]

Total Work (adjusted) = 1051 Speed (vertical meters/minute)
+ 142.7 Wt. (kilos) - 22,709.

This value, divided by 1000 for convenience, served as the subject's score on the treadmill test. For example:

Subject J.D. weighs 80.0 kilograms. His treadmill record shows the following measurements:

1. Time of run -- five minutes.
2. Rate of ascent attained -- 14.59 meters per minute.
3. Total Ascent.

$$5 \times \frac{\text{Rate 1st min. (8.11) plus Rate 5th min. (14.59)}}{2} = 56.75 \text{ m.}$$

4. Total Work:

Total Ascent (56.75 m.) x Body Wt. (80.0 kg) = 4540 VKgM

5. Treadmill Score for subject J.D. equals:

$$\frac{(1051)\ (14.59) \text{ plus } (142.7)\ (80.0) \text{ minus } 22,709}{1000} = 4.04$$

PHYSIOLOGICAL TESTS

Blood Pressure Readings. Systolic and diastolic blood pressure readings were taken by standard auscultatory techniques by means of Tycos anaeroid sphygmomanometers. The appearance of sound was taken as the systolic criterion and the point of muffling of sound as the diastolic criterion. Standard bell-type stethoscopes were used to detect the Korotkov sounds.

Heart Rate Determinations. From continuous electro-cardiotachometric records taken during the warmup period, the rest period, and the maximal run, the following heart rate measures were recorded:

1. Standard Exercise Heart Rate. The standard exercise heart was recorded as the half-minute rate during the last minute of the preliminary walking (warmup) period.

---

[5] In a preliminary study with 100 cases it was found that the correlation between rate of ascent reached and total work was .928 ± .014 and between body wt. and total work .170 ± .099. These figures would indicate that body wt. causes little variation in the amount of work done by the subjects. However, it was arbitrarily decided to make the correction indicated in the text in order to take this consideration into account in a minor degree. See: Taylor, C. Studies in physical fitness I, 1941. (This report is on file with the NRC Committee on Selection and Training of Aircraft Pilots.)

2. Maximal Run Heart Rates. Heart rate counts were made for each half-minute of the test run beginning at 15 seconds after the start of the run and continuing until the run was completed.[6]

Flarimeter Blow (FB). The Flarimeter[7] is actually an adaptation of the U-Tube mercury manometer. At stated periods during exercise or during a period of recovery from exercise, the subject is asked to inhale and then to "blow" into a mouthpiece which communicates with a fluid column of mercury. A very slow leak in the side of the mouthpiece permits the air to escape at the approximate rate of 50-100 c.c.s. per minute. (Actually, the lung air is far from exhausted during the blow, especially when conducted after exercise, and the test is essentially one of "breathholding.") This small orifice makes easy the timing of duration of of the "blow" and the procedure is simple for the subject to master if given a short practice period before the study is begun. Duration of the flarimeter blow was recorded in seconds and tenths of seconds by means of a stopwatch.

In the studies presented in this section the flarimeter technique was varied somewhat for the different groups as follows:[8]

Group	No. Blows	Press. mm Hg.	When Taken
I, II, III	1	20	1 minute after standard exercise.
IV-A, IV-B	3	20	45 seconds after standard exercise. 2nd and 3rd blows follow 10 seconds after the preceding. Score is the summation of the three blows.
V	2	20	45 seconds after standard exercise. 2nd blow follows 10 seconds after lst. Score is the summation of 2 blows.

The subjects were urged to maintain the blow as long as they possibly could. Slight fluctuations about the mark are of little consequence, but as the "breaking point" approaches, erratic excursions of the fluid column indicate the struggle to maintain the blow. Unless these signs occur, it

---

[6]These heart rates have been fitted by a cubic equation, as will be shown later in the paper, and with the exception of the "terminal" heart rate and the heart rate at the midpoint of the exercise period, etc., they do not appear individually in the analysis of the data. Rather, various functions of the "fitted" curves are used.

[7]McKenzie, L. F., Wells, P. V., Lewis, E. G., and Ylvisaker, L. S. Flarimeter tests of circulatory fitness. Amer. J. Med. Science. 1930, 180, 372-386.

[8]For purposes of the present study these variations in flarimeter technique are not considered important, although in subsequent studies the procedure will be investigated in some detail.

is likely that the subject has not exerted himself sufficiently. Abnormally short blows (e.g., less than 10 seconds durations) if not accompanied by signs of respiratory struggle, may indicate that the physiological end-point has not been reached.

THE SUBJECTS

All of the subjects involved in the present studies were between the ages of 18 and 28 years and all were "normal" in the sense that they possessed no known defects or illness which might have influenced their tolerance for exercise. All were either college students, laboratory workers, or members of the Armed Forces.

Subjects were divided automatically into the following groups according to the date on which the tests were given:

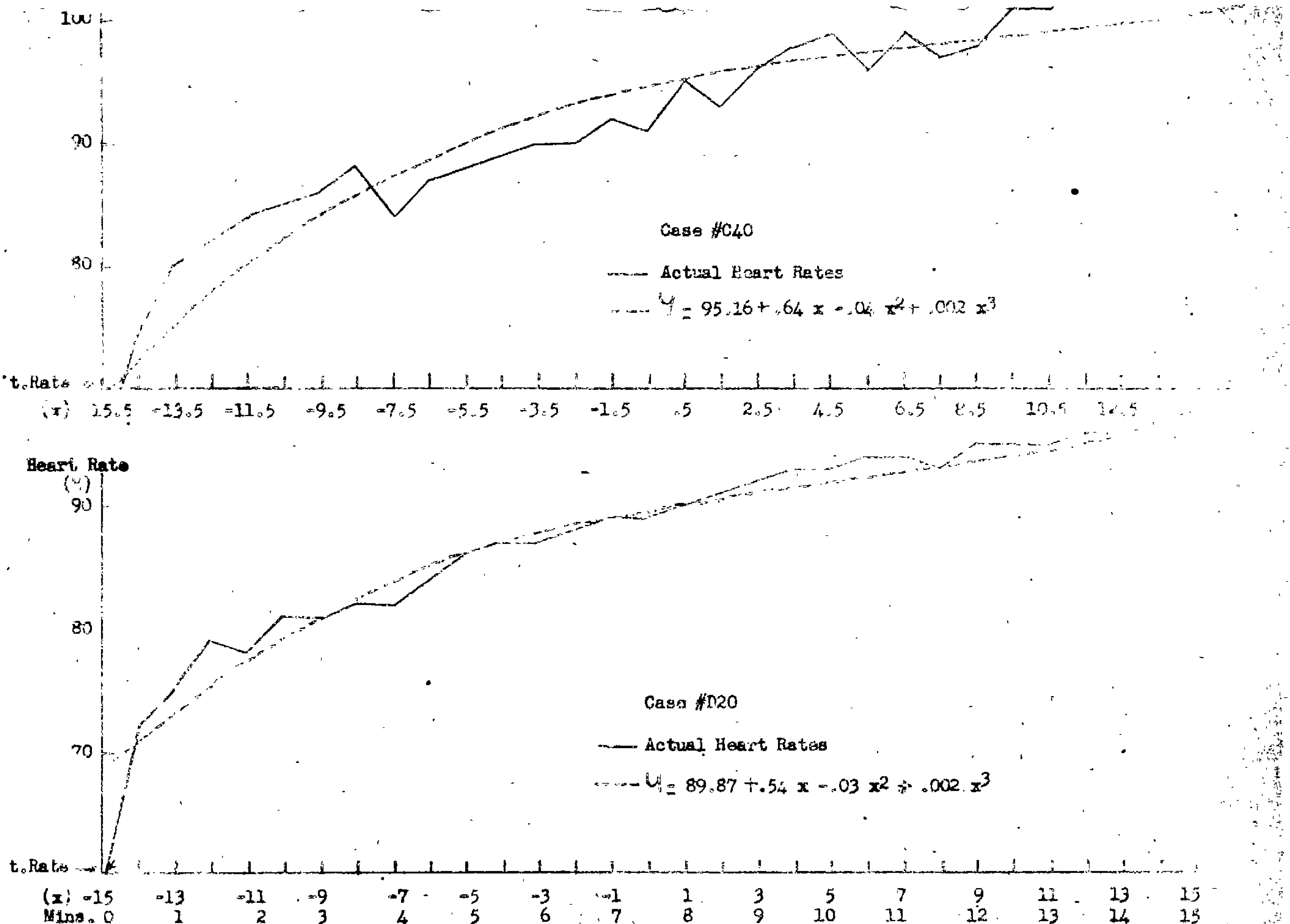
Group	N	Subjects
I	98	Stanford University students and soldiers in basic infantry training at Moffett Field, California
II	97	
III	45	
IV-A	29	Stanford University students
IV-B	(retest)	(Retested as group IV-B)
V	83	Stanford University students and soldiers (ASTP unit at Stanford)

All tests were conducted in a laboratory where temperature was maintained within the limits of 66-76 degrees Fahrenheit. No subject was tested within 1½ hours of meal time and all were asked to eat lightly in the preceding meal, although such control of diet could not be guaranteed. Subjects were garbed in tennis shoes and trunks and were rested 10-30 minutes before beginning the tests.

RESULTS

Mathematical Treatment of Maximal Run Heart Rates. Individual graphs were plotted for the 240 subjects in Groups I, II, and III, showing half-minute heart rates plotted against time of running. In Figures 1 and 2 are shown representative curves for two physically fit and two physically unfit subjects. It will be noted that, in general, the curve rises steeply but with decreasing slope and then gradually begins to level off though rising slightly. Finally, an upward curvature appears as exhaustion ensues.

The curve for the physically fit subjects can be seen to show a more marked initial rise; a longer, lower, and flatter intermediate period; and a less pronounced or non-existent final upward curvature when compared to the curves for unfit subjects.



Case #C40
Actual Heart Rates
$y = 95.16 + .64x - .04x^2 + .002x^3$
100
90
80
t.Rate
(x)
Case #D20
Actual Heart Rates
$y = 89.87 + .54x - .03x^2 + .002x^3$
Heart Rate (y)
90
80
70
t.Rate
(x) -15 -13 -11 -9 -7 -5 -3 -1 1 3 5 7 9 11 13 15
Mins. 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

FIGURE 1

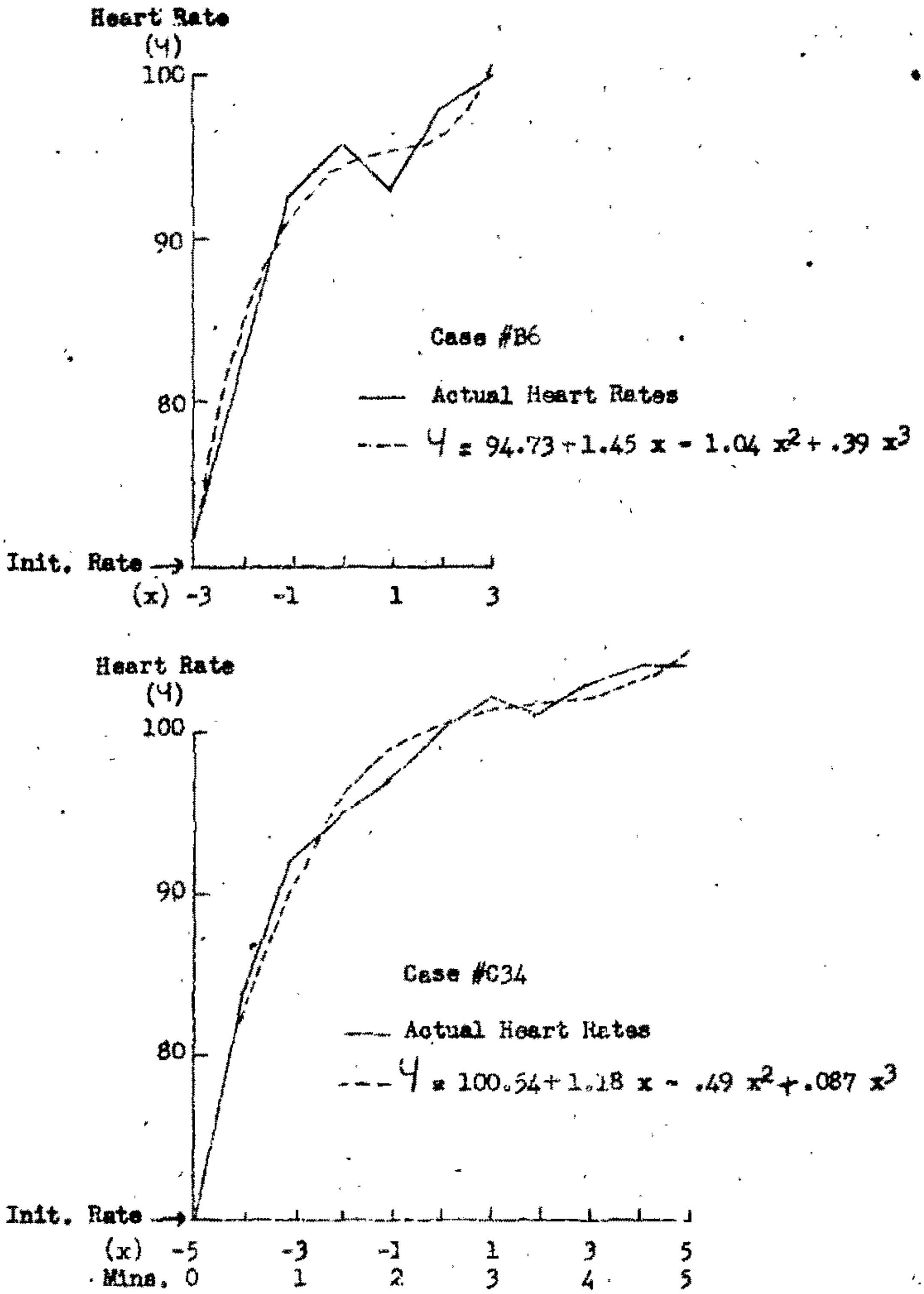


FIGURE 2

HALF-MINUTE HEART RATES OF PHYSICALLY UNFIT MEN
WHILE PER[illegible]

Physiologically, the physically fit man reaches his initial adaptation to strenuous exercise more rapidly, then approaches the condition of a steady state. Since the physical work involved is steadily increasing not even the most fit subjects can attain a completely steady state. The unfit man, however, tends rapidly toward a maximal level of heart rate and early exhaustion.

Attempts were made to determine mathematically the line of best fit for each individual curve. It was apparent from inspection of the curves that a cubic fit (3 degree parabolas) was probably best suited to the data. However, both quadratic and cubic fits were tried out. While there was a tendency for the curves of the physically fit subjects to approach a quadratic curvature, and even linearity, the cubic fit has the widest applicability. Equations of the form $Y = a+bx+cx^2+dx^3$ were calculated for 240 of the cases and fitted by the method of least squares. Examples of these equations are given in Figures 1 and 2.[9,10]

<u>Statistical Analysis of the Data Relative to Validity and Reliability</u>. The next step in the analysis of these data was to obtain single and multiple correlations between measures and the treadmill criterion. It seems obvious that nothing would be gained by the applications of such treatments if the complete maximal treadmill test were administered since then the treadmill score itself would be available.

In the analysis of the data information on the following points was sought:

1. The extent to which the maximal treadmill performance can be defined in physiological terms such as heart rates and breathholding.

2. The patterns of physiological response which are most valid as items of exercise tolerance as measured by the treadmill test and the relative importance of each component in the battery of measures.

3. The measures of treadmill exercise tolerance pertinent to a shorter, simpler, and submaximal type of fitness test which will predict performance on the complete or maximal test.

---

[9] These data as well as the curves for the individual subjects are on file with the NRC Committee on Selection and Training of Aircraft Pilots.

[10] <u>Editor's note</u>. Although cubics seem to be the most correct mathematical expression, it should be noted that the "upward swing" on the end of the curves is of approximately the same magnitude as (or at least no greater than) the random fluctuations throughout the curve and therefore may not represent a true inflection point. Nevertheless, the cubics did provide the best mathematical expression of acceleration for these data. Whenever there was any doubt regarding the adequacy of the cubic fit the curvilinear correlation was computed. Rational equations were investigated also but proved inadequate.

Since the terminal abscissa of the curves (Figures 1 and 2) is a measure of exercise tolerance, correlations of parameters of the cubics with their deviation will determine which cardiac factors such tolerance depends upon.[11] It is apparent that the curves may differ in (1) initial pulse, (2) pulse at the midpoint, (3) terminal pulse, (4) slope, (5) acceleration, and (6) the zero acceleration point. All of these measures have been obtained by differentiating the equations which have been determined as the best fitting curve for the observations on each individual.

Tables 1, 2, and 3 present the linear and curvilinear intercorrelations of the following variables for each of Groups I, II, and III:

(1) Score on the maximal treadmill test.

