8-2008

Occupational Relevance and Body Mass Bias in Military Physical Fitness Tests

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

Occupational Relevance and Body Mass Bias in Military Physical Fitness Tests

(Accepted for Publication, Medicine and Science in Sports and Exercise)

Published 2008

Paul M. Vanderburgh, EdD, FACSM

Department of Health and Sport Science

300 College Park

University of Dayton

Dayton, OH, 45469-1210

TEL: 937.229.4213

FAX: 937.229.4244

vanderburgh@udayton.edu

Running Title: Bias and Relevance in Military Fitness Tests

ABSTRACT

Recent evidence makes a compelling case that U.S. Army, Navy, and Air Force health-related physical fitness tests penalize larger, not just fatter, service members. As a result, they tend to receive lower scores than their lighter counterparts, the magnitude of which can be explained by biological scaling laws. Larger personnel, on the other hand, tend to be better performers of work-related fitness tasks such as load carriage, heavy lifting and materiel handling. This has been explained by empirical evidence that lean body mass and lean body mass to dead mass ratio (dead mass = fat mass and external load to be carried/lifted) are more potent determinants of performance of these military tasks than the fitness test events such as push-ups, sit-ups or two distance run time. Since promotions are based, in part on fitness test performance, lighter personnel have an advancement advantage, even though they tend to be poorer performers on many tests of work-related fitness. Several strategies have been proposed to rectify this incongruence including balanced tests, scaled scores, and correction factors - yet most need large scale validation. Because nearly all subjects in such research have been men, future investigations should focus on women as well as elucidate the feasibility of universal physical fitness tests for all that include measures of health- and work -related fitness while imposing no systematic body mass bias.

Key Words: allometric scaling; work physiology; body size; load carriage

INTRODUCTION

Paragraph 1. The primary military services of the U.S. Army, Navy and Air Force, require regular physical fitness tests (PFT) of all active duty and reserve service members. Though not identical, each of the services' PFTs includes events of upper body and trunk muscle strength/endurance as well as overall cardio-respiratory endurance in the form of a distance run. Specific test formats by service are shown in Table 1 and the minor differences between events (e.g., the sit-ups vs. curl-ups) can be found in the official service regulations regarding physical fitness tests (26-28). The Marine Corps PFT was not listed since it includes a pull-up test for men and a flexed-arm-hang for women, two events not well-studied with regard to the present topic. Widely considered to be measures of health-related fitness (22) the events of these tests also are conducive to mass testing and require little to no equipment, a key feature for a military PFT that often involves the testing of hundreds of participants at one time. Annual testing is mandatory for every service member and PFT test scores are one of a number of determinants of promotion. The PFT, then, for each service member, is a high-stakes test with important consequences. Noteworthy, however, is the fact that although all three services also employ additional and distinct evaluations of body composition, each uses different assessment methods, evaluation standards, and administrative procedures. Therefore, this review focuses only on the body mass bias and occupational relevance of the performance-related fitness events shown in Table 1.

BODY MASS BIAS

Paragraph 2. Research evidence suggests that the events of each of these tests impose a body mass penalty against larger, not just fatter, service members. Crowder & Yunker (7) used allometric scaling to determine that, in a sample of 238 fit and lean service academy male cadets, the combined score representing push-ups, sit-ups and two-mile run performance in the Army PFT (Table 1) imposed a systematic bias against larger cadets. The magnitude of this bias persisted in separate analyses of each event. In 59 male cadets from the same population, though a different sample, Vanderburgh & Mahar (30) reported 0.49 and 0.32 ($p < 0.05$) correlations between two-mile run time vs. body mass (M) and fat-free mass, respectively. Markovic & Jaric (21), assessed the influence of body size on 18 common tests of movement performance, including the one minute push-ups and sit-ups tests, with 77 male physical education students (ages 18-26). Their findings corroborated not only the existence but the magnitude of the body mass bias reported in the other studies (7, 30). For example, they determined that the push-ups and sit-ups scores exhibited a significant and negative correlation with body mass and that multiplying these scores by $M^{1/3}$ produced an expression that exhibited zero correlation with body mass, thereby eliminating bias.

Table 1. Muscle Strength/Endurance and Aerobic Capacity Tests of the Three Primary Armed Services (22-24)

	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

Paragraph 3. Such empirical evidence of body mass bias has important theoretical bases, beginning with laws of biological proportionality and scaling. The two basic relationships are those between maximal strength (S), maximal oxygen uptake (VO_{2peak}, in $ml \cdot min^{-1}$) and body

mass. Astrand & Rodahl (1) concluded that since muscle strength and VO_{2peak} are directly proportional to muscle and blood vessel cross sectional area, respectively, then strength and VO_{2peak} must be proportional to $M^{2/3}$. The implications of this suggest that commonly used expressions such as strength as $S \cdot M^{-1}$ or VO_{2peak} as ml $O_2 \cdot min^{-1} \cdot M^{-1}$ make too much of an adjustment for body mass and, therefore, penalize heavier individuals (10, 24, 30, 34). Said differently, the correlations between these ratio expressions (i.e. dividing by body mass to the first power) and body mass is statistically significant, and in the direction of being advantageous toward lighter personnel. More importantly, with such expressions, comparisons of VO_{2peak} and/or strength between individuals of different body mass are unduly influenced by body mass and can lead to inaccurate conclusions regarding physical performance (12, 29, 33).

