2006

Body Weight Penalties in the Physical Fitness Tests of the Army, Air Force, and Navy

Paul M. Vanderburgh
University of Dayton, pvanderburgh1@udayton.edu

Todd A. Crowder
United States Military Academy

Pages: 16
Words: 3802
Tables & Figures: 2
Photos: 0
References: 27
Contact: Dr. Paul Vanderburgh
Email: vanderburgh@udayton.edu
Guarantor: Dr. Paul Vanderburgh

BODY WEIGHT PENALTIES IN THE PHYSICAL FITNESS TESTS OF THE ARMY, AIR FORCE, AND NAVY (ACCEPTED AND FINAL VERSION)

Paul M. Vanderburgh, EdD
Department of Health and Sport Science
University of Dayton
300 College Park
Dayton, OH 45469-1210
(937) 229-4213
vanderburgh@udayton.edu

Todd A. Crowder, PhD
Department of Physical Education
United States Military Academy
West Point, NY 10996
(845) 938-2667

Key words: allometry, fitness testing, body weight bias

Abstract

Recent research has empirically documented a consistent penalty against heavier service members for events identical or similar to those in the physical fitness tests of the Army, Air Force, and Navy. These penalties, not related to body fatness, are based on biological scaling models and have a physiologic basis. Using hypothetical cases, we quantified the penalties for males, 60 vs. 90 kg body weight, and females, 45 vs. 75 kg, to be 15-20% for the fitness tests of these three services. Such penalties alone can adversely impact awards and promotions for heavier service members. To deal equitably with these penalties in a practical manner, we offer two recommendations: 1. Implementation of revised fitness tests with balanced events: penalties of one event against heavier service members are balanced by an equal and opposite bias against lighter service members, or 2. Development of correction factors which can be multiplied by raw scores to yield adjusted scores free of body weight bias.

Introduction

The need for standardized, regular physical fitness tests for the military services is well established. Each of the three primary services (Army, Air Force, and Navy) assesses the primary fitness dimensions of aerobic capacity and muscular strength/endurance as part of its official test of physical fitness. Recent scientific evidence, however, suggests that the tests used to assess these fitness dimensions (e.g., push-ups, sit-ups, abdominal crunches, and distance runs) impose a physiological penalty on heavier service members. Therefore, the objective of this paper is to quantify the magnitude of that penalty on the resulting fitness scores of these tests.

Background

Aerobic capacity, also called cardio-respiratory fitness, is measured via a timed run test. Muscular strength/endurance, for each service, is divided into upper body and trunk. The former is measured by a timed push-up test and the latter by either a timed sit-up, curl-up or abdominal crunch test. Table 1 shows a comparison of the three services' physical fitness tests for otherwise healthy service members (1,2,3). Noteworthy is the fact that some services have other events that measure another component of health-related fitness such as flexibility or body composition. This table only includes the events dealing with muscular strength/endurance and aerobic capacity because these are the only fitness components impacted by the body mass bias emphasis of this investigation.

/ Insert Table 1 about here /

These tests are practical for mass testing because they are time-efficient and require little, if any equipment. They also have passed certain tests of validity for the component measured. The push-ups, pull-ups, abdominal crunches, curl-ups and sit-ups, for example, are widely accepted as valid indicators of muscular strength/endurance according to the principle of face validity. That is, these tests, "on the face of it" are clearly indicators of either muscular strength and/or muscular endurance. The distance runs have been correlated with the gold standard of aerobic capacity: VO_{2max}, the maximum rate at which the body can consume oxygen during exercise (4,5).

One key issue of recent interest in physical fitness testing is the undue influence of body size on outcome measures, that is, the extent to which body size is being measured above and beyond fitness level. This can be easily understood in considering the one-repetition bench press exercise (1RM) , the maximum amount of weight one can lift one time. While this measure is clearly an index of muscular strength, it is also influenced by one's body mass (M); larger individuals generally have more muscle mass and, therefore, an advantage. Said another way, the 1RM bench press test not only measures muscular fitness but body mass as well. A simple correction is to divide the 1RM number by M. While appealing in its simplicity, this ratio adjustment has been criticized for penalizing heavier individuals (6-13). Similarly, use of raw data such as repetitions and distance run time have also been shown to penalize heavier and lighter individuals, respectively, (5,14,15).