(2) Acceleration at the 4th half-minute, i.e., the 2nd derivative of cubics set at the 4th half-minute (change in rate of increase in heart rate).

(3) Acceleration at the 6th half-minute, i.e., the 2nd derivative of cubics set at the 6th half-minute.

(4) Standard exercise heart rate, i.e., heart rate during the last minute of the preliminary walk.

(5) Equation constant, i.e., heart rate at the midpoint of the run.

(6) Zero acceleration point, i.e., the time at which the final upward curvature began.

---

[11] Editor's note. The objection might be raised that the correlations between the criterion and the heart rate scores are rendered unduly high because of the fact that they are not independent, in as much as the length of the abscissa of the heart rate curve essentially corresponds to the criterion. Thus, when the criterion is one of the parameters of the curve it would be expected that this parameter could be predicted with a high degree of accuracy from other selected parameters of the same curve. Although this was admittedly true, it should be noted that one purpose of this investigation was to determine which of the parameters best predicts the criterion in order that a "submaximal test" might be developed which would approach the validity of materials employing exercise to exhaustion. The principal exploratory hypothesis was concerned with defining a function of exercise tolerance in terms of early performance by analyzing the relationships of other parameters of the heart rate curve to the abscissa. According to the authors "a vital conclusion which developed from this method of analysis was that you could predict this degree of exercise tolerance from the curvature or lack of it at the beginning of the heart rate curve." Final validation of such a submaximal test necessarily should be carried on with data only from the early part of the curve.

(7) Slope of curve at 4th half-minute, i.e., the 1st derivative of cubics set at the 4th half-minute (rate of increase of heart rate).

(8) Terminal heart rate, i.e., heart rate during last minute of the maximal run.

Examination of Tables 1, 2, and 3 reveals that variable (2) acceleration at the 4th half-minute is correlated about .6 with the criterion, the maximal treadmill score in all three sets of data. Acceleration at the 6th half-minute is less valid in terms of the treadmill score and less consistent (r's = .49, .54, and .01 for Groups I, II, and III, respectively).[12]

Standard exercise heart rate correlates about -.45 with the treadmill score in two sets of data but only -.06 in the third. There is some evidence, however, that this finding is distinctly atypical. To check this point the data were redivided into two sets of 120 cases each, and the correlations recalculated. It was found that in both of the new sets of data the r was about -.30.

Further proof of the abnormality of this low correlation between standard exercise heart rate and treadmill score is offered in a later section of this paper. It may be stated here, however, that efforts to replace standard exercise heart rate with a heart rate during the run were disappointing because a later heart rate, such as at the 4th half-minute, though more highly correlated with the criterion, is at the same time more highly associated with the other cardiac measures during the run and so does not contribute as well to the multiple.

The validity of the other measures in terms of the treadmill score can most easily be judged by their effect on the multiple correlations (Table 4). In Section A of this table are given the best multiples of 1st to 6th order utilizing cardiac measures alone. It is apparent that inclusion of measures (3), (6), and (8) does not substantially add to the multiple and (7) adds very little. $R_{1.2457}$ is .86 and .85 for the two sets in which standard exercise heart rate is correlated with the criterion, but only .66 where it is not.

Since addition of (3), (6), and (8) does not materially raise this last multiple, flarimeter blow was added to the battery and the best multiples of several orders were computed (see Section B of Table 4). The most noteworthy finding here is that flarimeter blow, while not affecting multiples for Groups I and III, does raise R with (2), (4), (5), and (7) for Group II from .66 to .74. This suggests that the flarimeter blow contributes substitute variance.

---

[12]The third derivative (a constant in cubics) is also correlated with the criterion.

TABLE 1

CORRELATIONS WITH TREADMILL SCORE AND INTERCORRELATIONS OF FUNCTIONS OF CUBIC HEART RATE TRENDS DURING EXERCISE

Group I (N = 98)

			(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accel. at 4th Half-minute	(2)	r	.62						
		eta	.80						
Accel. at 6th Half-minute	(3)	r	.49	.94					
		eta	.61	.96					
Standard Ex. Heart Rate	(4)	r	-.46	.08	.22				
		eta	.57	.42	.43				
Equation Constant (Heart Rate at the midpoint of the run)	(5)	r	-.30	-.40	-.41	+.43			
		eta	.53	.50	.53	.57			
Zero Accel. Point	(6)	r	.27	-.02	.08	-.18	.04		
		eta	.36	.50	.49	.45	.43		
Slope at 4th Half-minute	(7)	r	-.17	-.68	-.83	-.55	.34	.06	
		eta	.33	.75	.86	.68	.46	.48	
Terminal Heart Rate	(8)	r	-.01	-.12	-.16	.17	.67	-.01	.31
		eta	.30	.31	.37	.37	.71	.45	.44

TABLE 2

CORRELATIONS WITH TREADMILL SCORE AND INTERCORRELATIONS OF FUNCTIONS OF CUBIC HEART RATE TRENDS DURING EXERCISE

Group II (N = 97)

			(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accel. at 4th Half-minute	(2)	r	.62						
		eta	.72						
Accel. at 6th Half-minute	(3)	r	.54	.82					
		eta	.60	.86					
Standard Ex. Heart Rate	(4)	r	.06	.13	.21				
		eta	.30	.40	.44				
Equation Constant (Heart Rate at the midpoint of the run)	(5)	r	-.16	-.30	-.42	.37			
		eta	.47	.49	.56	.54			
Zero Accel. Point	(6)	r	.30	.33	.26	.06	.24		
		eta	.44	.54	.56	.52	.49		
Slope at 4th Half-minute	(7)	r	-.42	-.68	-.84	-.46	.41	-.09	
		eta	.58	.85	.88	.54	.53	.62	
Terminal Heart Rate	(8)	r	-.03	-.19	-.09	.19	.68	.08	.19
		eta	.37	.44	.36	.44	.75	.33	.44

TABLE 3

CORRELATIONS WITH TREADMILL SCORE AND INTERCORRELATIONS
OF FUNCTIONS OF CUBIC HEART RATE TRENDS DURING EXERCISE

Group III (N = 45)

			(1)	(2)	(3)	(4)	(5)	(6)	(7)
Accel. at 4th Half-minute	(2)	r	.51						
		eta	.66						
Accel. at 6th Half-minute	(3)	r	.01	.33					
		eta	.61	.96					
Standard Ex. Heart Rate	(4)	r	-.45	.21	.45				
		eta	.59	.57	.65				
Equation Constant (Heart Rate at the midpoint of the run)	(5)	r	-.18	-.27	-.09	.32			
		eta	.55	.54	.63	.60			
Zero Accel. Point	(6)	r	.39	.41	.04	-.09	.20		
		eta	.70	.69	.71	.52	.69		
Slope at 4th Half-minute	(7)	r	.03	-.51	-.78	-.70	.24	.01	
		eta	.64	.66	.91	.81	.53	.74	
Terminal Heart Rate	(8)	r	.10	-.12	-.03	.10	.75	.21	.20
		eta	.46	.62	.56	.57	.84	.53	.57

TABLE 4

MULTIPLE CORRELATIONS

Subscripts

Treadmill Score	1
Acceleration at 4th Half-minute	2
Acceleration at 6th Half-minute	3
Standard Exercise Heart Rate	4
Equation Constant (Heart Rate at the midpoint of the run)	5
Zero Acceleration Point	6
Slope at 4th Half-minute	7
Terminal Heart Rate	8
Flarimeter Blow	9

Multiple	Group I N = 98	Group II N = 97	Group III N = 45
A. Heart Rate Measures			
$R_{1.24}$	.80	.63	.77
$R_{1.245}$	.83	.64	.80
$R_{1.2457}$	.86	.66	.85
$R_{1.234567}$	.89	.67	.87
$R_{1.2345678}$	.91	.67	.88
B. Heart Rate Measures with Flarimeter Blow			
$R_{1.249}$	.83	.68	.77
$R_{1.2459}$	.85	.69	.81
$R_{1.2479}$	.83	.70	.77
$R_{1.24579}$	.87	.74	.88

The etas shown in Tables 1, 2, and 3, were convincingly different from r's mainly for variables which have not been selected for further treatment. Since the N here was small, it was not deemed worthwhile to develop multiple curvilinear correlations. The assumption of linearity throughout the data seems adequately supported by this and subsequent evidence.

Reliability Studies (Groups IV and V). The preceding analyses demonstrate that a combination of heart rate measures and flarimeter blow will yield high multiple correlation with the maximal treadmill score. Two further studies were carried out to verify this validity and to determine reliability. Such reliability is of two sorts:

1. Test-retest reliability of the individual functions.

2. Multiple regression reliability, determined by cross-application of regression coefficients both on test-retest data and on different samples of single test data. This is a particularly crucial proof of significant relationship since it assesses both the universality of the regressions and the degree to which they characterize the response of the individual.

The subjects of Groups IV-A and IV-B were given treadmill tests twice, two to four days intervening. They were not told their score on the first test, and had no simple way of judging the passage of time during either run. The reliability coefficient for treadmill score (r_{11}) was .96, despite the relatively low N of 29. As with the previous groups, electro-cardiotachometer records were obtained, and the functions of cubic fits were determined. Figure 3 gives two representative examples of these test-retest curves.

The statistical data for heart rate functions are presented in Table 5. The test-retest reliabilities for cardiac functions, though not as high as that of the criterion, are acceptable with the exception of slope at 4th half-minute. This measure was dropped at this point because of its low reliability and also its low correlation with the criterion.

Thus, three measures, (2), (4), and (5) were isolated which possessed sufficient reliability for use in a battery, in addition to flarimeter blow, which from other studies has been found to have a coefficient of about .80

The interaction of such physiological factors, however, may often be such that the reliability of the composite is greater than the individual measures. Accordingly, predicted scores were computed for each subject's

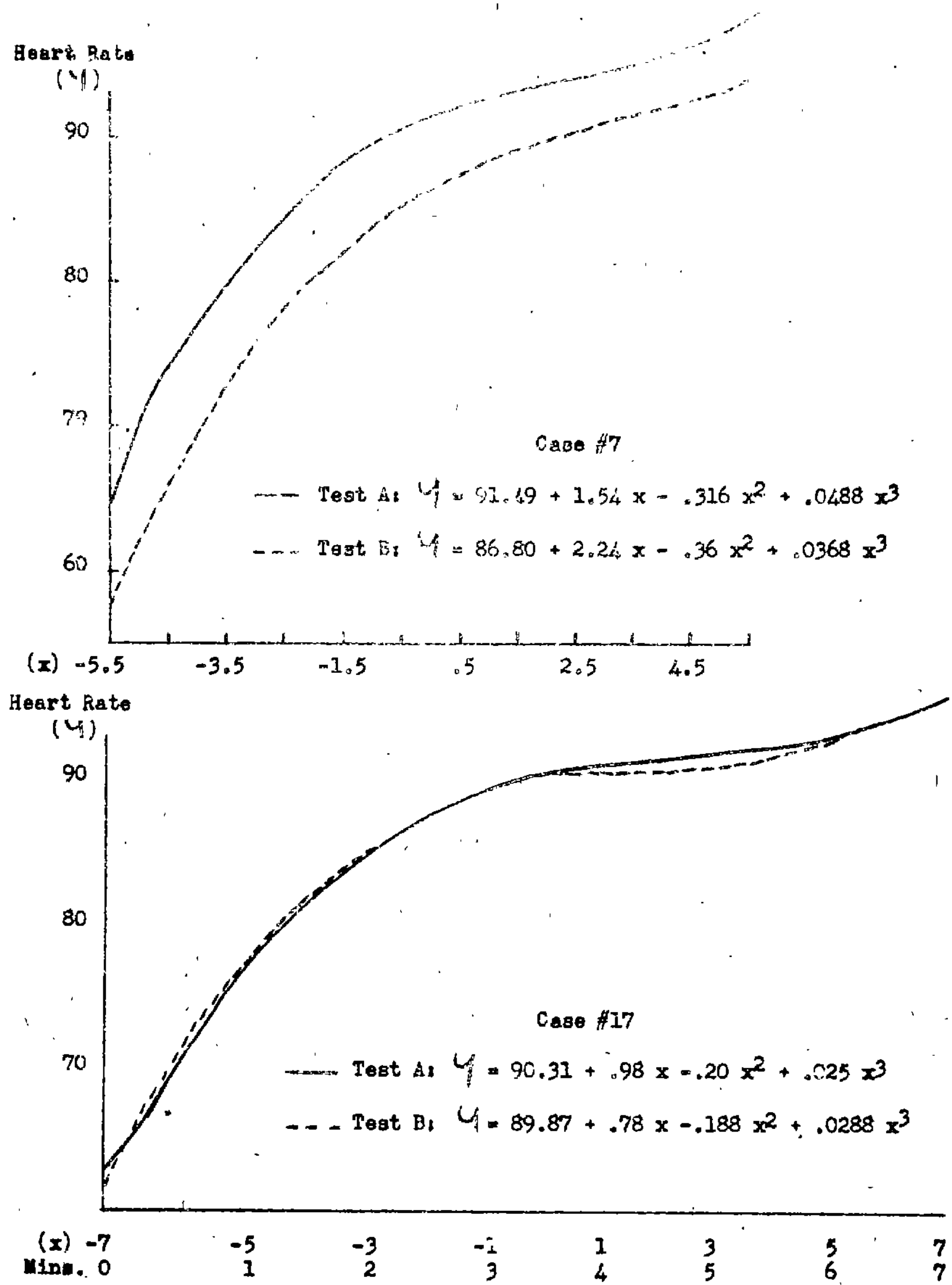


FIGURE 3

RELIABILITY OF CUBICS FITTED TO TEST RE-TEST HALF-MINUTE HEART RATES DURING EXERCISE

TABLE 5

ZERO ORDER CORRELATIONS AND RELIABILITY COEFFICIENTS OF GROUP IV CASES HAVING TEST-RETEST TREADMILL SCORES
(Test-Retest Correlations in Parentheses)

		Tread-mill (1)	(2)	(4)	(5)	
	TEST A (N = 29)	(.96)				
(2)	Acceleration at 4th Half-Minute	.77	(.63)			
(4)	Standard Exercise Heart Rate	-.47	-.09	(.77)		
(5)	Equation Constant	-.03	-.02	.61	(.85)	
(7)	Slope at 4th Half-Minute	-.01	-.33	-.58	-.02	(.30)
	TEST B (N = 29)					
(2)	Acceleration at 4th Half-Minute	.78				
(4)	Standard Exercise Heart Rate	-.53	-.36			
(5)	Equation Constant	-.11	-.39	.56		
(7)	Slope at 4th Half-Minute	-.23	-.20	-.35	.04	
Group V (N = 33)						
(2)	Acceleration at 4th Half-Minute	.73				
(4)	Standard Exercise Heart-Rate	-.33	.04			
(5)	Equation Constant	.20	.07	.56		
(7)	Flarimeter Blow	.39	.02	-.16	.14	

1st and 2nd tests, using the heart rate measures alone. The correlation between these (r_{11}) is now .80 for the 29 cases. This correlation rises to .9[illegible] when the grossly divergent case is removed. A single case could exert such an effect, of course, only when the N is small.

Since only one set of flarimeter readings was available for Group IV, the test-retest r with regressions including flarimeter score could not be obtained, but it is clear that the inconsistency of the one case would have been tempered by such an inclusion. These considerations tend to indicate that the composite will characterize the response of the subject more reliably than any of the single measures.

The validity (using the score on the maximal treadmill test as the criterion) of the multiple regression prediction attained in both 1st and 2nd tests of Group IV is even better than in Groups I, II, and III. The multiples are developed in the first section of Table 6. These higher coefficients in Groups IV and V may be largely attributed to technical improvements in the conduct of the tests. They therefore yield more trustworthy regressions than Groups I, II, and III.

Regression Reliability from Cross-Application to all Sets of Data. The final step in the estimation of the validity and stability of the multiple regressions consisted in computing multiple correlations for each set of data using alien regressions. It should be clear that if the regressions state the true relationships between the heart rate functions and flarimeter blow and treadmill performance, the multiple R obtained by application of the regression for Group V, for example, to the data of Group IV should not be much lower than the r obtained from the regressions inherent in Group IV.[13]

The R for the cross-applied regression has been calculated from the formula:

$$R = \frac{B_a r_{12} + B_b r_{13} + B_c r_{14} + \ldots B_n r_{1.n}}{\left(B_a^2 + B_b^2 + B_c^2 + \ldots B_n^2 + 2(B_a B_b r_{23} + B_a B_c r_{24} + B_b B_c r_{34} + \ldots B_{n-1} B_n r_{(n-1)n})\right)^{1/2}}$$

---

[13] Editor's note. As pointed out in the footnote on page 10, these correlations may be unduly high because of the fact that the predictor and the criterion variables are not independent. However, the object of this cross-application was to indicate whether the relations which were found were functions of a particular group or existed independent of sampling error. The logic was that if relationships were evident whose interpretation would be helpful in further development of submaximal tests it should be ascertained whether the relationships would be apparent in other samples.

where B_a, B_b, B_c, B_n = standard regression coefficients (betas) from an alien sample, and r's are inherent in the data to which the regression is applied.