Paragraph 4. These foundational relationships suggest, then, that the more proper expressions of VO_{2peak} and strength adjusted by body mass would be ml $O_2 \cdot min^{-1} \cdot M^{-2/3}$ (1, 12, 24) and $S \cdot M^{-2/3}$ (9, 13-16, 21). In more general terms, for similarly proper adjustment of the influences of body mass, any outcome physical performance variable, Y (e.g., push-ups repetitions, sit-ups repetitions, distance run time, etc.) can be expressed as $Y \cdot M^{-a}$. Numerous investigations have examined the fit between theoretically- and empirically-derived body mass exponents for not only strength and VO_{2peak} but many other performance variables as well. While the details of determining such exponents are described in detail elsewhere (2, 9, 24, 29, 32), ascertainment of fit is based on the theoretical exponent being with the 95% confidence interval of the empirically determined exponent. For example, the empirically determined body mass exponent for the total lift score (the sum of maximal bench press, squat and deadlift performances) among elite women powerlifters was determined to be 0.750 $\pm$ the SEE of 0.052 (34). While this value was not the expected 2/3 exponent, its 95% confidence interval ($0.750 \pm 1.96 \cdot SEE$) was 0.648 - 0.852 and, thus, contained the 2/3 value.

Paragraph 5. Elite powerlifters have often been chosen as subjects for such research because all are highly trained and, regardless of body mass, tend to be very lean, thereby reducing the extent to which body fat and training level may confound results. Furthermore, the powerlifting events are tests of one's one-repetition-maximum, the maximum weight that can be lifted one time, arguably a better indicator of strength than Olympic style weightlifting events which are likely more influenced by power and technique (34). For measures of maximal strength, the 2/3 body mass exponent has empirical support for young men and women (14, 15, 21), and elite male and female powerlifters (33, 34), but not in all cases. While 2/3 was within the 95% CI for the bench press, squat, and total lift for men and all events for women among elite powerlifters, the exponent for the men's deadlift was 0.480 ± 0.050, with the 2/3 exponent not within the 95% CI (33). The authors posited that the lower deadlift exponent may have been due to the influence of grip strength in that event and the finding that the grip strength exponent among adult men and women was 0.51 (29). This finding for the men's deadlift was replicated elsewhere (9). In a small sample of elite female world record holders in powerlifting, the bench press body mass exponent was 0.867 ± 0.053, within which the 2/3 exponent was also not found (34). This may have been due to the fact that only the current world record lifts (N = 9, excluding the heavyweight division, which had no upper weight limit), were considered in the allometric modeling. As such, the exponent, which also happens to be the slope of the best-fit curve, can be changed considerably based on one particularly superlative performance. These examples are illustrative of the variability of empirically derived exponents due to population specifications, sample size, training, and body composition. Nonetheless, the body mass exponent of 2/3 for strength measures has generally been well supported empirically (13).

Paragraph 6. These body mass exponent values should not be confused with those obtained via isokinetic dynamometry, in which maximal torque (N·m) is measured, not force. Torque, the product of a force (proportional to body mass to the 2/3 power) and a length (proportional to body to the 1/3 power) should theoretically be proportional to body mass raised to the first power (13). Indeed, investigations have determined the body mass exponents for torque to be no different from 1.0 for men (15) and elderly men and women, corrected for body fat (8).

Paragraph 7. The push-ups, abdominal crunches and sit-ups events of the military PFT's are not, however, measures of absolute muscular strength. They are timed events measuring maximal number of repetitions with the resistance force being a fraction of one's body mass. Accordingly, Jaric et al. (16) proposed that since the force needed to perform these exercises was directly proportional to body mass raised to the 2/3 power, and indirectly proportional to body mass, then test performance should be proportional to body mass raised to the 1/0.67 or -1/3 power. Empirical evidence supports this notion. Crowder and Yunker (7), in the aforementioned sample of 238 fit, young, male military academy cadets, determined that -1/3 was within the empirically derived body mass exponent's 95% CI for push-ups (-0.18 - -0.58) and nearly for sit-ups (-0.12 - -0.32) performance. For 77 male physical education students, Markovic & Jaric (21) concluded that push-ups and sit-ups performance should be normalized using the body mass exponent of -1/3. This means that, since body mass is negatively correlated with push-ups and sit-ups performance, the maximal number of repetitions should be multiplied by body mass to the 1/3 power before comparisons between individuals are made since dividing by $M^{-1/3}$ is the same as multiplying by $M^{1/3}$. No published data exist, however, on empirically derived body mass exponents for women in the push-ups and sit-ups test.