Such penalties are based on principles of allometry, the study of the relative growth of a part of an organism relative to the growth of the whole. In exercise science, allometry can be used to understand how certain outcome variables of physical performance are related to one-, two-, or three-dimensional factors such as height, cross sectional area, or weight, respectively (7). Despite the well documented effects of gender and age on physical performance, these principles apply similarly within populations, both theoretically and empirically (4,8,9,11). This means, for example, that despite the well-known body structure differences between men and women, these principles are equally applicable within each gender. An allometric relationship states that an outcome variable, Y, is related to a dimensional variable, X as follows:

$$Y \sim bX^{a} \qquad \text{(Eq 1)}$$

The parameters b and a are constants. In the case of muscular strength as the outcome variable (1RM), and body mass (M) as the dimensional variable, allometric scaling laws dictate that 1RM is directly proportional to muscle cross-sectional area (CSA), and CSA is proportional to $M^{2/3}$. Therefore, the following holds true:

$$1RM \sim M^{2/3} \qquad \text{(Eq 2)}$$

Equation 2 has very important application to strength testing. It dictates that, as M increases, strength increases but not in direct proportionality. Said another way, an exact scale model of an individual but 20% heavier, could not lift 20% more weight but could lift only 13% more. This number is calculated by rewriting a 20% improvement as 1.2 (a 20% improvement is equal to 1.2 times the original value) and raising it to the 2/3 power. Eq. 2 is also important because it yields the optimal body mass adjusted fitness score: $1RM \cdot M^{-2/3}$ (8,10). Such scores can be compared between individuals in a way that eliminates the body mass bias.

The research literature is replete with evidence of muscular strength not being directly proportional to body mass or fat-free mass (FFM). Rather, most studies have reported empirically that $1RM \sim M^x$ where x is significantly less than one and greater than zero. For example, for simple measures of strength (1RM or dynamometer), M exponents of approximately 2/3 and significantly different from 0 or 1 have been reported for young athletes (12), rugby athletes (16), elderly men and women (9), and elite male and female powerlifters (17,18). Finally, Jaric (11), in an extensive review of the scaling of strength literature, concluded that the optimal overall mass exponent is 2/3 for maximal strength.

While muscular strength is measured by some form of 1RM, muscular endurance is most often measured as repetitions (REPS) of a fixed resistance or one's body weight. In the case of the service fitness tests (e.g., push-ups, pull-ups, sit-ups, abdominal crunches or curl-ups), in which multiple repetitions of one's body weight are expected, these are generally considered tests of muscular endurance. While scaling's applicability to these tests is explained elsewhere (15), the predicted body mass exponent yields the optimal expression of $REPS \cdot M^{1/3}$. Because M has been shown to be negatively correlated with REPS, this index yields a correction that increases the score for increased M, thus mitigating the penalty for being larger. Markovic & Jaric (15) as well as Crowder et al. (14) have shown empirically that muscular endurance fitness tests that move one's body weight corroborate the $REPS \cdot M^{1/3}$ as the optimal body weight adjusted outcome score.

Allometric scaling laws also predict the same for the key measure of aerobic capacity, VO_{2max}, for reasons a bit more complicated but well explained elsewhere (19). In fact, the allometric relationship is the same as eq. 2 above except 1RM is replaced by VO_{2max} (expressed as a volume rate):

$$VO_{2max} \sim bM^{2/3} \qquad \text{(Eq 3)}$$

Interpreted similarly, a 20% heavier scale model of an individual would only have a 13% greater VO_{2max}. Empirical support for this body mass exponent approximating 2/3, and not 0 or 1, exists across different populations (19,20,21,22). Some findings suggest, however, that the 2/3 exponent value is not easily reproducible in certain populations (23,24) and may be partly due to the undue influence of fat mass (19,23).

Because VO_{2max} is one of the key determinants of distance run time, then these scaling laws apply to distance run times as well but not with the same exponents. The theoretical exponent for run time has been calculated as 1/3, which means that the optimal expression of a distance run time (T) should be $T \cdot M^{-1/3}$ (5). Notice, that as mass increases, the overall index decreases, which, for a score in which smaller is better, reduces the penalty for larger M. Empirical support for scaling of a distance run test exists as well. For a two-mile run time test of military service academy cadets, Vanderburgh (5) and Crowder (14) reported the optimal body mass exponents to be -0.41 and -0.26, respectively, both different from 0. Given that these were two separate studies, and the midpoint of these exponents was -0.33 (that predicted by theory), one can make the case that some scaling of distance run time is warranted.