The concern here is the theoretical exposition rather than practical application of the regressions. The betas, which express regression relations independently of variability in samples, are presented in Table 6. Those for acceleration and standard exercise heart rate, $B_{12.459}$ and $B_{14.259}$, are quite stable, while those for equation constant and flarimeter blow, $B_{15.249}$ and $B_{19.245}$, are somewhat more variable. Again, as would be expected, the composite represented by the regression is more stable than the individual weight, as shown by the results of cross-applications to the multiple R's in Table 7.

It will be recalled that in Group IV, though full heart rate data were available for the 1st and 2nd tests, only one set of flarimeter readings was available. Thus, in Table 7 there are two sets of heart rate multiples and regressions for Group IV, but only one regression and multiple based on flarimeter. The following alien regressions are therefore available:

A. Group IV-A heart rates
   Group IV-B heart rates
B. Group IV-A heart rates plus flarimeter
C. Group V heart rates
D. Group V heart rates plus flarimeter

Regressions for Groups I to III were not cross-applied to the later sets because the later regressions are considered to be based on somewhat superior data, but the latter regressions were applied to the data of Groups I to III..

If now the R's obtained with inherent regressions are compared with those obtained by cross-application of alien regressions, substantial agreement is found in Table 7. In most instances, R's differ by not more than .05. While, with use of heart rates alone, differences up to -.12 appear in applications to Group II and III data, the addition of flarimeter blow decreases these differences.

The two regressions (Group IV and V involving flarimeter blow) deliver multiple R's over .90 with Groups IV and V data regardless of which is used.[14] They both give multiples of over .80 for Groups I and III and about .70 for Group II which are very near the values given by the inherent regressions. Regarding the latter, it was previously concluded that the notably lower multiple R was due to an abnormally low correlation between standard exercise heart rate and the criterion. It is now clear that it could not have been due to an abnormal regression because alien

---

[14]Actually, if heart rate data alone are considered, there are three regressions which deliver comparable multiple R's, i.e., IV-B included.

TABLE 6

MULTIPLE CORRELATIONS AND STANDARD REGRESSION COEFFICIENTS
(Betas)

Subscripts:

Treadmill Score	1
Acceleration at 4th Half-Minute	2
Standard Exercise Heart-Rate	4
Equation Constant	5
Flarimeter Blow	9

Multiples	Group IV (N = 29) A	Group IV (N = 29) B	Group V (N = 83)
$R_{1.24}$	.87	.82	.81
$R_{1.245}$	.90	.91	.91
$R_{1.249}$	.89	–	.88
$R_{1.2459}$	.91	–	.95
A. Heart Rate Measures			
Betas			
$B_{12.45}$	.7264	.7778	.7186
$B_{14.25}$	-.5781	-.5151	-.6352
$B_{15.24}$	.2394	.4782	.5051
B. Heart Rate Measures with Flarimeter Blow			
$B_{12.459}$	.6159		.7157
$B_{14.259}$	-.5693		-.5689
$B_{15.249}$	.2773		.4334
$B_{19.245}$	.1941		.2546

TABLE 7

MULTIPLE CORRELATIONS OBTAINED WITH INHERENT AND ALIEN REGRESSIONS

Multiple	Group	N	Inherent Regressions	Alien Regressions	
A. Heart Rate Measures ($R_{1.245}$)*				Group IV-A Regression	Group IV-B Regression
	I	98	.83	.83	.79
	II	97	.64	.57	.57
	III	45	.80	.78	.72
	IV-A	29	.90	-	.89
	IV-B	(retest)	.91	.89	-
	V	83	.91	.90	.91
B. Heart Rate Measures with Flarimeter ($R_{1.2459}$)				Group IV-A Regression	
	I	98	.85	.85	
	II	97	.69	.60	
	III	45	.81	.79	
	IV	29	.91	-	
	V	83	.95	.93	
C. Heart Rate Measures ($R_{1.245}$)				Group V Regression	
	I	98	.83	.80	
	II	97	.64	.52	
	III	45	.80	.74	
	IV-A	29	.90	.88	
	IV-B	(retest)	.91	.91	
	V	83	.91	-	
D. Heart Rate Measures with Flarimeter ($R_{1.2459}$)				Group V Regression	
	I	98	.85	.82	
	II	97	.69	.62	
	III	45	.81	.76	
	IV	29	.91	.90	
	V	83	.95	-	

*Subscripts refer to the same variables as in Tables 5 and 6.

regressions give about the same multiples and there is no reason to alter the original conclusion. The above demonstrated stability of the regression equations when applied to alien data shows that the relations they represent are present in all groups.

SUMMARY AND CONCLUSIONS

From complete heart rate records during a preliminary walk and an exhausting treadmill run on 381 young men, aged 18 to 28 and without known pertinent medical defect, a pattern of cardiac response to the activity has been demonstrated. Various functions of a cubic fit to the heart rates during the maximal run have been correlated with treadmill score, a performance criterion of exercise tolerance. Two of these functions -- acceleration at the 4th half-minute, and heart rate at the midpoint of run (Equation Constant) -- are both valid (in terms of treadmill scores) and reliable.[15] Acceleration is particularly important, correlating above .6 in five out of six groups.

Two additional measures, breathholding with the flarimeter, and a standard exercise heart rate taken during the last minute of the preliminary walk, incorporated with acceleration and heart rate at the midpoint of the run, yield a highly valid and reliable relationship with the treadmill test. The proofs are as follows:

1. Of six samples, the multiple correlations between these measures and the criterion are higher than .90 in three, higher than .80 in two, and about .70 in one. There is satisfactory evidence that the last result is due to an atypical relationship between standard exercise heart rate and treadmill score.

2. In one series of test and retest experiments the reliability of the regression formula was found to be .80.

3. Regression equations developed from two sets of data (and in part from a third) were cross-applied to the other sets of data, yielding multiple correlations only slightly lower than those obtained from regressions inherent in each set of data.

It is clear that acceleration in heart rate response to exercise or a close correlate cannot be excluded if a satisfactory test of exercise tolerance is to be devised. While the degree of statistical validity (using the treadmill test as the criterion) attained by the present materials is not approached by any test in the literature and the outcome is of the greatest theoretical significance, the need for practical solution still remains.

It is yet to be demonstrated whether a measure of acceleration can be developed from responses to a "submaximal" test, which approaches the validity and stability of the present one taken from the second derivative of the third degree curve.

---

[15] Editor's note. Heart rate at the midpoint of the run cannot be considered a submaximal test since it cannot be determined except by continuing the run to exhaustion. However, as noted in the footnote on page 10, although it was expected that this research would uncover promising leads for the construction of submaximal tests, none of the original variables may be used directly for such a purpose, but in all cases comparable measures which can be used in a submaximal test must be sought.

SECTION II

EXPLORATORY STUDIES OF SUBMAXIMAL TESTS OF EXERCISE TOLERANCE[16]

INTRODUCTION

In contrast to maximal tests of exercise tolerance wherein the subject was required to work to the point where fatigue causes a considerable decrement in the individual's capacity to maintain the rate of work (exhaustion endpoint), submaximal tests feature a standard amount of work well within the capacity of all or nearly all subjects.

It is apparent that such a submaximal test technique is the only ultimate, practical solution to the measurement of exercise tolerance, not only for reasons of ease of administration, but because complete control is in the hands of the examiner and the test scores do not depend so heavily on the judgment and motivation of the subject. Such submaximal tests are particularly necessary where routine repeat tests are carried out and where the results of the tests are of sufficient importance to the subject that large differences in motivation would affect maximum test performance.

Inasmuch as test performance (work output) is a constant in the submaximal test, it is obvious that the subject's score on the test must be based upon measures of physiological response to the standard exercise and not on the amount of work performed.

The central problem, therefore, is to measure reliably the physiological responses to standard work situations and to validate these measures in terms of an acceptable (standard) criterion of exercise tolerance. To these ends, various physiological measures have been evaluated in terms of two criteria of exercise tolerance: (1) the maximal treadmill test described in Section I of this report, and (2) physical training over a three-month period, i.e., scores on athletic fitness tests (e.g., pullups, pushups, 300-yard dash, etc.) recorded for a three-month period.[17]

In addition to measures of physiological response to exercise taken during and after exercise, the validity and reliability of certain measures taken in the resting state, while reclining, sitting, and stand-

---

16 Originally presented in: Taylor, C., and Franzen, R. Submaximal tests of exercise tolerance. CMR, OSRD. June 1943.

17 These fitness tests and physical training scores are described in: Taylor, C., and Morris, R. Studies of the effect of training on scores in a submaximal test of exercise tolerance. CMR, OSRD. June 1943.

ing, and during changes in posture have also been determined.[18]

A. HEART RATE AND RESPIRATORY METABOLISM DURING STANDARD EXERCISE

Since the standard exercise commonly used in fitness tests has been that of stepping up and down on a stool of defined height for a certain number of times, an exploratory study was undertaken in order to determine in some detail the primary differences between the physiological responses of known criterion groups of subjects.

The subjects employed in this study were classified on the basis of athletic fitness scores given at Stanford University over a period of three months. The High Fitness Group consisted of 10 subjects chosen from the top tenth of the distribution of scores and the Low Fitness group of 11 students from the lowest tenth of the distribution.

That these groups of subjects represent the extremes of physical fitness was further verified by subjecting them to a "maximal pack test" with the results as noted in Table 8.[19]

TABLE 8

DIFFERENCE IN CRITERION GROUPS ON MAXIMAL PACK TEST

	High Group	Low Group	Difference
Mean	535	259	276
Sigma	69.1	70.5	t = 8.61
Range	410-600	[illegible]0-370	p = .001

The lack of overlap in these groups and the high mean differences on the pack test as well as the percentile difference on the athletic fitness scores leaves little doubt that they comprise true extreme criterion groups of physical fitness.

Test Procedure

1. After a preliminary rest period of 10 to 15 minutes, the subject was required to sit quietly for 5 minutes during which the respiratory and heart rate measures were taken.

---

[18] Evidence for the value of existing cardiovascular and respiratory measures of fitness in standard "fitness" tests such as the Schneider, McCurdy-Larson, etc. has also been collected. Some of these data are presented in Appendix [illegible] of the present report.

[19] The maximal pack test is fully described in: Taylor C., and Franzen, R. A maximal pack test of exercise tolerance. CMR, OSRD. June 194[illegible].

2. The subject was then required to mount a stool 15 inches in height 60 times in 90 seconds.[20] An ascent on the stool consisted in moving from a standing position before the platform (stool) to a standing position on the platform, followed by a descent to the original position.

3. Following this exercise period the subject sat down and his physiological recovery was followed for 20 minutes. This 20-minute period was divided into three different stages: Rc(1), the first 2 minutes; Rc(2), the next 3 minutes recovery; and Rc(3), the last 15 minutes.

We have, thus, five experimental periods during which physiological measures were taken: rest, work, Rc(1), Rc(2), and Rc(3).

Measures Recorded

1. Heart Rate
2. Respiratory Rate
3. Total Ventilation
4. Oxygen Consumption
5. Respiratory Quotient
6. Carbon Dioxide Production
7. Per cent Oxygen
8. Per cent Carbon Dioxide

Heart rates, as in the previous studies, were continuously recorded by means of the Henry Cardiotachometer.

Rates of ventilation up to 130 liters per minute and the necessity for continuously following respiratory and metabolic responses required special equipment. While, in general, the technique employed was that usually described as the Tissot-Haldane open circuit method, the gasometer and mask are unique in many respects.

The mask consists of a cupped hard rubber cap with a pneumatic rim to fit the contour of the chin, cheek, and nose, and in which is fixed a one-way valve system to permit drawing inspired air from the room and diverting expired air through an airway to the gasometer. The latter has two chambers and follows the general principle of the Krogh hinged chamber basal metabolism apparatus. However, the chambers are complete quadrants of a cylinder, and contain, when full, 200 liters of gas.

The double chamber feature is especially adaptable to continuous measurements of expired air. Collections are begun in Chamber 1. When that is filled, a valve is thrown, diverting the expired air to Chamber 2 while total volume is measured and gas samples are drawn from Chamber 1. When Chamber 2 is filled the valve is again turned to the chamber which has been emptied and made ready for collection again.

---

[20] See the diagrammatic sketch of the platform in Appendix C of this report.

In this way, the chambers may be used alternately without interruption of the collection of expired air. Gas samples are drawn over mercury and set aside for analysis later. Analyses are carried out in a standard Haldane apparatus following the procedures outlined by Boothby and Sandiford.[21]

Expired air was collected for each of the periods (Rc(1), (2), and (3) and gas samples were drawn from the total air collected during each period. The data are expressed as averages over the particular period being assessed.

This procedure does not allow sensitive depiction of change of state but was adopted because "block" collection of expired air is more accurate than shorter sampling periods. Gas samples were analyzed in duplicate in the Haldane apparatus.

RESULTS

The average heart rate, respiratory ventilation, respiratory quotient, and oxygen consumption for the high and low fitness groups are presented in Figures 4, 5, 6, and 7 respectively.

Examination of these figures reveals that the heart rate differences between the groups (Figure 4) are least during rest and greatest at Rc(1), the first 2 minutes and Rc(2), the next three minutes. This finding serves to illustrate a general trend shown in all the figures, i.e., the tendency for the greatest separation in the physiological measures of these groups to occur during the early periods of recovery from exercise.

Ventilation and oxygen consumption (Figures 5 and 6) all show the greatest mean difference in Rc(1) and all four functions agree in showing the most significant difference at this point.

The statistical differences at the various periods are presented in Table 9. Fisher's t statistic was calculated for the differences in the mean scores of the two groups at each point the measurements were recorded.

Table 9 reveals that all four functions show statistically significant differences in Rc(1) as indicated in Figures 4, 5, 6, and 7. Differences in oxygen consumption are not significant through the experiment although some separation does occur during Rc(1) and Rc(2) as illustrated in Figure 7. This low t-value for oxygen consumption might be taken to indicate that the groups did not differ markedly in the efficiency with which they performed the work. Heart rate and

---

[21] Boothby, W. M., and Sandiford, I. Laboratory manual of the technic of BMR determinations. Philadelphia: W. B. Saunders Co., 1920.