Paragraph 8. For measures of VO_{2peak} the body of empirical evidence is somewhat supportive of the 2/3 body mass exponent for adult men and women. Nevill et al. (24) reported an exponent of 0.67 for 204 recreationally active men and women. Heil (12), controlling for the effects of gender, age, percent body fat, height and self-reported physical activity among 440 men and women, determined the exponent to be 0.65 (0.530 – 0.776) and 0.76 (0.651 – 0.862) with and without height in the model, respectively. Other findings support the 2/3 body mass exponent but not when fat-free mass was considered. Batterham et al. (3), in a sample of 1314 men, calculated a 2/3 body mass exponent but a fat-free mass exponent not different from 1.0, when the effects of age and self-reported physical activity levels were controlled for. Similarly, for 98 women, Vanderburgh & Katch reported the same trend when scaling VO_{2peak} by body mass and fat-free mass (31), but without control for other variables.

Paragraph 9. Nonetheless, others have used this 2/3 exponent for VO_{2peak} to explain how the body mass exponent for distance run time should be 1/3 (30, 36, 37). Because distance run time has been shown to be indirectly proportional to peak oxygen update (VO_{2peak}), expressed per unit of M, or ml $O_2 \cdot M^{-1} \cdot min^{-1}$ (24) and VO_{2peak} has been shown to be proportional to $M^{2/3}$, then distance run time should be proportional to $M^{2/3} \cdot M^{-1}$, or $M^{-1/3}$. Since low score wins in run time (T), the correct scaling should then be $T \cdot M^{-1/3}$. Empirical evidence supports this derivation for adult men (6, 7, 26) and young adult men and women (24). In this latter investigation, the body mass exponent determination was not an objective but was instead derived by the present author based on available data presented. Providing credit for body mass may appear inappropriate if the fat mass constitutes a large percentage of the body mass. Recent evidence, however, makes a compelling case that body fat actually penalizes the $T \cdot M^{-1/3}$ values because the increase in run time due to fat is significantly larger than the handicap gained by the excess weight (6, 37).

Paragraph 10. Of key importance is the lack of published data on empirically derived exponents for women especially in the push-ups, sit-ups and distance run events. This may have been due, in part, to the relatively small percentage of women available in military units where much of such data collection has occurred. Others have expressed difficulty in seeking women subject volunteers at road races where body mass was to be measured (6). Nevertheless, given the similarity of body mass exponents for powerlifting events of strength between men and women (33) as well as those of VO_{2peak} (12, 24), one could readily hypothesize that body mass exponents for other fitness tests should be similar between men and women.

Paragraph 11. The impact of such body mass bias in the military physical fitness tests has been quantified. Vanderburgh & Crowder (36) calculated the difference in test scores between lighter and heavier men (60 vs. 90 kg) and women (45 vs. 75 kg) associated with physiologically equivalent performances. "Physiologically equivalent" was defined, for example, as the expected value of push-ups, sit-ups, or distance run score for a 90 kg man who was an exact scale model of himself but as a 60 kg man. Analyses indicated that the heavier service members' scores were 15 – 20% lower than their lighter counterparts and that this difference could be explained by body mass and not body fat differences. Because physical fitness test scores are an important element in the consideration of promotion, this body mass bias may be large enough to impose an unfair promotion disadvantage against larger men and women. Table 2 summarizes the body mass bias and exponents for common fitness tests of aerobic power, muscle strength, and muscle endurance and includes the resulting scaling expression that allows comparison of individuals or groups in a way that essentially eliminates the bias.

Paragraph 12. The consistent trend for body mass bias of the fitness tests events shown in Table 1 does not mean, however, that performance improvements are evidenced only with weight loss. In fact, Kraemer et al. (18) demonstrated that, in untrained women, a six-month resistance training protocol exercising all major upper and lower body muscle groups in power-type movements led to significant improvements in push-ups, sit-ups, and two-mile-run scores with a concomitant increase in body mass, explained at least partially by modest gains in lean body mass. In another investigation, Kraemer et al. (19) reported that total body resistance training plus endurance run training improved push-ups, two-mile run time, and loaded two-mile run time (carrying the standard load of soldier in the field: a 44.7 kg backpack while wearing boots and battle dress uniform) with no change in body mass. Such a training effect does not violate the laws of biological similarity because the trained individual is no longer a scale model of him or herself from the untrained or pre-trained state.

Table 2. Empirical evidence for body mass (M) exponents for common fitness measures.