Methods

Discussion of a bias or penalty is not complete without quantification. To better understand the magnitude of the penalties discussed above, we began with four hypothetical subjects: two men, 60 and 90 kg each, and two women, 45 and 75 kg each. We set the age at 28 years. Such classification represents the lighter and heavier ends of the weight distribution spectrum in a typical unit as well as an age well represented throughout all military services. For the lighter weights in each category (60 kg for men, and 45 kg for women), we also assumed a maximum performance on each event, based on the relevant official testing regulations of each service (1,2,3) and calculated the resulting equivalent performance for the heavier "subjects." From these, we determined the resulting points and calculated the penalty as a percentage.

An example for each gender illustrates this calculation for the Army physical fitness test. For a 60 kg 28 yr old male, he must perform 77 push-ups to achieve the maximum 100 points. This equates to a scaled score of $77\ (60)^{1/3} = 301.4$. The 90 kg man's equivalent performance (REPS) would be at the same scaled score. Therefore, $\text{REPS}(90)^{1/3} = 301.4$, yields REPS = 67. Similarly, for the 45 kg and 75 kg woman, the equivalent 2 Mile Run times, using the expression $T \cdot M^{-1/3}$, are 15:48 and 18:44, respectively.

Results and Discussion

Table 2 details the quantification of the penalties for each service's physical fitness test, using the same hypothetical subjects. For the body weights shown, which are reasonable estimates of lighter and heavier males and females in a unit, the physiological penalty for heavier individuals is between 14.7% and 20.0% across all three services.

/ Insert Table 2 about here /

The combination of theoretical and empirical evidence yields a compelling case that each of the above physical fitness test events imposes a significant body weight penalty on men and women in the armed services. One cannot overstate the clarification that the penalty is not one of fatness but is a real physiological penalty. Restated from above, two identical scale models of each other, one with M=60 kg who achieves maximum scores and the other with M=90 kg, would compare similarly as shown in Table 2. Stated differently, a 75 kg woman who scores 240 points on the Navy test could be performing at the same physiological level as the 45 kg woman who scores a perfect 300.

The real issue with the current testing system is that some heavier service members may be evaluated unfairly in the highly competitive awards and promotion systems of the different service branches. Because each test (1,2,3) clearly includes an awards-based point system, then these fitness test results are conducive to inclusion in the entire performance portfolio of the individual.

There are two important challenges, however, in implementing body-weight-bias-free fitness tests. The first is that such a system seems counterintuitive. One could easily see that a test requiring a person to carry a 100 lb rucksack from A to B clearly provides advantages for the larger service members. Yet the same argument could be made about the advantages to the smaller in a test that required just transporting oneself from A to B. In the latter, one might think that carrying one's body weight is body mass-independent because larger people have more muscle mass to carry the larger weight. While somewhat true, the aforementioned scaling laws and empirical evidence indicate that ability to carry weight does not increase in direct proportionality to body weight. Therefore, since each of these fitness test events involves body weight as the principal resistance, each is subject to similar body penalties.

The second key challenge of having a fitness test free of body weight bias is scoring. Since the scaling laws involve exponents, then expressing scores such as $1RM \cdot M^{-2/3}$, $REPS \cdot M^{1/3}$ or $T \cdot M^{-1/3}$, creates an impractical, less user-friendly testing scenario which necessitates the use of a calculator and creates scores with strange values and units. One simpler method is the use of correction factors, which are dimensionless numbers that are multiplied by one's raw score (reps, weight, time, etc) to yield an adjusted score devoid of any body weight bias. The International Powerlifting Federation has used such a score to determine the single overall best performer at large meets out of all weight classes.

This index, though based on a polynomial model, has been validated and can be used to adjust any 1RM strength scores (25). A simpler index based on the allometric model for strength has also been published (13) and includes an explanation on how correction factors can be developed for any type of fitness test. For a two-mile run test, Vanderburgh and Laubach (26) developed a time handicap for each gender based on body weight and age such that the handicap would be subtracted from the actual run time to yield an adjusted score that could allow comparison of any runners, regardless of age or body weight.