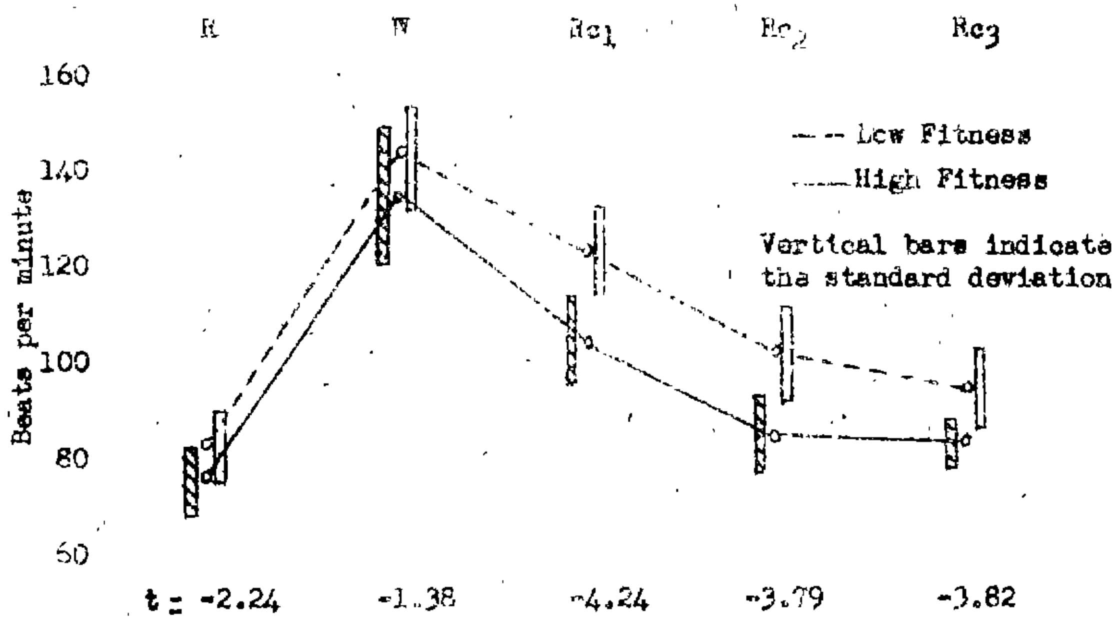


FIGURE 4 HEART RATE

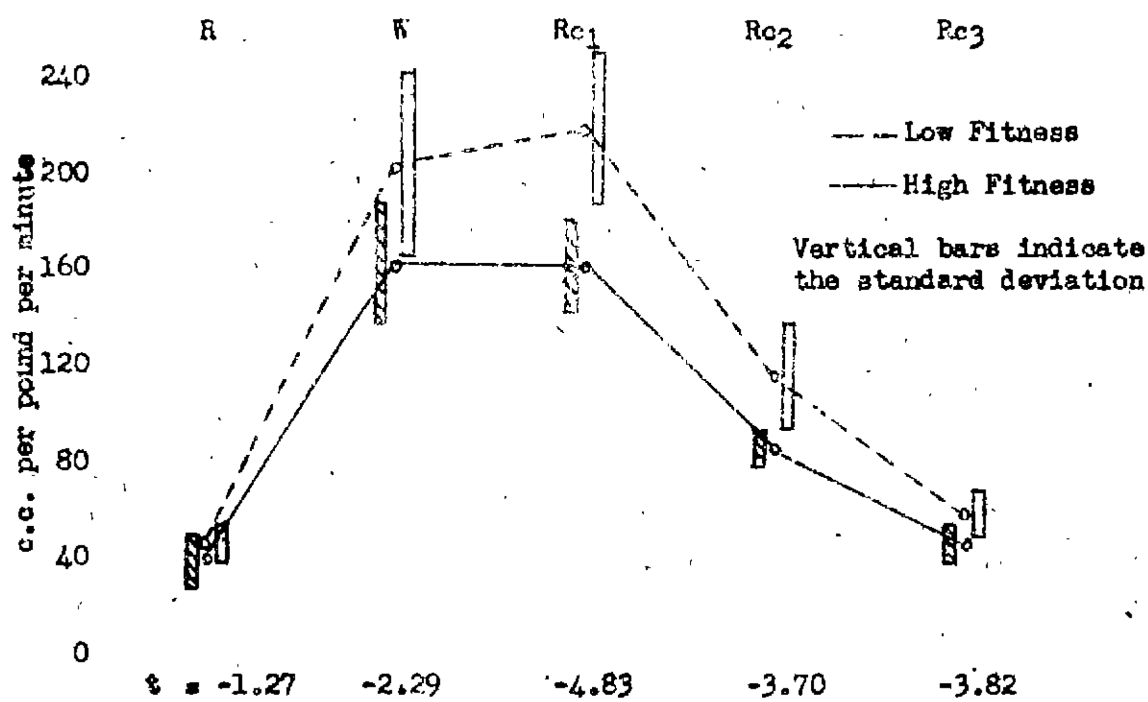


FIGURE 5 TOTAL VENTILATION

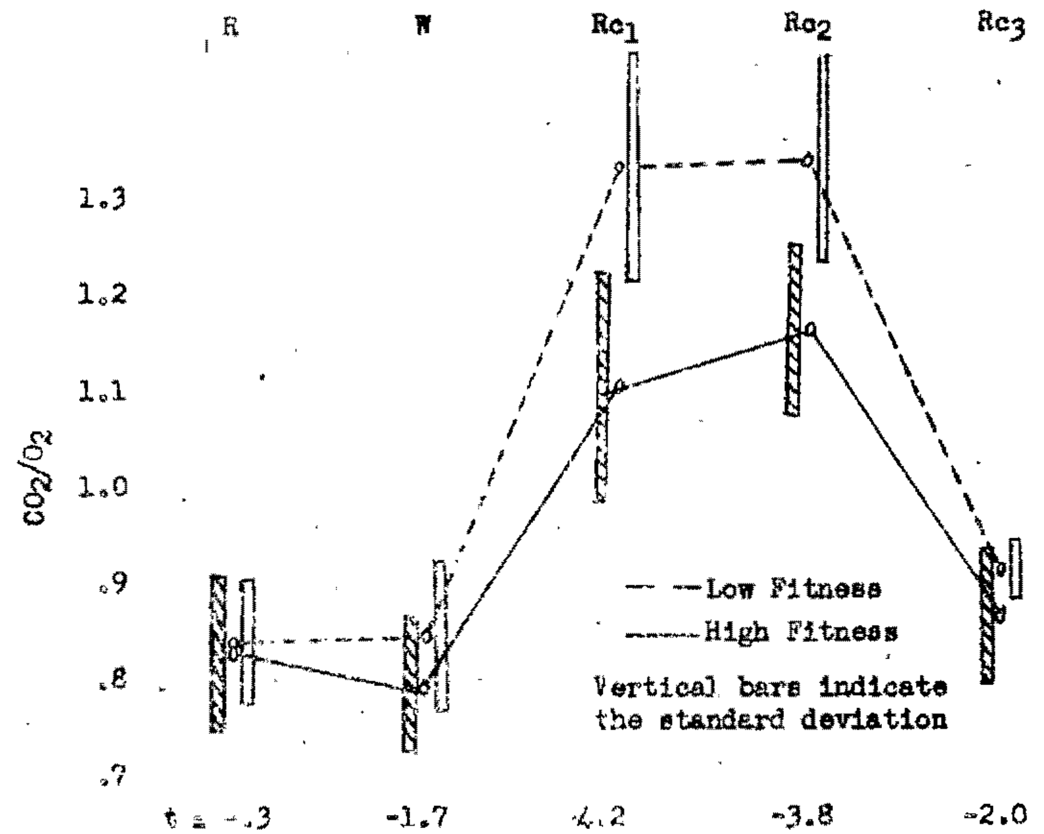


FIGURE 6 RESPIRATORY QUOTIENT

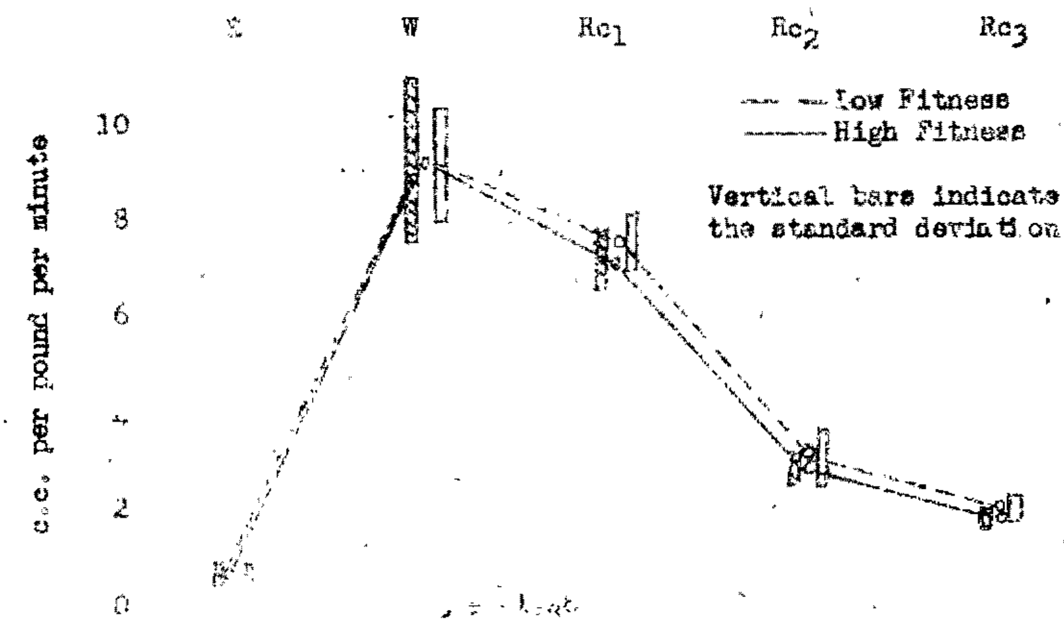


FIGURE 7

TABLE 9

MEANS, STANDARD DEVIATIONS, AND t-VALUES FOR HIGH AND LOW GROUPS

Function	Rest High	Rest Low	Work High	Work Low	Rc(1) High	Rc(1) Low	Rc(2) High	Rc(2) Low	Rc(3) High	Rc(3) Low
Respiratory Rate	11.4	15.1			16.8	22.3	15.7	22.1		
t	-1.04				-2.35		-2.51			
Heart Rate	76.4	84.3	136.9	145.8	107.1	126.3	97.2	103.8	84.9	97.1
	8.50	8.56	16.48	11.32	10.33	9.39	9.71	10.10	6.22	8.17
t	-2.24		-1.38		-4.24		-3.79		-3.82	
Total Ventilation	42.0	46.2	164.9	203.2	163.5	224.8	88.7	119.6	49.2	60.9
	7.91	7.07	26.90	43.73	21.06	34.65	9.30	23.54	6.46	9.36
t	-1.27		-2.29		-4.83		-3.70		-3.32	
Oxygen Consumption	1.87	1.96	9.39	9.32	7.29	7.70	3.15	3.26	2.08	2.18
	.29	.18	1.76	1.29	.67	.56	.38	.50	.19	.27
t					-1.70					
Carbon Dioxide Production	1.55	1.56	7.20	7.96	8.09	10.19	3.69	4.42	1.91	2.00
t	-.18		-1.33		-4.09		-2.69		-1.31	
Respiratory Quotient	.83	.84	.79	.85	1.11	1.33	1.17	1.35	.87	.92
	.08	.06	.07	.08	.12	.12	.09	.11	.07	.05
t	-.3		-1.7		-4.2		-3.8		-2.0	
Percent Oxygen	4.50	4.07	5.55	4.96	4.51	3.54	3.58	2.77	4.29	3.64
	.56	.34	.58	.50	.57	.38	.50	.31	.42	.13
t	2.05		2.56		4.41		4.24		4.41	
Percent Carbon Dioxide	3.72	3.40	4.36	4.08	4.93	4.59	4.18	3.73	3.71	3.33
t	2.27		1.87		1.39		2.47		2.31	

ventilation, on the other hand, are shown to much more clearly differentiate the groups.

SUMMARY AND CONCLUSIONS

These exploratory studies indicate that early recovery from exercise is the point of maximum divergence between high and low fitness subjects in the standard exercise type of test used in this study. They further indicate that oxygen consumption is not a significant variable in such differentiation and that respiratory quotient is so highly related to ventilation that the additional labor necessary to obtain it is not justified (the correlation between respiratory quotient and ventilation for these 21 cases was .84). Ventilation was dropped from further consideration in simple fitness testing because of its low reliability and the technical difficulties of securing reliable spirometry.[22]

From a comparison of the average responses of the high and low fitness groups to a standard exercise test it may be tentatively concluded (within the limitations of the sample) that:

1. Early recovery is the point of maximum divergence between the physiological responses of the criterion groups.

2. Heart rate and ventilation differentiate the groups markedly, but metabolic efficiency is not a significant differential.

B. VALIDITY OF CARDIOVASCULAR-RESPIRATORY MEASURES TAKEN BEFORE AND AFTER STANDARD EXERCISE TESTS

Introduction. This series of investigations was undertaken in order to provide further information on the cardiovascular-respiratory reactions of normal subjects to the short 'standard exercise test' and in order to determine the relationship between these measures and the maximal treadmill test scores.

Comparisons are made of the linear and curvilinear relations; partial correlations which might be meaningful as measures of fitness; multiple regression equations; and first, second, and third order multiple

---

[22] Data justifying these statements may be found in: Taylor, C., and Fransen, R. Maximal tests of exercise tolerance. CMR, OSRD. June 1943. This report is on file with the NRC Committee on Aviation Medicine and a preliminary report is on file with the Committee on Selection and Training of Aircraft Pilots. Calculations of various derived measures from the above data were also determined in order to ascertain whether greater differences than those shown could be obtained between the two groups. However, all were less valid than the primary measures from which they were secured. The derived measures included oxygen debt, oxygen pulse, ventilation cost, ventilation debt, and gross mechanical efficiency. These analyses are on file with the Committee on Selection and Training of Aircraft Pilots.

correlations.

Subjects. Five groups of subjects were employed in these studies. All of the subjects ranged in age between 18 and 28 years and all were "normal" in the sense that they possessed no known defects which might have influenced their tolerance for exercise. The general make up of the five groups of subjects is as follows:

Group[23]	N	Description
A	59	Soldiers from Moffett Field
B	60	College men from Stanford University
I	98	College men from Stanford University
II	97	College men from Stanford University
III	49	Soldiers from Moffett Field

Two sets of the data (Groups I and II) contain the same variables and are really random divisions of the materials of one complete experimental investigation. The other groups contain data which are slightly different; it was impossible during these studies to make all variables available for analysis.

Procedure. The following experimental routine was followed throughout these investigations in the administration of the 'standard exercise test' and the recording of the physiological measures.

1. Preliminary Reclining Rest. During this period the subject was merely required to lie down for 10 to 20 minutes. No observations were made.

2. Pre-exercise Standing. At the end of the Reclining period, the subject was required to stand for a period of 4 minutes during which the heart rate and the blood pressures were taken. All measurements were taken during the last 2 minutes of standing. Heart rates were 30 second counts from 15 to 45 seconds in the minute. Both systolic (appearance of sounds) and diastolic (muffling point) pressures were recorded.

3. Standard Exercise Tests. Following the Standing period, the 2 minute standard exercise test was administered. During these 2 minutes the subject was required to make 60 ascents and descents from a 15-inch stool as described earlier in this paper. Immediately upon completion of the test the subject sat down.

4. Recovery Period. This period lasted approximately 6 minutes following the standard exercise. During this time the following measurements were taken:

   a. Post-exercise Heart Rate. The heart rate (sitting) recorded during the first minute (15 to 45 seconds) following recovery.

---

[23] Groups A and B contain subjects not used previously. Groups I, II, and III are the same as those in Section I.

b. Flarimeter Blow (sitting). Starting at the second minute after the standard exercise.

At the conclusion of the flarimeter blow test, the subject was again required to stand up and the following measures were taken:

c. Heart Rate (recovery) and systolic and diastolic blood pressures. Measurements were taken during 3½ to 5½ minutes of recovery.

All subjects in Groups I, II, III, and A were also administered the maximal treadmill test described in Section I of this paper and their exercise tolerance scores calculated.

The 60 subjects in Group B, however, did not take the maximal treadmill test. They were given two submaximal (standard) tests according to the above procedure -- one week intervening between the tests. This procedure was followed with this group in order that the reliabilities (self-correlations) could be determined for the primary and derived measures of physiological response to the standard tests.

Results. Table 10 presents means and standard deviations for anthropometric and blood pressure measures for all experimental groups, except Group B. It will be noted that Group A differs somewhat but that Groups I, II, III are quite comparable.

The relation between the various measures and treadmill score is represented by r's and etas. Examination of the linear correlations (r's) reveals that none of the measures in Table 10 was significantly correlated with the criterion (score on the maximal treadmill test). No correlation is over .27, and even where such an r occurs in one sample (for example, chest width) it drops to insigificance in other samples. In several instances (see Tables 10 and 11) curvilinear regression was tested by making best cubic fits on the scatter plots and obtaining rho, that is, a measure of the goodness of this curvilinear fit. These rho's in all cases showed similarity to the Pearson r's, suggesting that the departures from linearity, shown by the etas, were for the most part due to irregular patterns which yield no more significant relationship than is given by linear correlation. The anthropometric and blood pressure measures may, therefore, be eliminated from consideration as possible predictors of exercise tolerance as measured by the maximal treadmill test.

Correlational matrices for heart rates and flarimeter blow are presented in Table 11 and multiple R's arranged in order of magnitude in Table 12, for Groups I, II, III, and A. Though all of the measures of heart rate tend to indicate significant correlation with the criterion (maximal treadmill test) pre-exercise standing and recovery heart rates are less consistent than post-exercise heart rate.