	Theoretical	Actual (Ref)	95% CI	Advantage	Scaling Index
Push-ups (REPS)	1/3	0.42^{+} (21) 0.38^{+} (7)	NR 0.18 - 0.58	Lighter	$REPS \cdot M^{1/3}$
Sit-ups (REPS)	1/3	0.32^{+} (21) 0.22^{+} (7)	NR 0.12 – 0.32*	Lighter	$REPS \cdot M^{1/3}$
Two-mile run (T)	1/3	0.26^{+} (7) 0.40^{+} (30) 0.356^{+++} (24)	0.14 - 0.38 0.23 - 0.57 NR	Lighter	$T \cdot M^{-1/3}$
5K run (T)	1/3	0.410^{+} (6)	0.199 - 0.622	Lighter	$T \cdot M^{-1/3}$
Bench press (1RM)	2/3	0.69^{+} (21) 0.671^{+} (33) 0.756^{++} (33)	NR 0.585 - 0.757 0.646 - 0.866	Heavier	$1RM \cdot M^{-2/3}$

1. The actual body mass exponent is the empirically-derived version of the M exponent in the Scaling Index (last column in the table) such that, in a sample, the correlation between the Scaling Index with the actual exponent and M is not different from zero. This would indicate no body mass bias. If the 95% CI of the actual exponent contains the theoretical exponent, then the Scaling Index would be appropriate for performance comparisons between individuals of the same gender but different body mass. In all cases, empirical support exists for the use of the Scaling Index with the exception of one study for (Ref 7: 1/3 is just outside the 95% CI for men's sit-ups).
2. $^{+}$Male only, $^{++}$Female only, $^{+++}$Both,
3. *Theoretical exponent not within 95% CI of actual
4. "Advantage" refers to which personnel, by body mass, receive an advantage in the raw fitness test score
5. NR = not reported
6. Ref 6 (N = 99 M): recreational 5K race competitors
7. Ref 7 (N= 238 M): fit, lean service academy cadets
8. Ref 21 (N = 77 M): college-age physical education students
9. Ref 24 (N = 112 M, 92 F): recreationally active young adults
10. Ref 30 (N = 59 M): fit, lean service academy cadets
11. Ref 33 (N = 30 M, 27 F for each event): world class powerlifting competitors

Paragraph 13. An interesting characteristic of these military physical fitness test events is that the primary resistance is body weight and little else. Typical physically demanding tasks in many military specialties, however, require individuals to move not only themselves but equipment, supplies, and/or weapons as well, requiring more absolute strength and power, often correlated with larger lean body mass (10). This suggests that performance of such military tasks may correlate only moderately with physical fitness test scores; and may be more strongly correlated with body mass such that larger service members are better performers. The empirical evidence supports these hypotheses.

Paragraph 14. In 93 Royal Navy (U.K.) personnel (52 male and 41 female) Bilzon et al. (5) examined the extent to which anthropometric and fitness variables explained variance in performance of simulated free carry and stretcher carry tests. While the optimal regression equation for the free carry contained the predictors of standing broad jump, lean body mass, dead mass (total weight lifted plus fat mass), 20m sprint time, push-ups, sit-ups and grip strength ($R = 0.89$), the lean body mass to dead mass ratio (LBM/DM) alone yielded correlations of 0.87 and 0.85 for the free carry and stretcher carry, respectively. Interestingly, this index, LBM/DM favors larger, leaner personnel, given that the external weight to be carried (i.e., the casualty) is independent of one's own weight.

Paragraph 15. This importance of LBM/DM as a determinant of load carriage was examined by Lyons et al (20). In 28 male volunteers, during heavy (40 kg) load carriage, LBM/DM and absolute VO_{2max} ($ml \cdot min^{-1}$) were the strongest single predictors of $\%VO_{2max}$, a useful indicator of the metabolic demand of load carriage. In fact, as load increased from light to heavy the correlation between absolute VO_{2max} and $\%VO_{2max}$ increased with a concomitant decrease in the correlation between relative VO_{2max} ($ml \cdot kg \cdot min^{-1}$) and $\%VO_{2max}$. Given the widely accepted use of distance run tests as surrogate measures of relative VO_{2max} (30), authors concluded that "application of these measurements would ensure selection criteria for load-carriage occupations are based on lean muscle mass rather than running speed." In a similar load carriage study, Bilzon et al. (4) determined that the correlation between loaded (18 kg load) treadmill running time to exhaustion and lean body mass was 0.71. Furthermore, in a steady state run with similar load at 9.5 $km \cdot h^{-1}$, there was no relationship between VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$) and the exercise tolerance time. These findings suggest that the distance run test, a surrogate measure of VO_{2max} in ml $O_2 \cdot kg^{-1} \cdot min^{-1}$, exhibits at best a moderate relationship with a typical military load carriage task and, according to the authors, "… incurs a systematic bias against heavier personnel," the very personnel who are better performers on load carriage tasks.