Perhaps one perceived limitation of any sort of "credit" for being heavier, is the extent to which one is rewarded for excess body fat. A brief analysis, however, of the comparison between the physiological cost of fat weight on distance run time, for example, vs. the handicap gained by that excess weight, shows that excess fat weight does, indeed, pose a net penalty. For example, if a 70 kg woman with a three mile run time of 30 minutes, gained an additional fat weight of 2 kg, then according to Nevill's 5K equation (4), her actual run time would increase to 30:53, yet she would have received only a 17 second credit for the extra weight. In essence, by carrying excess fat weight of 2 kg, she loses a net 35 seconds. This trend persists for any of the above mentioned physical fitness tests, each of which requires one to carry his/her body weight. One can conclude, then, that individuals who do not meet the body weight standards of the military due to excess fat weight, will be penalized appropriately with any such correction factors. Only when the additional weight approaches all lean body mass would the net penalty approach zero.

One other way to deal with the scoring issue is to use events that cancel the body weight penalty out or to use events with no bias. For example, the combined scores of a maximal repetition test with a fixed weight (e.g., all men bench press 175 lb as many times as possible) and a distance run test could cancel out the body mass bias. To our knowledge, no such specific test has been validated but theory clearly suggests its merits. A Backpack Run Test has been proposed for the military that, in theory, imposes no body weight bias and, arguably, has more occupational relevance for the military than an unloaded distance run test (27). Either approach is defensible in that fitness has both an absolute and relative relevancy, both of which can be measured in practical ways that are fair.

Conclusion

In this paper, we applied the relevant theoretical and empirical research evidence to quantify the resultant penalties for heavier service members taking the standardized physical fitness tests of the Army, Navy and Air Force. This penalty, in the 15 – 20% range, suggests that continued use of the current fitness evaluation system will likely result in inaccurate and unfair evaluation for awards and promotions. We propose two solutions: 1. Correction factors which can be multiplied by the performance score or 2. Revised fitness tests which include tests of absolute power and strength along with those that are body weight-supported, thereby negating this penalty.

References

1. U.S. Air Force Instruction 10-248. Fitness Program. January 2004; pp. 55-59.

2. U.S. Army Field Manual 21-20. Physical Fitness Training. October 1998; pp. 14-4 – 14-6.

3. U.S. Navy OPSNAVINST 6110.1G. Physical Readiness Test (PRT). October 2002; p.14.

4. Nevill AM, Ramsbottom R, Williams C: Scaling physiological measurements for individuals of different body size. Eur J Appl Physiol 1992; 65: 110-7.

5. Vanderburgh PM, Mahar MT: Scaling of 2-mile run times by body weight and fat-free weight in college-age men. Journal of Strength and Conditioning Research 1995; 9: 67-70.

6. Aasa U, Jaric S, Barnekow-Bergkvist M, Johansson H. Muscle strength assessment from functional performance tests: role of body size. Journal of Strength and Conditioning Research 2003 ; 17: 664-70.

7. Batterham AM, George KP: Allometric modeling does not determine a dimensionless power function ratio for maximal muscular function. J Appl Physiol 1997; 83: 2158-66.

8. Vanderburgh PM, Mahar MT, Chou CH: Allometric scaling of grip strength by body mass in college-age men and women. Res Q Exercise Sport 1995; 66: 80-4.

9. Davies MJ, Dalsky GP. Normalizing strength for body size differences in older adults. Med Sci Sports Exercise 1997; 29: 713-7.

10. Jaric S, Ugarkovic D, Kukolj M. Evaluation of methods of normalizing muscle strength in elite and young athletes. Journal of Sports Medicine and Physical Fitness 2002; 42: 141-151.

11. Jaric S. Muscle strength testing – Use of normalization for body size. Sports Med 2002; 32: 615-31.

12. Jaric S, Radosavljevic-Jaric S, Johannson H. Muscle force and muscle torque in humans require different methods when adjusting for differences in body size. Eur J Appl Physiol 2002; 87: 304-7.

13. Vanderburgh PM. A simple index to adjust maximal strength measures by body mass. Journal of Exercise Physiology^{online} 1999; 2: 7-12.

14. Crowder T,. Yunker C: Scaling of push-up, sit-up and two-mile run performances by body weight and fat-free weight in young, fit men. [Abstract.]. Med Sci Sports Exercise 1996; 28: S183.

15. Markovic G, Jaric S. Movement performance and body size: The relationship for different groups of tests. Eur J Appl Physiol 2004; 92: 139-49.

16. Atkins SJ. Normalizing expressions of strength in elite rugby league players. Journal of Strength and Conditioning Research 2004; 18: 53-58.

17. Dooman CS, Vanderburgh PM. Allometric modeling of the bench press and squate: Who is the strongest regardless of body mass? Journal of Strength and Conditioning Research 2000; 14: 32-36.