The latter, combined with flarimeter blow, yields the highest first order multiple in Table 12 for Groups A, I, and II, (R = .52, .66,

TABLE 10

CORRELATIONS OF ANTHROPOMETRIC AND BLOOD PRESSURE DATA WITH TREADMILL SCORES

	GROUP A (N = 59)				GROUP I (N = 98)				GROUP II (N = 97)				GROUP III (N = 49)			
Variable	Mn	S.D.	r	eta	Mn	S.D.	r	eta	Mn	S.D.	r	eta	Mn	S.D.	r	eta
Treadmill Score	47	9			51	11			52	9			49	10		
Age (years)	23	2	-.15	.43												
Height (cm.)	176	5	.10	.48	179	6	.01	.41	179	7	.25	.46	179	7	.26	.47
Chest Width (cm.)	30	2	.24	.34	31	2	-.13	.31	32	2	-.11	.44	32	2	.27	.41
Hip Width (cm.)	29	2	-.06	.44	29	2	-.17	.48	29	2	.04	.26	29	2	-.06	.39
Vital Capacity (liters)	5	1	.16	.41												
Systolic Pressures (mm. Hg.)																
Pre-ex. reclining	123	11	.04	.33												
Pre-ex. sitting	121	10	.05	.37												
Pre-ex. standing	114	11	-.01	.65*	129	12	.01	.34	126	12	.12	.43	130	12	.18	.42
Post-ex. standing					133	13	-.12	.38	132	14	.02	.23				
Diastolic Pressures (mm. Hg.)																
Pre-ex. reclining	78	8	-.11	.37												
Pre-ex. sitting	79	7	-.03	.39												
Pre-ex. standing	81	7	-.25	.53**	87	9	.08	.28	86	10	.15	.46	78	9	.19	.42
Post-ex. standing					87	10	.06	.37	87	10	.22	.51				

* rho = .18   ** rho = .25

TABLE 11

CORRELATION MATRICES: HEART RATE AND BREATHHOLDING DATA

(1) Treadmill Score
(2) Pre-exercise Standing Heart Rate
(3) Recovery Standing Heart Rate
(4) Post-exercise Heart Rate
(5) Flarimeter Blow
(6) Pre-exercise Sitting Heart Rate

Group		(1)	(2)	(3)	(4)	(5)	Mean	S.D.
Group A								
	(4)	-.36					110	17
	(5)	.46			-.27		22	8
	(6)	-.19			.48	.001	61	8
Group III								
	(2)	-.47(.47)*					92	15
	(4)	-.42	.72				124	17
	(5)	.22(.23)*	-.04		-.04		21	7
	(6)	-.51	.68		.27	.01	76	11
Group I								
	(2)	-.34(.42)*					88	14
	(3)	-.53	.78				96	16
	(4)	-.61	.55	.72			125	21
	(5)	.37(.39)*	-.07	-.17	-.22		25	10
Group II								
	(2)	-.06(.07)*					88	12
	(3)	-.19	.82				95	15
	(4)	-.33	.64	.65			125	17
	(5)	.43(.44)*	-.10	-.21	-.13		28	9

* rho computed from a cubic regression

TABLE [illegible]

MULTIPLE CORRELATIONS: HEART RATE AND BREATHHOLDING DATA

(1) Treadmill Score
(2) Pre-exercise Standing HR
(3) Recovery HR
(4) Post-exercise HR
(5) Flarimeter Blow
(6) Pre-exercise Sitting HR

Multiple	GROUP A (N = 59)	GROUP I (N = 93)	GROUP II (N = 97)	GROUP III (N = 49)
$R_{1.23}$		.54	.25	
$R_{1.24}$		.61	.38	.49
$R_{1.25}$		.49	.43	.51
$R_{1.26}$				.54
$R_{1.34}$		.63	.33	
$R_{1.35}$		.60	.45	
$R_{1.45}$	.52	.66	.51	.48
$R_{1.46}$	.35			.59
$R_{1.56}$	.50			.56
$R_{1.234}$		.63	.41	
$R_{1.235}$		.61	.46	
$R_{1.245}$		.66	.55	.53
$R_{1.246}$				.59
$R_{1.256}$				.58
$R_{1.345}$		.67	.52	
$R_{1.456}$	.52			.63
$R_{1.2345}$		.67	.55	
$R_{1.2456}$				.64

and .51, respectively). This is the highest multiple obtained in Group A, and nearly as high as the third order multiples for Groups I, and II. Group III differs from Group A, I, and II, in that it shows a relatively high criterion r for sitting heart rate. In this group, the highest multiples contain sitting heart rate, while flarimeter blow, correlating only .22 with the criterion, exerts much less influence. Little weight should be given to the results for Group III in this connection, however, because of the relatively small N (49) and because of the apparently atypical relations of sitting heart rate and flarimeter blow. Flarimeter blow usually has a higher correlation (Groups A, I, II) and sitting heart rate a lower correlation (Group A) with the criterion than was found for Group III. Generally speaking, then, post-exercise heart rate and flarimeter blow reveal the best correlations in the data.

That they are more stable also is shown by the betas for heart rate and flarimeter blow in Table 13. Betas involving pre-exercise standing, recovery, and sitting heart rate are susceptible to large fluctuation, even with change of sign in two cases.

The nearest approach to a reliable prediction of the criterion, which could be developed from these data, would be a battery of measures containing post-exercise heart rate and flarimeter blow, combined in the ratio 3:2. Unfortunately, the multiple R with the criterion of such a battery (approximately .55) and the variation in betas would very strongly limit the validity and reliability of such a practice. Hence, the conclusion here is of little practical significance at the moment.

Table 14 presents the test-retest reliability coefficients from Group B. Of the heart rates, post-exercise heart rate has the highest reliability ($r_{11} = .74$). This measure is also more reliable, than any of the derived heart rate indices, such as change on standing, change on exercise, and change on recovery. Systolic pressure is uniformly more reliable than diastolic but the systolic change on standing is very unreliable.

It is significant that post-exercise heart rate, the most reliable of the heart rates and flarimeter blow, the most reliable of all the measures, should also show greatest validity. When it is considered that the standard error of a measure[24] for a variable having a reliability coefficient of .80 is .44 of the sigma of a distribution, it is clear that reliabilities below .80 are very unsatisfactory for purposes of prediction. The only hope for the use of such physiological materials is that "batteries" of measures might, through compensatory interplay, yield higher reliability than the component measures taken singly.

__Summary.__ Four groups of subjects (303 cases) were given maximal treadmill tests and a submaximal standard test routine consisting of

[24] $SE_{meas} = \sqrt{1 - r_{11}}$

TABLE 13

BETAS (STANDARD SCORE COEFFICIENTS FOR MULTIPLE REGRESSION): HEART RATE AND FLARIMETER BLOW

(1) Treadmill Score
(2) Pre-exercise Standing HR
(3) Recovery Heart Rate
(4) Post-exercise HR
(5) Flarimeter Blow
(6) Pre-exercise Sitting HR

Beta	GROUP I (N = 98)	GROUP II (N = 97)	GROUP III (N = 49)	GROUP A (N = 59)
$Beta_{12.345}$	-.12	.33		
$Beta_{13.245}$	.27	-.10		
$Beta_{14.235}$	-.43	-.42		
$Beta_{15.234}$	.24	.39		
$Beta_{14.56}$			-.31	-.19
$Beta_{15.46}$			.24	.41
$Beta_{16.45}$			-.43	-.10
$Beta_{14.5}$	-.56	-.28	-.43	-.24
$Beta_{15.4}$	.25	.40	.24	.40

TABLE 14

RELIABILITY COEFFICIENTS: GROUP B

(N = 60)

		r_{11}
A. Primary Measures:		
Reclining:		
	(1) Heart Rate	.62
	(2) Systolic Pressure	.68
	(3) Diastolic Pressure	.51
Pre-exercise Standing:		
	(4) Heart Rate	.68
	(5) Systolic Pressure	.75
	(6) Diastolic Pressure	.55
Post-exercise Sitting:		
	(7) Heart Rate	.74
	(8) Flarimeter Blow	.80
Recovery Standing:		
	(9) Heart Rate	.66
	(10) Systolic Pressure	.79
	(11) Diastolic Pressure	.67
B. Derived Measures:		
(4) - (1)	Change on Standing	.57
(7) - (4)	Change on Exercise (a)	.68
(9) - (4)	Change on Exercise (b)	.50
(7) - (9)	Change on Recovery	.69
(5) - (2)	Change on Standing	.41

heart rates, blood pressures and breathholding. Another group (60 cases) were given the standard test routine twice, a week intervening between tests. From a correlational analysis of these data, with respect to validity in terms of the treadmill criterion and test-retest reliability, the following conclusions have been drawn:

1. Age, height, weight, hip width, chest width, and vital capacity do not correlate significantly with the treadmill criterion.

2. Systolic and diastolic blood pressures, reclining, sitting and standing, and standing after exercise similarly do not show significant relationship with the criterion and are not stable enough to provide evidence for a weighted scale.

3. Of the heart rates taken before and after exercise, the post-exercise count most consistently correlates with the criterion. r is about -.40. Post-exercise heart rate is also the most reliable of the heart rates taken.

4. Post-exercise breathholding with the flarimeter also correlates about .40 with the criterion and is more reliable than heart rates or blood pressures.

5. Breathholding and post-exercise heart rate yield a multiple correlation of approximately .55. The validity of this battery is considered, therefore, too low for predictive purposes under the present circumstances, particularly since it has little stability.

6. The use of curvilinear relations does not improve the regressions.

C. A FOUR-MINUTE (SUBMAXIMAL) TREADMILL TEST

Introduction. In the preceding sections of this report it has been demonstrated that the subject's exercise tolerance as measured by the maximal treadmill test cannot be measured to a completely acceptable extent by the physiological responses to standard tests of exercise tolerance such as stepping up and down on a chair or stool of prescribed height.

It was also shown, however, in Section I of the present report, that the second derivative of the cubics (for the heart rate measures) set at two minutes from the start of the test run correlated significantly with the duration of the run (score). On the basis of this evidence, it may be hypothesized that performance on the maximal treadmill test might be predicted from the trend of the heart rate measures recorded during an abbreviated treadmill test. It is obvious, however, that if shorter segments of the heart rate curve are used for prediction there must be some loss in reliability and validity due to variability or irregularity of the early heart rates.

Subjects. The data gathered on two groups of subjects were used in this investigation as follows:

... of 94 subjects who were given repeat ... the short form of the treadmill test (described below) in order ... the reliability of the measure.

... V. This group was made up of the 83 subjects of Group V who ... with the maximal treadmill test during the studies reported ... of this report. The validity of the short form of the test ... by determining the relationship between the early segment of the curve and the final treadmill scores.

Procedures. The 94 subjects of Group VI were tested on the short form of the test according to the following procedure (it will be noted that ... is the same as for the maximal treadmill test except that the subject is stopped at the end of 4 minutes):

1. Preliminary Walk. Each subject was required to walk on the treadmill for a period of 3 minutes. During this period, the speed of the treadmill was set at 108 meters per minute and the pitch at 5 per cent.

2. Intermediate Rest. At the end of the preliminary walk the subject was required just to stand for a period of 3 minutes. During this period the ... were made beginning 45 seconds after the preliminary walk.

3. Short Test Run. After 3 minutes rest (above) the subject was required to run for four minutes on the treadmill with the speed set at 162 meters per minute. The beginning pitch was set at 5 per cent and thereafter was raised 1 per cent after each one-minute period of running.

Results. All but two of the subjects were able to complete the run. For these two subjects then, the so-called short form amounted to a true maximal test of exercise tolerance. For the remainder of the subjects the test was definitely submaximal.

Since the maximal treadmill scores were not available for the 94 subjects who were given the test-retest for reliability purposes, the data from the 83 subjects of Group V (see Section I) were used to determine the validity of the 4-minute test. This analysis was completely feasible with the heart rates from the first four minutes of the maximal test. It must be assumed, however, that this segment of the maximal test for these subjects would not be essentially different were the subjects actually only run for 4 minutes.

Table 15 gives the zero-order and multiple correlations which express the validity of measures from the 4-minute test. The following points may be noted:

1. Measures of heart rate curvature (the second derivative of a second degree curve) and third-order differences, are correlated with the criterion to much less extent than the accelera-

TABLE 15

ZERO-ORDER AND MULTIPLE CORRELATIONS: FOUR-MINUTE TREADMILL TEST

(1) Treadmill Score
(2) Standard Exercise Heart Rate
(3) Flarimeter Blow
(4) Terminal Heart Rate
(5) 3rd Difference (Finite difference of the third order)
(6) 2nd Derivative (2nd Degree)
(7) Post-exercise Recovery Index

(N = 83 cases)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(2)	-.33	(.63)*					
(3)	.39	-.10	(.86)				
(4)	-.34	.75	.03	(.86)			
(5)	.30	-.33	.27	-.19	(.33)		
(6)	.23	-.51	.19	-.08	.29	(.33)	
(7)	-.29	.30	-.15	.19	-.30	-.22	(.63)

*Reliability r's (test-retest) in parentheses. N = 94 and consists of cases other than those used to determine the correlations in the matrix except for (7) which has an N of 60.

$R_{1.34} = .53$ $R_{1.347} = .55$

$R_{1.23} = .48$ $R_{1.345} = .54$

$R_{1.3457} = .56$

$1.3467 = .56$ $R_{1.234567} = .58$

tion computed for the cubic fit in the maximal test. Therein lies the failure of this modification. In abbreviating the test, a large part of the intermediate portion of the curve has been lost which, it will be recalled, was straighter and less sloped, the longer the subject was able to run (see Section I, p. 6 ff).

Of course, the terminal upswing of heart rate on approach to exhaustion has also been lost except in a very few cases. Hence, the third differences (taken from first, second, fourth, and eighth half-minute intervals) fail to correlate well with the criterion. Instead, in most cases only the initial curvature is available, describing in the four-minute period a second degree parabola. The second derivative of this curve does not, in itself, adequately differentiate the subjects ($r_{16} = .23$). Thus, it appears that the four-minute segment of the heart rate curve is too short in most cases to demonstrate the trend on which significant prediction could be based.

In further substantiation of these facts, Table 16 should be considered. Groups IV-A[25] and V heart rates at the second, fourth, and eighth half-minute intervals are intercorrelated. With moderate to low correlation with the criterion and high intercorrelation, partial variance, on which multiple R's must be based, is very small. This is, in part, why the curve functions lose validity in the abbreviated treadmill test.

2. Standard exercise heart rate and flarimeter blow show a combined relationship with the criterion ($R_{1.23} = .48$) comparable with results in other series of experiments (see Table 14). It is interesting to note that terminal heart rate is more reliable than standard exercise heart rate, though correlating highly with it.

3. Post-exercise recovery index is a measure of heart rate deceleration after the preliminary walk. Its correlation with the criterion ($r = -.29$) and contribution to the multiple are unsubstantial. The negative sign of the r, indicating lesser deceleration in fitter subjects, is contrary to the accepted dictum that pulse recovery is more rapid in the fit individual. As shown by the negative correlation between standard exercise heart rate (during the last minute of preliminary walk) the fit man departs less from his resting condition, under the stress of exercise. His heart rate thus rises less and recovers relatively more slowly than that of the less fit man. The reliability of this measure is low.

---

[25]Group IV-A is the same group of subjects analysed in the studies reported in Section I of this report (N = 29). (See page 16.)

TABLE 16

INTERCORRELATIONS OF EARLY HEART RATES

(1) Treadmill Score
(2) Standard Exercise Heart Rate
(3) Heart Rate at 2nd Half-Minute
(4) Heart Rate at 4th Half-minute
(5) Hearth Rate at 8th Half-minute

	(1)	(2)	(3)	(4)
Group IV-A (N = 29)				
(2)	-.47			
(3)	-.42	.84		
(4)	-.55	.78	.89	
(5)	-.53	.77	.83	.90
Group V (N = 33)				
(2)	-.25			
(3)	-.01	.66		
(4)	-.21	.77	.79	
(5)	-.30	.74	.72	.93

Group V Pack Pulses* (N = 33)

	TM	(1)	(2)	(3)
(1)	-.23			
(2)	-.26	.94		
(3)	-.31	.93	.94	
(4)	-.33	.86	.91	.96

*Described on page 47.

4. The fifth-order multiple ($R_{1.234567}$ = .58) indicates the low validity of the four-minute test. Since the zero-order r's are all low, the regressions would be expected to show instability in other applications.

5. The reliability coefficients, computed from the data on the 94 cases of Group VI range from .86 for both terminal heart rate and flarimeter to .33 for the second derivative. Calculation of the reliability for the batteries of measures was not made because of the doubtful validity of the four-minute test.

Summary. From an analysis of heart rates during the first four minutes of treadmill experiments on 83 subjects and test-retest four-minute treadmill experiments on 94 cases, it has been sought to develop measures of heart rate trend, analogous to acceleration in the previous maximal treadmill studies (Section I), which would predict the treadmill score. From these studies, the following tentative conclusions may be drawn:

1. Measures of heart rate curvature, third differences and the second derivative of a second degree fit of the four-minute heart rate curves, yield correlations no higher than .30 with the criterion.

2. While a fifth-order multiple of .58 was obtained with these data, none of the variables correlate higher than .4 with the criterion. Such a battery could hardly fail to be erratic if cross validated on other samples.

3. From these considerations it may be concluded that the four-minute treadmill test does not provide a satisfactory basis for prediction of maximal treadmill performance.

D. A SUBMAXIMAL PACK TEST

Introduction. Because of the simplicity of administration of a standard test of the "stool-stepping" variety, a further effort was made to validate a submaximal test of this sort in terms of the maximal treadmill test. There is even further reason to study this type of test further since evidence in the preceding studies casts some doubt on the feasibility of developing a reliable and valid short form of the treadmill test itself.

Accordingly, a test routine was devised, including post-exercise heart rate, flarimeter blow, and a series of work periods during which the subject was required to carry weights of increasing magnitude during the submaximal test. The graded series of work was included in the hope that characteristics of the heart rate increase on such graded pack test work might prove to be a valid index of physical fitness as indicated in the treadmill studies.