Paragraph 16. In a comprehensive review of the relationship between body size and composition to performance of certain military tasks, Harman and Frykman (10) concluded that load carriage, lifting, pushing, and exerting torque are closely related to lean body mass and that push-up, sit-up and 2-mile run scores are not potent determinants of physically demanding military task performance. Indeed, in their discussion of likely explanations for these conclusions, the authors pointed to the both aforementioned scaling laws (1) as well as the well documented advantages of being smaller and lighter for the push-up, sit-up and distance run tests of the military (23). Harman et al. (11) recently examined the ability to predict performance of simulated battlefield activities via simple field tests in 32 male U.S. Army soldiers. Results indicated that not only did the Army's PFT events of push-ups, sit-ups and two-mile-run scores demonstrate significant trends for poorer performance among larger men but the field expedient tests of vertical jump and horizontal jump did so as well. Furthermore, the simulated battlefield activities that were predicted reasonably well ($r = 0.77$ to 0.82, $p < 0.05$) were, with the exception of a casualty carry, events that required manipulation of their own weight with a light load (18 kg). The authors recognized that, "On the battlefield, there are activities other than casualty rescue that also involve the manipulation of relatively heavy loads, e.g., setting up field artillery, hauling heavy weapons and ammunition, and moving obstacles. These are activities at which larger soldiers, who may not excel at physical fitness tests, could also be at an advantage."

Paragraph 17. In a comprehensive large-scale study with 379 trained soldiers (304 men and 75 women), Rayson et al. (25) examined the relationships between physical performance, anthropometric tests, and criterion military tasks. The criterion tasks (score bases in parentheses) included a staged single lift of an ammunition box (maximum successful lift), a carry of one 20kg water can in each hand (time to failure to maintain a certain pace), a repetitive lift and carry of an ammunition box (time to failure to maintain a certain pace), and a loaded march (time to complete 12.8 km). The physical performance tests included pull-ups, push-ups, sit-ups, hand grip strength, lift power, dynamic muscular endurance (time of failure to maintain a lifting cadence at an absolute weight), aerobic capacity (time to failure of a paced shuttle run) and static muscular endurance (time of failure to maintain a static hold in position). The resulting multiple regression models indicated that, by far, performance measures of absolute strength, endurance, and power were more predictive of criterion task performance than were relative measures (those in which the primary resistance was body mass, e.g. push-ups, sit-ups). Furthermore, fat-free mass, the single most potent anthropometric predictor, was positively correlated with performance of each test. Finally, push-ups, sit-ups and estimated aerobic capacity (in this case a surrogate for distance run time), were moderate to poor predictors of criterion performance.

Table 3. Pearson product moment correlation coefficients between physically demanding military task performance, common fitness test scores and indices of body mass (references in parentheses)

	Stretcher Carry	**Free Carry**	**Load Carriage ($\%VO_{2peak}$)**	**Maximal Lift (1RM)**	**Body Mass**
Body Mass	0.42 (5)	0.40 (5)			
LBM	0.76 (5)	0.76 (5)	-0.62 M (20)	0.86 (25)	
LBM/DM	0.85 (5)	0.87 (5)	-0.60 M (20)		
Push-ups (timed max reps)	0.70 (5)	0.69 (5)		0.24 M (23) 0.32 F (23)	-0.20 M (21)*
Sit-ups (timed max reps)	0.58 (5)	0.56 (5)		0.06 M (23) 0.24 F (23)	-0.27 M (21)
Two-mile run time				0.06 M (23) 0.14 (23)	0.49 M (30)
Absolute VO_{2peak} ($ml \cdot min^{-1}$)			-0.76 M (20)	0.83 (25)	

1. All correlation coefficients for pooled samples of males (M) and females (F) and $p < 0.05$ unless otherwise noted
2. Ref 5 (N = 52 M, 37 F):
   a. Stretcher Carry = avg velocity to complete a prescribed route of a simulated individual's portion (41 kg) of a stretcher carry
   b. Free Carry = avg velocity to complete a prescribed route of a simulated individual's portion (37 kg) of the free carry of a casualty
3. Ref 20 (N = 28 M): Load Carriage = $\%VO_{2max}$ at 40 kg load, 1.11 $m \cdot s^{-1}$ walking speed, 0% grade
4. Ref 21: (N = 77 M)
5. Ref 23 (N = 751 M, 450 F)
6. Ref 25 (N = 181 M, 53 F): Maximum Lift = max weight lifted to a height of 1.45m, not to exceed 72 kg, using a progressive protocol (5kg added to each successful lift)
7. Ref 30 (N = 59 M)
8. *p = 0.08