18. Vanderburgh PM, Dooman C: Considering body mass differences, who are the world's strongest women? Med Sci Sports Exercise 2000; 32: 197-201.

19. Vanderburgh PM, Katch FI: Ratio scaling of VO_{2max} penalizes women with larger percent body fat, not lean body mass. Med Sci Sport Exercise 1996; 28: 1204-8.

20. Loftin M, Sothern M, Trosclair L, O'Hanlon A, Miller J, UIdall J. Scaling VO2peak in obese and non-obese girls. Obesity Research 2001; 9: 2906.

21. Heil DP. Body mass scaling of peak oxygen uptake in 20- to 79-yr-old adults. Med Sci Sports Exerc 1997; 29: 1602-8.

22. Buresh R, Berg K. Scaling oxygen uptake to body size and several practical applications. Journal of Strength and Conditioning Research 2002; 16: 461-5.

23. Batterham AM, Vanderburgh PM, Mahar MT, Jackson AS. Modeling the influence of body size on VO_{2peak}: Effects of model choice and body composition. J Appl Physiol 1999; 87: 1317-1325.

24. Janz KF, Burns Tl, Witt JD, Mahoney LT. Longitudinal analysis of scaling VO_2 for differences in body size: The Muscatine Study. Med Sci Sports Exercise 1998; 1436-44.

25. Vanderburgh PM, Batterham AM. Validation of the Wilks powerlifting formula. Med Sci Sports Exerc 1999; 31: 1869-1875.

26. Vanderburgh PM, Laubach LL. Modeling a two-mile run age and body weight handicap for men and women. Biomedical Sciences Instrumentation 2000; 36: 325-330.

27. Vanderburgh PM, Flanagan S. The Backpack Run Test: A model for a fair and occupationally relevant military fitness test. Mil Med 2000; 165: 418-21.

Table 1

	Upper Body Muscular Strength/Endurance	Trunk Muscular Strength/Endurance	Aerobic Capacity
Army	2 min Push-ups	2 min Sit-ups	2 Mile Run
Air Force	1 min Push-ups	1 min Sit-ups	1.5 Mile Run
Navy	2 min Push-ups	2 min Curl-ups	1.5 Mile Run

Table 1. Muscle Strength/Endurance and Aerobic Capacity Tests of the Three Primary Armed Services (1,2.3)

Table 2

Event	Male		Female	
	60 kg	90 kg	45 kg	75 kg
	Maximum performance (equivalent points)	Equivalent performance[1] (equivalent points)	Maximum performance (equivalent points)	Equivalent performance[1] (equivalent points)

ARMY

Push-ups 2 min	77 (100)	67 (89)	50 (100)	42 (90)
Sit-ups 2 min	82 (100)	72 (89)	82 (100)	69 (86)
2 mile run time	13:18 (100)	15:14 (78)	15:48 (100)	18:44 (74)
Total points (%diff.)[2]	**300**	**256 (14.7%)**	**300**	**250 (16.7%)**

AIR FORCE

Push-ups 1 min	57 (10)	50 (8.75)	41 (10)	35 (9.0)
Ab Crunches 1 min	53 (10)	46 (8.5)	47 (10)	40 (8.5)
1.5 mi run time	9:36 (50)	11:00 (42)	11:24 (50)	13:31 (40.5)
Total points (%diff.)[2]	**70**	**59.25 (15.4%)**	**70**	**58 (17.1%)**

NAVY

Push-ups 2 min	84 (100)	73 (85)	46 (100)	39 (80)
Curl-ups 2 min	101 (100)	88 (80)	101 (100)	85 (75)
1.5 mi run time	8:55 (100)	10:12 (85)	10:17 (100)	12:11 (85)
Total points (%diff.)[2]	**300**	**250 (16.7%)**	**300**	**240 (20%)**

[1]Equivalent performance is the raw score for the heavier individual to be physiologically equivalent to that of the lighter individual. For example, the 60 kg Army male's 77 push-ups are equivalent to 67 push-ups for the 90 kg male. Similarly, equivalent points are the corresponding points from the scoring tables (1,2,3).

[2]This percentage difference is the lighter vs. the heavier individual's total points. For example, the 60 kg Army's male's performance is equivalent to 256 points for the 90 kg male, a 14.7% difference.

Table 1. Calculation of equivalent points and the resulting penalties based on body weight differences of 60 to 90 kg for men, 45 to 75 kg for women, and an age of 28 yr.