Procedure. Specifically, the [illegible] Pack Test was conducted as follows:

1. Period of Rest. The subject was required to remain completely at rest for a 5-10 minute period prior to the test.

2. Standard Work Period. During this period, the subject was required to ascend and descend from a 17-inch stool at the rate of 40 cycles per minute for a period of 90 seconds.

3. At the end of the Standard Work Period the subject was stopped for approximately 5 minutes and the following measures recorded:

   a. Post-exercise Heart Rate. A 20-second count beginning 10 seconds after exercise.

   b. Recovery Heart Rate. A 20-second count beginning 4 minutes after exercise.

   c. Flarimeter Blows. The first flarimeter blow was begun 45 seconds after the Standard Work Period was stopped; the second blow was then begun 10 seconds following the end of the first. A 30 mm. Hg. manometer with a slow leak was employed. The score was taken as the sum of the two blows.

4. The subject was then required to don a pack sack and 5 minutes after the end of the Standard Work Period, was again required to step up and down on the same stool at the same rate for the following 1-minute periods:

   a. First Work Period. The subject stepped up and down on the stool, but no load was put in the pack sack.

   b. Second Work Period. During this 1-minute period the subject was required to carry 15 pounds of weight in the pack sack while performing the standard exercise.

   c. Third Work Period. Again the subject stepped up and down from the 17-inch stool for a period of 1-minute, but during this period he was required to pack an additional 15 pounds (making 30 lbs. of weight).

   d. Fourth Work Period. No additional weight was required during this 1-minute period. The subject merely performed the standard exercise again while packing 30 lbs. of weight in the pack sack.

5. After each of the above 1-minute work periods the subject was required to stand still for a period of 30 seconds while a 20-second heart rate count was made. Four heart rate scores were therefore recorded for each subject during the pack test.

Subjects. Eighty-three subjects were given both the submaximal pack test described above and the maximal treadmill test (see Section I), in order to obtain data for determination of validity of the measures recorded during the pack test. In addition, retests were carried out with many of the original 83 and pairs of tests on other subjects. In all, a total of 94 test-retest cases were available for the determination of the reliability of the various functions.

Results. Nearly all of the subjects were able to negotiate the entire series of 1-minute pack tests. Approximately 3% of them, however, were unable to complete the series and stopped when unable to maintain the required rate of ascent and descent.

The correlation matrix, and the highest multiples of first to fourth order, are given in Table 17. It is clear from the zero-order r's that none of the measures recorded during the submaximal pack test possess any considerable validity. Post-exercise heart rate and flarimeter blow combined yield a multiple R of only .42, although judging from the previous investigations (see Table 12), this correlation is usually about .50.

The heart rates, taken during the pauses in the graded series of work-loads ("pack pulses") possess high intercorrelation and for this reason only one of them, the fourth, has been carried into the matrix. This measure, the fourth pack heart rate, is more reliable and valid than post-exercise heart rate.

It is apparent that measures of heart rate behavior in relation to the graded work will not differentiate subjects as effectively as the curve functions in the treadmill test. This is well shown in Table 16. The intercorrelations of immediately contiguous heart rates are in all cases .94 or higher, and the first and fourth correlate .86. The complete failure of third differences to correlate with the criterion is further evidence that the shape of the curve described by the graded work heart rates does not differentiate high from low treadmill performers.

Considering the possibility of using the submaximal pack test as a screening device, "cut-off" points have been determined from inspection of the scattergrams of the variables with treadmill score. The significance of difference between the categories distinguishing extreme groups from the rest of the cases was then tested by chi-squared. These data are given in Table 18.

The high group was defined by four variables, three of which were in the correlation analysis. The fourth, P52, is the number of work periods in the graded series necessary to elevate heart rate to 52 or higher. Use of the selected combinations A, D, BC, ABC, ABD, ABCD (see Table 18 for definitions) detected 24 men who had higher than average treadmill score, and [illegible] with lower than average. On the other hand, it failed to classify 16 subjects who were above average on the treadmill. The chi-squared is 19.[illegible]4 (P [illegible] .00002). This table illu-

TABLE 17

CORRELATIONS: SUPMAXIMAL PACK TEST

(1) Treadmill Score
(2) Post-exercise Heart Rate
(3) Flarimeter Blow
(4) Fourth Pack Heart Rate
(5) Third Differences (Pack Pulses)
(6) Recovery Heart Rate

(N = 83)

	(1)	(2)	(3)	(4)	(5)	(6)
(2)	-.19	(.70)*				
(3)	.39	-.07	(.75)			
(4)	-.32	.73	-.20	(.82)		
(5)	.12	.08	.10	.09	(.06)	
(6)	-.05	.64	-.09	.55	.15	(.66)

$R_{1.34} = .46$ $\quad$ $R_{1.346} = .49$

$R_{1.23} = .42$ $\quad$ $R_{1.345} = .48$

$R_{1.35} = .40$ $\quad$ $R_{1.236} = .44$

$R_{1.36} = .39$ $\quad$ $R_{1.235} = .43$

$R_{1.3456} = .50$ $\quad$ $R_{1.23456} = .50$

$R_{1.2346} = .49$

$R_{1.2345} = .48$

* Figures in parentheses are test-retest r's. Reliabilities are not on the 83 cases represented in the matrix. N of reliabilities is 94, but not the same 94 that were used to determine reliability of the four-minute treadmill.

TABLE 13

SUBMAXIMAL PACK TEST CUT-OFFS

High Group

Categories: A = P52 4 and 5
B = 4th Pack Heart Rate under 52
C = Flarimeter over 41 seconds
D = Post-exercise Heart Rate under 41

	Above Av. Treadmill	Below Av. Treadmill	
High Group (A, D, BC, ABC, ABD, ABCD)	a 24	b 6	30
All other cases	c 16	d 37	53
	40	43	83

Δ = 888 - 96 = 792
Chi-squared = 19.04 P = < .00002

Low Group

Categories: A = P52 1 and 2
B = 4th Pack Heart Rate over 58
C = Flarimeter Blow under 35

	Below Av. Treadmill	Above Av. Treadmill	
Low Group (AB, C, ABC, F*)	a 24	b 10	34
All other cases	c 25	d 27	52
	49	37	86

Δ = 648 - 250 = 398
Chi-squared = 4.25 P = .04

* F represents subjects unable to complete graded series of work tests.

lustrates the issues involved in use of such a screening test. By broadening the criterion it is possible to define more high men, but at the same time, including more low men. Conversely, sharpening the cut-off criterion to exclude more of the poor men will pick fewer good men.

A similar cut-off procedure to classify the low men proved to be considerably less satisfactory. While these categories chose 24 low men, they also chose 10 high men and failed to pick 25 other low men. Chi-squared is 4.25 ($P = .04$).

This analysis of the submaximal test is not entirely adequate because the sample (83 cases of Group V) was not ideal for validation purposes, since the extremes were not well represented. Other cut-offs might have been tried. Nevertheless, the results indicate the serious limitations of the standard submaximal test and define the degree of confidence which can be placed in classification by such a test. Since a much more promising type of submaximal test is afforded by a variable length procedure further consideration of the present type has been abandoned, at least for the present.

Summary. A test routine was set up including Post-Exercise Heart Rate, Flarimeter Blow, and a series of work periods during which the subject was required to carry weights of increasing magnitude during the submaximal test. The characteristics of the heart rate increase on the graded pack test were examined to determine if they could serve as valid indices of physical fitness as indicated in the treadmill studies. Eighty-three subjects were given both the submaximal pack test and the maximal treadmill test. Retests were carried out with many of the original 83, while in all, a total of 94 test-retest cases were available for the determination of the validity of the various functions.

Briefly, the main findings of this section of the investigation showed that:

1. None of the measures recorded during the submaximal pack test appeared to possess any marked validity.

2. Measures of heart rate behavior did not appear to differentiate subjects as effectively as the curve functions in the treadmill test.

3. Cut-off points were determined from inspection of the scattergrams of the variables with treadmill score and the significance of the difference between the categories distinguishing extreme groups from the rest of the cases was tested by chi-squared. The chi-squared P values for both the high and low groups fell below .05. In the high group the selected combination of variables detected 24 men with higher than average treadmill score and 6 with lower than average, but failed to classify 16 subjects who were above average on the treadmill. In the low group the categories chose 24 low men and 10 high men, but failed to pick 25 other low men.

SECTION III

THE CEILING HEART RATE TREADMILL TEST

INTRODUCTION

In the preceding section it was shown that the four-minute treadmill test was too short to provide valid measurement of the trend of adaptation to exercise. Although it was possible to identify the poorer fitness group, i.e., the men who could run only 4-6 minutes before exhaustion, those subjects with average tolerance (able to run 8-10 minutes before exhaustion) could not be differentiated from the superior subjects (those able to run 12-15 minutes).

It was then hypothesized that if each subject could be run on the treadmill test to a given proportion of his total curve a test of sufficient length would be available to give heart rate curves of sufficient validity to predict maximal performance.

The Ceiling Heart Rate test described in this section is the outcome of the preliminary studies undertaken with a view to identifying a segment of the heart rate curve which would be reasonably predictive of the total curve for each individual.

Procedure. Following the hypothesis stated above, the first problem involved is the selection of a "cut-off" criterion point which might result in segments of the heart rate curve proportion to the total. Such a criterion is not only essential to the theoretical analysis of the proportional curve hypothesis, but is even more important from the standpoint of developing a submaximal test of exercise tolerance which might prove of practical value in the field.

A "cut-off" point to be practicable must be determinable during the course of the test run on the treadmill so that the test can be stopped at some point before actual exhaustion is reached.

At first, a standard maximum heart such as 85 or 95 beats per half-minute was considered. However, inspection of the data and preliminary trial tests using this technique showed that such a cut-off point would fail to provide proportional lengths of the curve because individual differences in general level of the heart rate curve were too great.[26]

For example, two equally good subjects (as measured by length of run on the treadmill test) may have heart rates, though similar in trend, different by 5 to 10 beats per minute throughout the entire test, including the final heart rate attained. It must be assumed, therefore, that

---

[26] The absolute "cut-off" scheme does however yield data of considerable validity as will be shown later in this paper.

with equal treadmill test performance such differences do not necessarily reflect variation in cardiac efficiency. It then becomes important that a "cut-off" criterion be determined which takes into account the individual heart rate pattern, i.e., an individual "cut-off" criterion founded on a baseline for that individual.

Resting heart rates, while giving some clue to heart rate level during work, fail badly in individual cases. A considerably more representative baseline may be obtained during a period of strenuous but standardized exercise.

The average half-minute heart rate, taken from the period 1' 45" to 3' 15" of the treadmill run, hereinafter to be designated "Early Run Heart Rate" (ERHR), was chosen for study. With ERHR as an estimate of heart rate level, a "cut-off" heart rate could be defined which reasonably suits the individual being tested.

Since heart rate approaches a maximum with the onset of exhaustion, which terminates the maximal test performance, it is evident that if a reliable forecast of this exhaustion level of heart rate from ERHR can be made, a given heart rate short of the maximum to serve as a "cut-off" criterion can be predicted so that the individual will not have to be run to exhaustion on the test. Such prediction can be made according to the formula:

$$\bar{Y}c = r_{xy} \frac{\sigma y}{\sigma x} X + \left\{ M_y - r \frac{\sigma y}{\sigma x} M_x \right) - K \right\}$$

where, Y = maximum heart rate (the average of three last half-minute rates)
X = ERHR (Early Run Heart Rate)
M and σ = means and standard deviations of X and Y
K = a subtrahend determining the point below maximum heart rate at which "cut-off" will be made
$\bar{Y}c$ = predicted "cut-off" criterion for the individual

The accuracy of this system depends, of course, on the correlation coefficient, (r), which determines the standard error of the estimated Yc. The final validity of the method, however, rests upon the magnitude and stability of multiple regressions obtained from three variables which develop from the data when the cut-off is used. These variables are: (a) ERHR, (b) the second derivative of the heart rate curve extending from start of run to the cut-off point,[27] and (c) CT, which is the time to reach the cut-off heart rate.

---

[27]<u>Editor's note</u>. It is important to note that this cut-off point which is defined as a given heart rate short of the maximum at the exhaustion level was predicted by the above formula and not obtained by running subjects to exhaustion.

K, the number of heart beats per half-minute below the maximum, at which the cut-off was to be made was selected as one standard error[28] on the judgment that it stops most of the subjects sufficiently short of maximum performance to make the test submaximal, but not so early that the subject is stopped before reliable heart rate trends are established.

Subjects and Data. The necessary data for the statistical analyses undertaken in these studies were available on all of the subjects of Groups I, II, III, and IV tested during the investigations reported in Section I of this report. At this time it was felt unnecessary to subject another large group of subjects to the maximal treadmill test. It will be necessary, however, to provide new groups of subjects for a rigid cross validation of the results of these analyses.

Results. The essential data on which these cut-offs were based, are presented in Table 19. The correlations between ERHR and maximum heart rate at point of cut-off are very consistent over the several groups of data, averaging approximately .55. A fairly high, but inconsistent negative correlation, is found between the criterion and ERHR while correlations with maximum heart rate are for the most part insignificant.

The standard error of estimating maximum heart rate (and also the cut-off heart rate which is a constant subtrahend) based on a correlation of only .55 is relatively large and would seem to seriously limit the validity of the cut-off method. However, the partial correlations ($r_{13.2}$), i.e., the correlation between treadmill score and maximum heart rate with ERHR held constant, indicate that high treadmill scores tend to be above the regression of maximum heart rate on ERHR, and low scores below. In other words, the good men tend to run beyond their maximum heart rate predicted from ERHR while poor men stop short of it. As can be seen in Table 19 these partial correlations are above .70 in four of the six samples and approximately .30 in the other two.

The standard error of estimate then is systematic to the extent of these partial correlations, and affects the correlation of CT with treadmill score considerably less than would otherwise be assumed from such an error of estimate. The cut-off criterion, which is a parallel of the regression line will, therefore, tend to run the poor men a longer time, and the good men a shorter time than would be true if ideal proportions of total heart rate span could be determined. This curtails the CT variance, but does not seriously lower the correlation with total treadmill.

The prediction equations are also given in Table 19. It will be noted that the regression coefficients and constants vary somewhat with the means and standard deviations from sample to sample. One explana-

---

[28] That is, $\sigma_x\sqrt{1 - r^2_{xy}}$ when x is maximum heart rate and y is early run heart rate.

TABLE 19

INTERCORRELATIONS BETWEEN EARLY AND MAXIMUM HEART RATE AND TREADMILL SCORE WITH PREDICTION EQUATIONS

(1) Treadmill Score
(2) Early Run Heart Rate
(3) Maximum Heart Rate

Variable	Groups	N	Mean	S.D.	Correlations (1)	(2)	$r_{13.2}$
ERHR (2)	I	98	88.0	6.26	-.68		
	II	97	87.7	5.73	-.37		
	III	45	91.6	7.01	-.62		
	IV-A	29	89.2	5.22	-.53		
	IV-B	(retest)	87.6	4.78	-.58		
	V	83	87.5	4.47	-.37		
Maximum Heart Rate (3)	I	98	100.9	4.89	-.12	.54	.39
	II	97	101.1	4.17	.02	.50	.26
	III	45	97.1	4.27	.17	.53	.75
	IV-A	29	98.4	3.61	.27	.62	.90
	IV-B	(retest)	97.7	3.93	.20	.51	.70
	V	83	97.6	3.76	.43	.52	.79

PREDICTION EQUATIONS

I	98	Yo = .422 (x) + (63.8 - 4*)
II	97	Yo = .366 (x) + (69.0 - 4)
III	45	Yo = .325 (x) + (67.4 - 4)
IV-A	29	Yo = .431 (x) + (60.0 - 3)
IV-B	(retest)	Yo = .416 (x) + (61.3 - 3)
V	83	Yo = .440 (x) + (59.1 - 3)

*Differences in these cut-off constants are due to differences in the standard error of estimate.

a very adequate basis for general application of the cut-off point to men of this age range.

The zero-order and multiple correlations between the measures developed by the cut-off method and the treadmill score (criterion) are given in Table 20. The data are arranged to emphasize three points: (a) the comparability between multiples containing CT and those containing the second derivative; (b) the validity of multiples containing CT; and (c) the validity of the simplified measure, T 85-95, the treadmill duration starting with a pulse of 85 and ending when the pulse reaches 95.

It will be recalled that one of the primary objectives in developing the cut-off procedure was to obtain a segment of the total curve which would portray a greater proportion of the heart rate trend during work. It was anticipated that longer curves would yield validity approaching that for the total curves (as shown in Section II).