Paragraph 18. As summarized in Table 3, evidence suggests that performance of physically demanding military tasks is well-correlated with absolute measures of physical performance and lean body mass and moderately correlated with performance tests such as those used in the U.S. military physical fitness tests. In other words, while the ability to moves one's weight either in a muscular endurance or aerobic power event contributes to some success in certain physically demanding military tasks, the ability to exhibit absolute amounts of muscular strength and endurance (i.e. repetitions of fixed external weights) and aerobic power (i.e. absolute VO_{2peak}), are even stronger determinants of military occupational fitness. Additionally, the evidence consistently indicates that performance of occupationally relevant military tasks favors larger personnel yet the physical fitness test events favor the smaller. Therefore, this body mass bias tends to reward the better performers on the high stakes physical fitness tests of health-related fitness and penalize the better performers of occupationally relevant physically demanding tasks.

STRATEGIES AND REMEDIES

Paragraph 19. The apparent incongruence between physical fitness test and occupational task performance has been addressed via potential remedies in the literature. These include: balanced tests, scaled scores, and correction factors though the intent of each is generally to remove body mass bias, not use tests that are advantageous to heavier personnel. This is because zero body mass bias is clearly between that of the bias against heavier personnel in the health-related physical fitness tests and the bias against lighter personnel of the occupationally relevant tests. Given the defensible notion that health-related fitness and occupational fitness are both desirable, a zero body mass bias test appears to be a reasonable remedy.

Paragraph 20. Two versions of the balanced fitness test, the first proposed remedy, have been offered. The first (36) is a test with multiple events such that one event advantageous to lighter personnel is balanced by another event advantageous to heavier personnel. Said differently, the health-related fitness event is balanced by the occupationally relevant test. While strikingly simple in purpose, such a test has neither been validated nor used by any of the military services as a mandatory fitness test. This may be due to the fact that occupationally relevant tests require equipment for each individual test, and are, therefore, not conducive to mass testing. Nonetheless, for example, a maximal one repetition maximum bench press could be accompanied by a distance run test. A person performing well in both must have a relatively large lean body mass, helpful in some key military tasks, and well-developed aerobic capacity, characterized by his/her ability to move body weight over a long distance in a short time period. This individual, then, from a health-related and occupationally relevant fitness perspective, would be a very valuable asset.

Paragraph 21. A backpack run test has been modeled (35) as the second type of balanced fitness test, one comprised of a single event in which the primary resistance includes one's body mass and an absolute amount of additional mass that is constant between individuals. In this case, the event is a timed, distance run test with a backpack that mimics the load soldiers would be expected to carry in training or wartime situations. The model, based on actual distance run time data from 59 lean, fit service academy male cadets, was developed using metabolic equations to estimate the run speed of carrying additional loads. As load increased from zero to 40 kg, the body mass bias went from positive (against heavier personnel, as in a typical distance run) to zero. At 20 kg, the body mass bias was not significantly different from zero. Based on modeling of actual distance run times, these results make a compelling case that, at some level of load, the body mass bias would be zero. While this backpack run test demonstrates apparent face validity by closely simulating a physical performance skill that has occupational and health-related fitness relevance, it has neither been field tested nor validated with large samples. Furthermore, though it does require equipment that each service member would be expected to have, the injury risk of training for such a load carriage test may increase to unacceptable levels (17).

Paragraph 22. Not all attempts to create balanced fitness tests are successful. A popular fitness event in the U.S. pairs 5K distance running with a bench press exercise. Each competitor not only completes the run as fast as possible but also, prior to the run, executes as many repetitions of a bench press as possible (39). For each repetition, 30 sec is subtracted from the race time to yield an adjusted run time. Because the bench press weight is a percentage of one's own body weight adjusted by age, the maximal repetition test becomes essentially similar to the push-up test, with its aforementioned body mass biases. Vanderburgh & Laubach (39) empirically examined this possibility with 312 competitors (258 M, 54 F) in such an event. Indeed, the correlations (r^2) between adjusted run times and body mass were 0.28 and 0.35 ($p<0.01$ for both) for men and women, respectively, thus indicating substantial body mass bias. Using correction factors, based on body mass, the authors reduced this bias to zero.

Paragraph 23. Due to the logistical advantages of no equipment needed in the current physical fitness tests shown in Table 1, another remedy has been proposed that simply removes the body mass bias of the Table 1 test scores (13, 21, 29). This "scaling" solution entails dividing the raw score by body mass raised to a certain exponent and are those previously discussed and shown in

Table 2. Those achieving the best scaled scores in a unit could be considered the most fit overall for health and occupational purposes. This is based on the shifting of the disadvantageous body mass bias away from heavier personnel not to lighter but to a point of zero bias, the midpoint. There are limitations to using such scaled values. First, they create a strange currency of values. For example, the proper scaled score for a push-ups score of 45 repetitions in one minute would be 178.4 $\text{reps}\cdot\text{kg}_{\text{body mass}}^{1/3}$ for a 65 kg woman. Interpretation of this value is complicated by the scarcity of norms using these units. Second, because of the exponent, the calculation is problematic without a calculator. Third, using scaled values calculated from different body mass exponents can lead to erroneous results. For example, based on validation studies with female world class powerlifters (34), one may be tempted to add the bench press, squat and deadlift scaled scores using the exponents of 0.87, 0.72, and 0.63, respectively. Different exponents, however, yield different units and, therefore, such scaled values cannot be added (32).

Paragraph 24. The correction factor remedy is the means by which scores can retain the same units as the original raw data, thereby facilitating more meaningful interpretation. Discussed in detail elsewhere for measures of strength (33) and for the common military fitness test events (38), correction factors are dimensionless numbers that are multiplied by a raw score to compute an adjusted score. For example, a woman, 79 kg body mass and 24 yrs of age, executes 34 push-ups in two minutes. Normally, this would yield a score of 83 points based the Army's standards (27). The correction factor, based on what she would have scored had she been an exact model of herself but at a lighter "reference body mass" of 56.7 kg (details explained in ref. 38), would be 1.12. Her 34 push-ups multiplied by 1.12 yields 38.1 or 38 push-ups, for a new score of 89 points. This represents a 7% improvement over the non-corrected score.

Paragraph 25. The use of correction factors is not new to sport or fitness testing. The sport of powerlifting uses the Wilks correction factor to compute the best overall lifter of a meet, across all body weight divisions. While the Wilks algorithm is based on a second order polynomial model, it has been shown to appropriately remove body mass bias in nearly the identical manner as the allometric model, upon which the 2/3 body mass exponent is based (33). A recently published (37) and validated (6) handicap model, yields a correction factor for 5K run time based on one's body mass and age. This handicap allows physiologically valid comparisons between individuals of differing age and body mass. That is, the correction factor allows credit for the decrement in performance expected by the *independent* effects of age and body mass, not the confounding effects of lifestyle, effort, or body composition.

Paragraph 26. Correction factors applied to military fitness testing, however, create a situation in which everyone's score either remains unchanged (for lighter personnel) or improves (for heavier personnel). This disrupts the normative bases upon which score standards have been established (26-28). To maintain normative-based standards, a re-scaling of scores based on correction factors should be considered (38). For criterion based standards of occupational fitness, however, future research investigating the threshold levels of corrected scores below which occupational fitness would be generally insufficient to perform physically demanding work tasks is recommended.

SUMMARY

Paragraph 27. The body of research evidence, especially for men, makes a compelling case that the current physical fitness tests of the U.S. Army, Air Force, and Navy are unduly advantageous to lighter personnel. Most physically demanding military tasks, however, are better performed by those with larger lean body mass – the same individuals who tend to be penalized by the high-stakes physical fitness test scores. Given that these tests are measures of health-related fitness and that occupational fitness is better measured via load carriage, lifting, and/or materiel handling tests, the removal of body mass bias appears to a reasonable “middle ground” remedy. Although balanced fitness tests, scaled values and/or correction factors can remove this bias, none is without limitations. Future research should focus on women as well as the development of test events that are fair, practical, and predictive of fitness for work and health for all military personnel.

ACKNOWLEDGMENTS

The results of the present study do not constitute endorsement by ACSM.

REFERENCES

1. Astrand PO, Rodahl K. *Textbook of Work Physiology*. New York, NY: McGraw Hill; 1986. p.399-405.

2. Batterham AM, George KP. Allometric modeling does not determine a dimensionless power function ratio for maximal muscular function. *J Appl Physiol.* 1997;83(6):2158-66.

3. 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(4);1317-1325.

4. Bilzon JL, Allsopp AJ, Tipton MJ. Assessment of physical fitness for occupations encompassing load-carriage tasks. *Occup Med.* 2001;51(5):357-361.

5. Bilzon JL, Scarpello EG, Bilzon E, Allsop J. Generic task-related occupational requirements for Royal Navy personnel. *Occup Med.* 2002;52(8): 503-510.

6. Crecelius A, Vanderburgh PM, Laubach LL. Contributions of body fat and effort in the 5K run age and body weight handicap model. *J Str Cond Res.* (in press).

7. 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 Exerc.* 1996;28(5):S183.

8. Davies MJ, Dalsky GP. Normalizing strength for body size differences in older adults. *Med Sci Sport Exerc.* 1997;29(5):713-717.

9. Dooman CS, Vanderburgh PM. Allometric modeling of the bench press and squat: Who is the strongest regardless of body mass? *J Str Cond Res.* 2000;14(1):32-36.