Such has proven to be the case. Multiples containing the derivative of a second degree fit to the heart rate from beginning of run to the cut-off point proved to be in the neighborhood of .7 when combined with standard exercise heart rate and flarimeter blow.[29] This constitutes a statistical demonstration of the logical connection between the total curve validity and the present cut-off curve. However, the combination of CT, Standard Exercise Heart Rate, and Flarimeter Blow correlates slightly higher with the criterion in three of the four samples used for this comparison, and the close comparability of the multiples, $R_{1.234}$ and $R_{1.235}$ demonstrates (pending cross-validation of the results) that the cut-off measure, CT, might be substituted for the curve function. It may be, in fact, a slightly different measure of the same tendencies.

Summary Analysis of Stability of the Ceiling Heart Rate Tests. The multiple correlations between the treadmill criterion and the three independent variables -- Standard Exercise Heart Rate, Flarimeter Blow, and Time to Cut-off representing submaximal measures -- were high enough to justify using them as a workable battery predicting maximal performance. This battery is hereafter referred to as "Ceiling Heart Rate Test A."

The multiple correlations between the treadmill criterion and the three independent variables -- Standard Exercise Heart Rate, Flarimeter Blow, and T 85-95 -- likewise are sufficiently high to justify these as a workable battery. This latter battery is hereafter referred to as "Ceiling Heart Rate Test B."

Ceiling Heart Rate Test A has certain theoretical advantages over Ceiling Heart Rate Test B and the multiples with the criterion are in all cases but one higher, though the difference is slight throughout. The next logical step in the treatment of these data was therefore the analysis of the stability of these two batteries of measures.

---

29 Since only a portion of the total curve is now used a second degree curve is adequate. The second derivative is a constant in these second degree curves.

TABLE 20

CORRELATION MATRICES AND MULTIPLE R'S: CEILING HEART RATE TEST

(1) Treadmill Score
(2) Standard Exercise Heart Rate
(3) Flarimeter Blow
(4) Time to Cut-off
(5) Second Derivative
(6) T 85-95

GROUP I (98)

	(1)	(2)	(3)	(4)
(2)	-.46			
(3)	.37	-.10		
(4)	.62	-.47	.17	
(5)	.51	-.07	.11	.66
(6)	.50	-.36	.10	

$R_{1.24} = .65$ $R_{1.26} = .58$

$R_{1.34} = .68$ $R_{1.36} = .59$

$R_{1.234} = .70$ $R_{1.236} = .66$

$R_{1.235} = .72$

GROUP II (97)

	(1)	(2)	(3)
(2)	-.06		
(3)	.46	.004	
(4)	.55	-.33	.25
(6)	.53	-.22	.20

$R_{1.234} = .65$ $R_{1.236} = .64$

TABLE 20 (Continued)

CORRELATION MATRICES AND MULTIPLE R'S: CEILING HEART RATE TEST

GROUP III (45)

	(1)	(2)	(3)
(2)	-.42		
(3)	.24	-.12	
(4)	.52	-.39	.15
(6)	.53	-.31	.16

$R_{1.234} = .59$ $R_{1.236} = .61$

GROUP IV-A (29)

	(1)	(2)	(3)	(4)
(2)	-.47			
(3)	.59	-.06		
(4)	.60	-.51	.41	
(5)	.40	-.11	.55	.31
(6)	.52	-.43	.40	

$R_{1.24} = .63$ $R_{1.26} = .59$

$R_{1.34} = .71$ $R_{1.36} = .66$

$R_{1.234} = .75$ $R_{1.236} = .74$

$R_{1.235} = .73$

GROUP IV-B (retest)

	(1)	(2)	(3)	(4)
(2)	-.57			
(3)	.59	-.20		
(4)	.75	-.62	.51	
(5)	.60	-.75	.54	.61
(6)	.69	-.52	.47	

$R_{1.24} = .78$ $R_{1.26} = .71$

$R_{1.34} = .79$ $R_{1.36} = .75$

$R_{1.234} = .80$ $R_{1.236} = .73$

TABLE 20 (Continued)

CORRELATION MATRICES AND MULTIPLE R'S: CEILING HEART RATE TEST

GROUP V (83)

	(1)	(2)	(3)	(4)
(2)	-.33	-.10		
(3)	.39	-.63	.26	
(4)	.59	-.42	.13	.73
(5)	.51	-.45	.27	
(6)	.57			

$R_{1.24} = .59$ $R_{1.26} = .57$

$R_{1.34} = .64$ $R_{1.36} = .62$

$R_{1.234} = .64$ $R_{1.236} = .63$

$R_{1.235} = .62$

Each of the tests may best be represented by the multiple regression equation with the criterion (treadmill test). This equation will, of course, determine the score of individuals to whom the test will be applied in the future. The question might then be asked whether the equation of any one set of materials is applicable to another set.

It is known that the multiple regression equation represents the best weighting of the factors involved in the group from which the equation itself was derived. How well it applies to another group is a product of the reliability of the elements and the changing conditions of the sample. Tests might be acceptably reliable and still show low stability because conditions existing in any one sample are different from those existing in another.[30]

Table 21 gives the Betas of the two Ceiling Heart Rate batteries in the three large groups of subjects. It was thought better not to deal with Betas of the three smaller groups, though cross-application of

---

[30] Editor's note. Decisions on the suitability of these tests for field use should, of course, be based on further validation through the use of additional criteria.

the other multiple regression batteries to them was made. It is apparent from inspection of the Betas in Table 21 that there is a good deal of similarity in the standard score weights obtained from these three groups.

Cross-application of Betas provides the definitive results. It is difficult to judge the importance of a discrepancy between Betas except when the cross-applications are made. The R for the cross-applied regression has been calculated from the formula:

$$R = \frac{B_a r_{12} + B_b r_{13} + B_c r_{14} + \ldots B_n r_{1.n}}{\left(B_a^2 + B_b^2 + B_c^2 + \ldots B_n^2 + 2(B_a B_b r_{23} + B_a B_c r_{24} + B_b B_c r_{34} + \ldots B_{n-1} B_n r_{(n-1)n}\right)^{1/2}}$$

where B_a, B_b, B_c, B_n = standard regression coefficients (Betas) from an alien sample, and r's are inherent in the data to which the regression is applied.

By using this formula all hazards which are due to changed variance are excluded. Since the problem of variance to be expected from groups to which a final test may be applied must be studied as a separate problem, it is proper here to make the cross-application with Betas. The stability of patterns of standard score weights as derived from different groupings is then being tested.

Table 22 presents the results of these cross-applications. It is apparent that the Group I regression used in Ceiling Heart Rate Test A, when applied to other groups, gives almost the same multiple correlation that is obtained when each group uses its own regressions. An exception to this is the application to Group II which seems to be atypical. The same is true of the application of Group V regressions but this time without the exception. Group II regressions, however, fail to recover a multiple as high as the one which is inherent in the original relations. In all cases except the application to Group V, Group II regressions obtain multiple correlations considerably lower than the inherent ones.

The same relations hold for cross-applications of regression to Ceiling Heart Rate Test B. Here, again, the multiples obtained by using the regressions of Group I are only a little lower than those that are inherent, except in the case of Group II. All multiples using Group V regressions are only a little short of the inherent ones, with the possible exception of Group IV-A which falls from .74 to .67. Use of Group II regressions again results in considerable diminution of multiple correlations from those that are obtained when best weightings are computed for each set.

It may tentatively be concluded from these results that regressions obtained in Group V represent a stable pattern. These regressions are

TABLE 21

THE BETAS OF THE CEILING HEART RATE BATTERIES IN THE THREE LARGE GROUPS

(1) Treadmill Score
(2) Standard Exercise Heart Rate
(3) Flarimeter Blow
(4) Time to Cut-off
(6) T 85-95

	GROUP I (N = 98)	GROUP II (N = 97)	GROUP V (N = 83)
Using Ceiling Treadmill			
$B_{12.34}$	-.21	.11	.05
$B_{13.24}$	.27	.33	.26
$B_{14.23}$	.48	.50	.55
$R_{1.234}$	.70	.65	.64
Using T 85-95 Treadmill			
$B_{12.36}$	-.30	.04	-.09
$B_{13.26}$	.30	.37	.26
$B_{16.23}$	.36	.47	.46
$R_{1.236}$	.66	.64	.62

TABLE 22

MULTIPLE R'S OBTAINED FROM THE APPLICATION OF ALIEN REGRESSION COEFFICIENTS TO CEILING HEART RATE BATTERIES

(1) Treadmill Score
(2) Standard Exercise Heart Rate
(3) Flarimeter Blow
(4) Time to Cut-off
(6) T 85-95

TIME TO CUT-OFF

Groups	$R_{1.234}$ Inherent	$R_{1.234}$ using: Group I Regression	Group II Regression	Group V Regression	N
I	.70	.70	.63	.68	98
II	.65	.58	.65	.64	97
III	.59	.58	.48	.52	43
IV-A	.75	.73	.66	.75	29
IV-B	.80	.80	.76	.78	29
V	.64	.62	.63	.64	83

T 85-95

Groups	$R_{1.236}$ Inherent	$R_{1.236}$ using: Group I Regression	Group II Regression	Group V Regression	N
I	.66	.66	.56	.63	98
II	.64	.55	.64	.62	97
III	.61	.59	.51	.58	43
IV-A	.74	.72	.64	.67	29
IV-B	.78	.79	.71	.73	29
V	.63	.61	.61	.63	83

very much like those obtained in any of the other groups, even atypical Group II. They represent the formulae for arriving at scores in Ceiling Heart Rate Tests A and B.

It next seemed desirable to test the reliability of this regression equation directly. Using a multiple regression (not Betas but b's) of Group V, predictions were made on IV-A and IV-B. The scores obtained represent two applications of Ceiling Heart Rate Test A to 29 cases. Since the Flarimeter Blow was only taken once, it was necessary to include the same score for both applications. However, since the test-retest correlation of this test is .85 or better, the spurious addition to reliability of the regression prediction from this source is small. The correlation between the two regression scores, computed separately, is .91.

The principle of the ceiling tests is to select a heart rate span in the progression toward exhaustion and to determine work done during this span. The less curvature there is in the heart rate development and the more gradual the slope, the more work there will be.

Ceiling Heart Rate Test A uses a varying cut dependent upon heart rate level and Test B uses the same cut-off for all individuals.

It is important that further studies be undertaken to investigate further Ceiling Heart Rate Tests which include the same cut-off for all cases. If this method could be employed to gain results as good as those for Ceiling Heart Rate Test A, it would incorporate certain practical advantages for tests of this type.

APPENDIX A

EXPLORATORY STUDIES

APPENDIX

EXPLORATORY STUDIES

INTRODUCTION

Previous to undertaking the studies described in the text a variety of exploratory investigations were carried out in order to determine the feasibility of employing the maximal treadmill test as a measure of exercise tolerance and to test the predictive value of the standard measures of cardiovascular-respiratory response to exercise or physical exertion. It was possible also to investigate the relationship between certain anthropometric measurements, the treadmill test, and the physiological measures.

The studies described in this appendix are admittedly exploratory in nature. They served primarily to define the direction of further experimental investigations aimed at the construction of a battery of measures which would reliably predict exercise tolerance as measured by the treadmill test.

SUBJECTS

A group of 100 subjects was selected at random from an available 2000 enlisted men at Moffett Field, California. All of the subjects were certified as physically fit, i.e., they had been classified as I-A by the draft boards which insured their being relatively free from any defects which would handicap them for military duty. All of the men had had a minimum of three weeks military training and so had experienced the major part of the "toughening up" process. They ranged in age from 18-22 years (a period during which large variations in physical capacity are not expected to occur).[31]

PROCEDURE

All subjects were subjected to the maximal treadmill test and their treadmill score calculated. Cardiovascular and respiratory measures and certain anthropometric measurements (e.g., chest width, age, height, etc.) were recorded for each man.[32]

---

[31] These facts should indicate beyond a doubt that the group is relatively homogeneous. A test of exercise tolerance which will provide a large spread in measurements taken in such a group could therefore be expected to differentiate quite well between subjects in the general population.

[32] The treadmill test and the method for computing the scores on this test are described in Section I of the report. Cardiovascular, respiratory, and anthropometric measures and procedures for recording them are described elsewhere in the text, either in the body of the report or in this Appendix.

These measurements were then subjected to detailed statistical analyses in order to determine the relationship between such measures and the treadmill scores (exercise tolerance). It was hoped that out of such analyses could be derived a maximized predictive battery of physiological measures which could be more reliably employed in the field as a test of 'Fitness' than those tests in current use.

RESULTS

The mean treadmill score for the 100 cases employed in this study was 10.41 with a standard deviation of 3.59 and a range of scores from 3 to 22. The treadmill test is capable of differentiating between subjects even in such a homogeneous group as this. Differentiation should be even better therefore in a random population of individuals.

Anthropometric Measures. Anthropometric measurements were analyzed in this study for two primary reasons:

1. To determine the relationship between body dimensions and exercise tolerance as measured by the treadmill test.

2. To determine the interrelationships between the treadmill test scores, the physiological measures, and the anthropometric measurements.[33]

The following anthropometric measurements were recorded for all 100 subjects:

1. Age, recorded to the nearest year.

2. Body weight to the nearest one-half kilogram.

3. Height to the nearest centimeter, standing erect with back to measuring rod.

4. Leg length. Derived from the difference between standing and sitting height and with no pressure.

5. Chest width. Measured to nearest centimeter at the nipple level.

6. Hip width. Measured with firm pressure at the greatest width of bi-iliac diameter to the nearest centimeter.

7. Deviation from Pryor Weight. The deviation of body weight from

---

[33] It was reasoned that if such relationship was found to exist in any positive degree in this homogeneous group of subjects that it might be necessary to standardize the marks and measurements in terms of the anthropometric factors as well as exercise tolerance and physiological response.

the standard [illegible] from Pryor Tables.[34]

8. Surface Area. Measured in square meters of body surface, taken from the DuBois chart.[35]

9. Vital Capacity. The best of three trials measured in liters and corrected for body temperature.

10. V.C./S.A.
surface   Vital capacity in liters per square meter of body surface.

In Table 23-A are presented the correlations of each of the anthropometric measures and the maximal treadmill score for the 100 cases. Also presented is the standard deviation (S.D.), the mean, and the coefficient of variation (C.V.) for each of the measures.

Of these measures, only chest width can be considered highly significant. On the whole, the results are not such as to justify any conclusion concerning the relationship between treadmill performance and anthropometric measures.[36]

In Table 23-B are given the correlations between weight, height, and chest width and the short test physiological measures. Here, only the correlation between weight and the ratio P.P./D.P. was found to be significant. According to the data given in this table, then, the relationship between anthropometric measures and functional response to standard exercise is too low to be considered of importance to the present study.

Since vital capacity has been given so much prominence as a measure of physical fitness the analysis of these scores was pursued somewhat further in the present investigation. Accordingly, the relationship between vital capacity (a functional measure) and surface area, height, weight, and chest width (structural measures) was determined (see Table 24).

Considering the high correlation between vital capacity and surface area (.509), the measurements were put on a comparable basis by expressing them in terms of liters per square meter of body surface. When this is done, the correlation between treadmill score and vital capacity dropped from .239 to .161. Similarly, when chest width was partialled

---

34 Pryor, H. B. Width-weight tables for boys and girls from 1 to 16 years; for men and women from 17 to 24 years. Stanford: Stanford University Press, 1936.

35 Carpenter, T. M. Tables, factors, and formulas for computing respiratory exchange and biological transformations of energy. 3rd Ed. Washington, D. C.: Carnegie Institution, Publication No. 303B, 1939.

36 Editor's note. It should be noted, however, that such measures could not be expected to show marked differentiation in a group as homogeneous as the one employed in this study. The possibility of such relationship existing in the general population cannot therefore be discarded on the basis of these data.

TABLE 23

ANTHROPOMETRIC MEASURES

A. Correlation with Treadmill Score (N = 100)

Measure	Mean	S.D.	r	C.V.
Age	22.93	1.95	-.172	7.82
Standing Height	176.48	5.47	.201*	3.10
Leg Length	84.90	4.05	.200*	4.77
Body Weight	70.34	7.65	.168	10.88
Chest Width	30.17	1.86	.341**	6.15
Hip Width	29.04	1.46	.035	5.03
Surface Area	1.86	.11	.190	5.97
Vital Capacity	5.03	.58	.239*	11.44
V.C./S.A.	2.70	.26	.161	9.72
Dev. Pryor Wt.	-7.90	13.07	.030	

B. Correlation with Short Test Measures (N = 59)

	Weight	Height	Chest Width
Recovery Heart Rate	.163	.010	.064
Heart Rate Change	.116	-.117	.034
Flarimeter Blow	-.121	.160	.157
Ratio P.P./D.P.	.239*	(1)	.169

(1) Correlation between height and standing diastolic pressure is .042.

* Attains 5 per cent level of significance.

** Attains 1 per cent level of significance

TABLE 24

CORRELATION BETWEEN VITAL CAPACITY AND STRUCTURAL MEASURES

Surface area and vital capacity	.509
Body height and vital capacity	.498
Body weight and vital capacity	.418
Chest width and vital capacity	.366

out, the correlation with treadmill score dropped to .129.

It may be tentatively concluded, therefore, that for this group of subjects, vital capacity has little relationship with the treadmill performance apart from such structural measurements as surface area and chest width and consequently has little place in a test of the functional capacities of men of this type.

<u>Resting Circulatory Functions</u>. Resting circulatory measures were taken in a quiet room. The subject, upon reporting to the laboratory in the basal state, reclined on a cot while the reclining heart rate was counted and blood pressures obtained. After 5 minutes observation he took a sitting position for five minutes followed by a seven-minute standing period. The circulatory observations were made during each minute for the three positions. Heart rates were counted with a stethoscope over the heart. Counts were begun at the 15th second and ended at the 45th second point of each minute. The rates were expressed as beats per minute. Systolic (1st point) and diastolic (4th point) readings were made as closely to the 30 second point as possible.

Observations were made in each of the 17 minutes of the resting period. The readings for the first two minutes, during which adjustment takes place, were omitted in the computation of the average heart rates and blood pressures.

Table 25 presents the means, standard deviations, and coefficients of correlation for the individual cardiovascular measures and the correlation between these measures and the maximal treadmill performance scores. Also given are the critical ratios of the difference between certain standing and reclining measures.

The results in Table 25 suggest that there is an inverse relationship between fitness scores and resting heart rate (the r, however, is not statistically significant in this population). Since this conclusion has been borne out by other studies[37] the only explanation of the lack of significance of the r found here is the relative homogeneity of the group. It will be noted that the average heart rate for these 100 cases was [illegible] beats per minute which is regarded as characteris-

---

[37] Hambly, W. D., Hunt, [illegible], Parker, [illegible], Mumby, M. S., and Warner, S. G. Tests for physical efficiency. <u>Guy's Hospital Report</u>, [illegible], 357-385.

[illegible] Exercise. 3rd Ed. [illegible]

TABLE [illegible]

RESTING CIRCULATORY FUNCTIONS

Measure	Correlation with treadmill Score	N	Mn	S.D.	C.V.
Reclining					
Heart rate	-.166	98	61.45	8.20	13.35
Systolic	.039	98	123.69	11.01	8.90
Diastolic	-.171	98	78.41	7.85	10.01
Pulse pressure	.154	98	46.18	9.57	20.72
Sitting					
Heart rate	-.137	100	65.80	8.32	12.64
Systolic	.132	100	122.68	10.72	8.74
Diastolic	.000	100	80.56	7.01	8.70
Pulse pressure	.097	100	42.72	9.46	22.14
Standing					
Heart rate	-.138	98	77.85	10.70	13.74
Systolic	.007	98	114.70	11.37	9.91
Diastolic	-.250*	98	81.10	7.52	9.27
Pulse pressure	.198*	98	34.06	9.84	28.29
P.P./D.P.	.258*	59	.42	.14	33.33
Change Reclining to Standing					
Heart rate	-.006	98	17.12	8.56	
Systolic	+.001	98	-8.82	8.46	
Diastolic	-.103	98	3.47	5.80	
Pulse pressure	+.060	98	-11.51	9.19	

*Attains 5 per cent level of significance.

		Reclining	Standing	Difference	S.E. (diff.)	C.R.*
HR.	Mn	61.45	77.85	+16.40		
	SD	8.20	10.70		13.48	1.22
SP.	Mn	123.69	114.70	-8.99		
	SD	11.01	11.37		15.83	.57
DP.	Mn	78.41	81.10	+2.69		
	SD	7.85	7.52		10.87	.25
P.P.	Mn	46.18	34.06	-12.12		
	SD	9.57	9.84		13.73	.88

$$*C.R. = \frac{M_1 - M_2}{S\ (diff.)}$$

tic of a very fit person.

It will be noted that there is no significant relationship between any of the reclining or sitting, cardiovascular measures and no relationship between the change in these measures when the subject changes from a reclining to a standing position. Diastolic blood pressure, pulse pressure, and P.P./D.P. do, however, show significant correlation with the treadmill score.

These data, therefore, indicate that, within this sample of supposedly "fit" individuals, and as compared to the treadmill score, high fitness is characterized by a tendency to high pulse pressure, lower diastolic pressure and a higher pulse pressure/diastolic pressure ratio, and possibly by a somewhat lower heart rate.

Before accepting the foregoing analyses as conclusive, it should be noted that the circulatory reactions to active standing may deal with an aspect of fitness not properly measured by the treadmill test and, therefore, weighting assigned to such variables as standing diastolic pressure or the ratio P.P./D.P. may not be appropriate in the assessment of general physical fitness. It should also be recalled that those measures which show no relationship in this highly selected population might prove to be of some predictive significance in a more heterogeneous group.

<u>The Double Work Test</u>. An attempt was also made with this group of subjects to determine the validity (in terms of relationship with treadmill scores) of the physiological responses to shorter tests of exercise tolerance. Although several possibilities were explored, the Double Work Test seemed the most promising during these preliminary studies.

In this test the subject was required to ascend and descend a 15-inch exercise platform at the rate of 30 times per minute. Every subject was run as follows: Five minutes of rest; 1 minute of work; 2 minutes of rest; 2 minutes of work; and then 10 minutes of recovery.

<u>Cardiovascular Measures</u>. Four indices of heart rate and blood pressure were selected as the physiological responses to exercise to be analysed in these studies:

- "A" The sitting pre-exercise values.
- "B" The average of the two readings obtained in the minutes following the first and second work periods, i.e., 2nd and 6th minutes.
- "C" The difference between values used in "B" served as a measure of slope.
- "D" A measure of recovery, represented by the mean of readings taken from the 2nd to the 5th minute after the second work period.

The means and standard deviations of these indices along with their correlation with the maximal treadmill test score are presented in Table 26.

TABLE 26

CIRCULATORY DATA - DOUBLE WORK TEST

Measure	Mean	S.D.	Correlation with Treadmill Score
Heart Rate:			
Index "A"			-.137
" "B"	102.00	13.78	-.299**
" "C"	18.10	7.52	-.259**
" "D"	75.75	14.58	-.216*
2nd min. recovery	77.35	17.55	-.293**
2nd min. recovery - resting HR	13.83	14.26	-.270**
Pulse Pressure:			
Index "A"	42.72	9.46	.165
" "B"	90.08	16.93	.174
" "C"	17.20	12.12	.104
" "D"	63.06	12.52	-.033

*Attains the 5 per cent level of significance.
**Attains the 1 per cent level of significance.

It will be noted that Indices B, C and D, and 2nd min. recovery heart are the only measures which possess significant relationship with the treadmill score.[30]

[30]These data are further analyzed and discussed in the original report by Craig Taylor, entitled Studies in physical fitness I, 1941. Op. cit. (Referred to in Footnote 5.)

This report also presents data which demonstrate the relationship between respiratory and metabolic measurements and the treadmill score and the correlation between breathholding tests (Flarimeter Blow) and the treadmill scores. Most significant among these data are the following correlations with treadmill performance: (a) Gross mechanical efficiency correlates .595 with the treadmill and is the only respiratory-metabolic measure attaining significance (b) A smaller r (-.304) was found for the difference between resting and recovery oxygen consumption. (c) Flarimeter Blow II (the last blow taken during the recovery period following a 15-minute rest period) correlated .450 with the treadmill score (N = 59).

APPENDIX B

RELATIONSHIP OF STANDARD "FITNESS" TESTS TO THE TREADMILL TEST

APPENDIX B

RELATIONSHIP OF STANDARD "FITNESS" TESTS TO THE TREADMILL TEST

INTRODUCTION

As part of the exploratory studies discussed in Appendix A, certain of the standard tests being currently employed in the field as measures of physical fitness were administered to as many of the subjects as were available for the study and the relationship between these tests and the maximal treadmill test of exercise tolerance determined.[39]

STANDARD FITNESS TESTS

The following standard tests were administered to various subjects selected from the original group of 100 involved in the early exploratory studies on whom scores on the maximal treadmill test were available:

1. The Schneider Index[40]
2. The McCurdy-Larson "Organic Efficiency" Test[41]
3. Crampton Test[42]
4. The McCloy Two-Variable Test[43]
5. The Stone Test[44]
6. The Pulse Ratio Test[45]

---

[39] A brief summary of some of the pertinent literature relative to these tests and other minor investigations involving these tests are given in a report by Dr. Craig Taylor entitled <u>Studies in physical fitness I</u>, 1941. <u>Op. cit.</u> (Referred to in Footnote 5.)

[40] Schneider, E. C. A cardiovascular rating as a measure of physical fatigue and efficiency. <u>J. Amer. Med. Ass.</u>, 1920, 74, 1507-1510.
Schneider, E. C. <u>Physiology of muscular activity</u>. 2nd Ed. Philadelphia: W. B. Saunders Co., 1941

[41] McCurdy, J. H., and Larson, L. A. Measurements of organic efficiency for the prediction of physical condition. <u>Res. Quart. Amer. Ass. Health and Phys. Educ.</u> Ass. (Supplement), 1935, 6, 11.
Larson, L. A. A study of the validity of some cardiovascular tests. <u>J. exp. Educ.</u>, 1939, 7, 214-220.

[42] Crampton, C. W. A test of condition. <u>Med. News</u>, September 1905.

[43] McCloy, C. H. <u>Tests and measurements in health and physical education</u>. Craft, 1939.

[44] See: McCurdy, J. H., and Larson, L. A. <u>Op. cit</u>. This test was scored according to their revisions.

[45] Tuttle, W. W., and Dickinson, R. E. A simplification of the pulse-ratio technique for rating physical efficiency and present condition. <u>Res. Quart. Amer. Ass. Health and Phys. Educ.</u>, 1938, 9, 73-80.

RESULTS

The relationship between these standard tests and those of the measures on which the tests are based are presented in Table 27.

TABLE 27

RELATIONSHIP OF STANDARD TESTS TO THE TREADMILL TEST

Standard Test	r with Treadmill	N
1. Schneider Index	.148	50
2. McCurdy-Larson Test	.061	32
a. Sitting Diastolic	.000	100
b. Standing Diastolic	-.250	100
c. Length of Flarimeter Blow after 20 Sec. exercise	.082	32
d. Diff. between Normal Standing Pulse and Pulse Rate 2 Min. after exercise	-.188	32
e. Standing Pulse Pressure	.198	98
f. Vital Capacity/Body Surface	.061	32
3. Crampton Test	-.086	50
4. The McCloy Test	-.014	50
5. Stone Test	.146	50
6. The Pulse Ratio Test	-.231	50

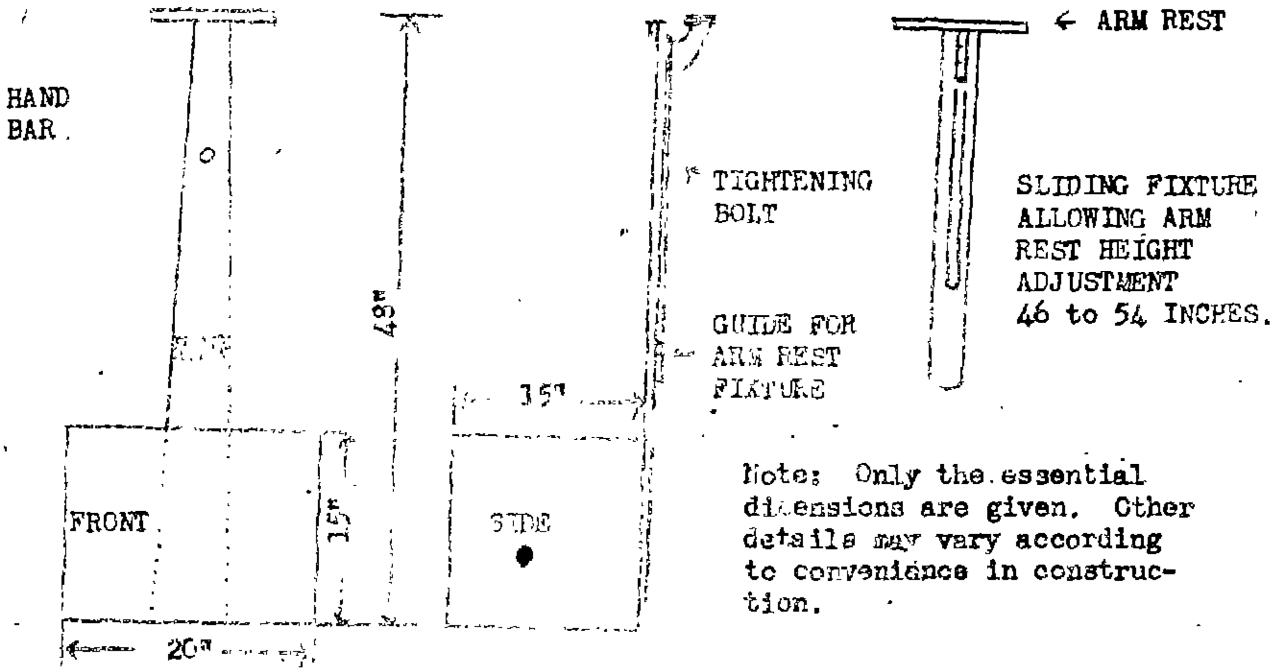
It will be noted that none of the total test scores on the standard tests correlate significantly (at the 5 per cent level) with the treadmill test. Although this does not prove the superiority of the treadmill test as a measure of physical fitness it does indicate nevertheless that the standard tests are not assessing the same cardiovascular-respiratory that are being measured by the treadmill test. These data seem also to justify dropping further experimental work in so far as studying the relationship between these tests and maximal treadmill performance.

APPENDIX C

FIGURE 8. THE EXERCISE PLATFORM
FIGURE 9. THE EXERCISE PATTERN
FIGURE 10. STETHOSCOPE POSITION FOR HEART RATE COUNT

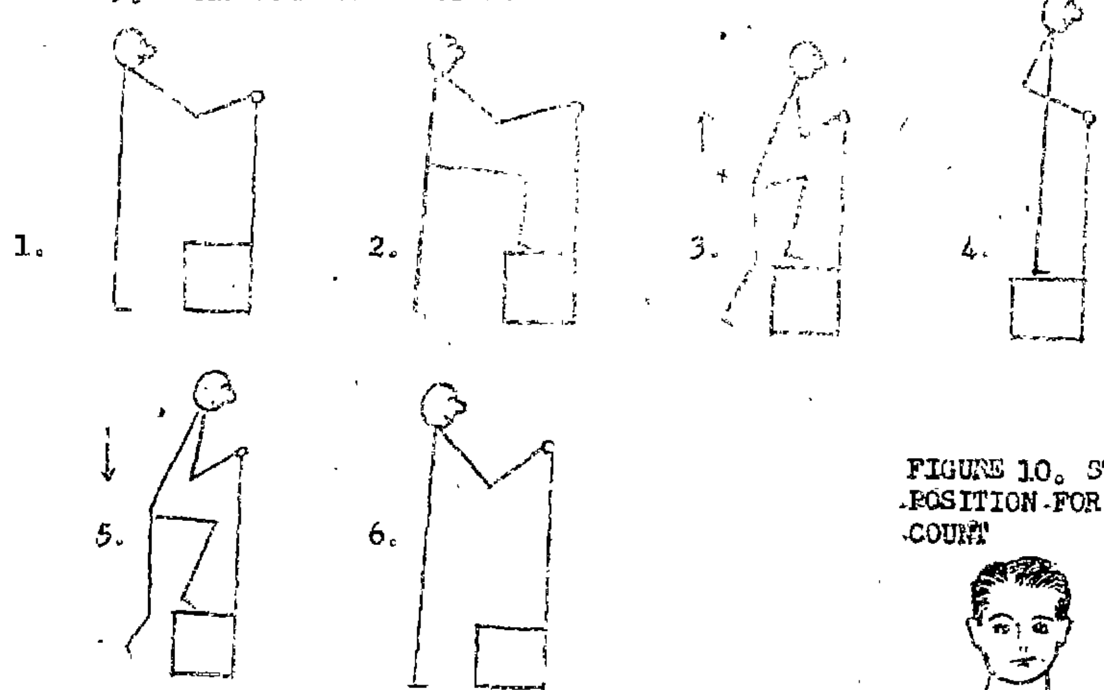
APPENDIX C

FIGURE 8. THE EXERCISE PLATFORM



Note: Only the essential dimensions are given. Other details may vary according to convenience in construction.

FIGURE 9. THE EXERCISE PATTERN



1 and 6: off-the-platform
4: on-the-platform

The first leg in 2 depends on preference of the subject; in 5 the same leg is started down first.

FIGURE 10. STETHOSCOPE POSITION FOR HEART RATE COUNT