10. Harman EA, Frykman PN. The relationship of body size and composition to the performance of physically demanding military tasks. In: *Body Composition and Physical Performance.* Washington DC: National Academy Press; 1992. p.105-18.

11. Harman EA, Gutekunst DJ, Frykman PN, Sharp MA, Nindl BC, Alemany JA, Mello RP. *Mil Med.* 2008;173(1):36-41.

12. Heil DP. Body mass scaling of peak oxygen uptake in 20- to 79-yr-old adults. *Med Sci Sport Exerc.* 1997;29(12):1602-1608.

13. Jaric S. Muscle strength testing – Use of normalization for body size. *Sports Med* 2002;615-631.

14. Jaric S, Ugarkovic D, Kukolj M. Evaluation of methods of normalizing muscle strength in elite and young athletes. *J Sports Med Phys Fit.* 2002;42(2):141-151.

15. Jaric S, Radosavljevic-Jaric S, Johansson 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.

16. Jaric S, Mirkov D, Markovic G. Normalizing physical performance tests for body size: A proposal for standardization. *J Str Cond Res.* 2005;19(2):467-74.

17. Knapik JJ, Reynolds KL, Harman E. Soldier load carriage: Historical, physiological, biomechanical, and medical aspects. *Mil Med.* 2004;169(1):45-56.

18. Kraemer WJ, Mazzetti SA, Nindl BC, Gotshalk LA, Volek JS, Bush JA, Marx JO. Effect of resistance training on women's strength/power and occupational performances. *Med Sci Sports Exerc*. 2001;33(6):1011-25.

19. Kraemer WJ, Vescovi JD, Volek JS, Nindl BC, Newton RU, Dziados JE, French DN, Hakkinen K. Effects of concurrent resistance and anaerobic training on load-bearing performance and the Army Physical Fitness Test. *Mil Med*. 2004;169(12):994-9.

20. Lyons J, Allsopp A, Bilzon J. Influences of body composition upon the relative metabolic and cardiovascular demands of load-carriage. *Occup Med.* 2005;55:380-384.

21. Markovic G, Jaric S. Movement performance and body size: The relationship for different groups of tests. *Eur J Appl Physiol*. 2004;92:139-149.

22. Maud PJ. Fitness assessment defined. In: Foster C, Maud PJ. *Physiological Assessment of Human Fitness.* 2nd ed. Champaign (IL): Human Kinetics; 2006. p.6-7.

23. Myers DC, Gebhardt DL, Crump CE, Fleishman EA. Validation of the military enlistment physical strength capacity test. *Technical Report 610.* 1983. U.S. Army Research Institute for the Behavioral Sciences, Natick, MA.

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

25. Rayson M, Holliman D, Belyavin A. Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics.* 2000;43(1):73-105.

26. *U.S. Air Force Instruction 10-248. Fitness Program*. January 2004; p. 55-9.

27. *U.S. Army Field Manual 21-20. Physical Fitness Training.* October 1998; p. 14-4 – 14-6.

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

29. Vanderburgh PM, Mahar MT, Chou CH. Allometric scaling of grip strength by body mass in college-age men and women. *Res Quart Exerc Sport.* 1995;66(1):80-84.

30. Vanderburgh PM, Mahar MT. Scaling of 2-mile run times by body weight and fat-free weight in college-age men. *J Strength Cond Res.* 1995;9(2):67-70.

31. Vanderburgh PM, Katch FI. Ratio scaling of VO_{2max} penalizes women with larger percent body fat, not lean body mass. *Med Sci Sport Exerc.* 1996;28(9):1204-8.

32. Vanderburgh PM. Two important cautions in the use of allometric scaling: The common exponent and group differences principles. *Meas Phys Ed Exerc Sci.* 1998;2(3):153-63.

33. Vanderburgh PM, Batterham AM. Validation of the Wilks Powerlifting formula. *Med Sci Sports Exerc.* 1999;31(12):1869-75.

34. Vanderburgh PM, Dooman C. Considering body mass differences, who are the world's strongest women? *Med Sci Sports Exerc.* 2000;32(1):197-201.

35. Vanderburgh PM, Flanagan S. The Backpack Run Test: A model for a fair and occupationally relevant military fitness test. *Mil Med.* 2000;165(5):418-21.

36. Vanderburgh PM, Crowder TA. Body weight penalties in the physical fitness tests of the Army, Air Force, and Navy. *Mil Med.* 2006;171(8):753-756.

37. Vanderburgh PM, Laubach LL. Derivation of an age and weight handicap for the 5K run. *Meas Phy Ed Exerc Sci*. 2007;11(1):49-59.

38. Vanderburgh PM. Correction factors for body mass bias in military physical fitness tests. *Mil Med.* 2007;172(7):738-742.

39. Vanderburgh PM, Laubach L. Body mass bias in a competition of muscle strength and aerobic power. *J Str Cond Res.* (in press).