9-2011

Load Carriage Distance Run and Pushups Tests: No Body Mass Bias and Occupationally Relevant

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

Nicholas S. Mickley
University of Dayton

Philip A. Anloague
University of Dayton, panloague1@udayton.edu

Kimber Lucius
University of Dayton

Load carriage distance run and pushups tests: No body mass bias and occupationally relevant

Accepted for Publication, May 20, 2011, Military Medicine

Paul M. Vanderburgh, EdD
Nicholas S. Mickley, BS
Philip A. Anloague, DHSc
Kimber Lucius, BS

Department of Health & Sport Science
University of Dayton
300 College Park
Dayton, OH 45469-1210

KEYWORDS
Allometric Scaling
Backpack
Weight Bias

Introduction

The importance of regular physical fitness tests (PFT) in the armed forces is unquestioned, as is the need for regular physical conditioning related to the rigors of the profession. Due to the large numbers of service members in each of the services, PFT events have been designed such that no equipment is needed and mass testing is feasible. The no-equipment requirement essentially dictates that body weight is the primary resistance for each event. Accordingly, all the armed services of the United States have at least one version each of a distance run, timed push-ups (PU), and timed sit-ups tests (1).

Recent evidence, however, suggests that these three events impose a body mass (M) bias against heavier, not just fatter men and women and demonstrate poor correlations with the ability to perform physically demanding occupational tasks. Using a sample of lean and fit male service academy cadets with a large variance in body size, Vanderburgh & Mahar (2) reported correlations between two-mile run time and body mass and fat-free mass of 0.49 and 0.32, respectively. Any positive or negative non-zero correlation in this case would indicate M-bias. Crowder and Yunker (3), also testing the same population but with a sample size of 238, reported a similar bias against larger cadets for the PU, sit-ups, and two-mile run (2MR) events. Markovic and Jaric (4) found comparable biases in 77 males, age 18-26 for one-minute timed PU and sit-ups.

Such M-bias is theoretically based on well-documented allometric relationships between physical performance and body dimensions. Astrand & Rodahl (5) posited that since muscular strength (S) and maximal oxygen uptake (VO_{2max}, in $L \cdot min^{-1}$) were each directly proportional to blood vessel and muscle cross sectional area, respectively, then both S and VO_{2max} must be proportional $M^{2/3}$. This is based on the principle that mass is a 3-dimensional construct and cross sectional areas are 2-dimensional. As a result, the case has been made that the optimal method of expressing VO_{2max} and S is $ml \cdot M^{-0.67} \cdot min$- and $S \cdot M^{-0.67}$ respectively (1,2). Therefore, to express them in their conventional way using an M exponent of 1.0 for VO_{2max}, or an exponent of 0 for strength, is to introduce M-bias.

The impact of this on how to express PU and 2MR without M-mass bias has both theoretical and empirical support (1), the details of which exceed the scope of this paper. Nonetheless, others have determined that the optimal M exponent for expressing pushups (PU) and 2MR scores is 1/3 and is as follows: $PU \cdot M^{0.33}$ and $2MR \cdot M^{-0.33}$ (2,3,4,6). These follow from the findings that $PU \sim M^{-1/3}$ and $2MR \sim M^{1/3}$. These scaled scores allow comparisons of PU and 2MR performances between individuals of different M such that M-bias is removed. Note that the signs of the exponents in the scaled scores are due to the fact that larger numbers are better for PU but smaller are better for 2MR. The practical effects of these M-biases have been quantified. If a 90 kg man scored the physiological equivalent of a 60 kg man, same for a 75 kg vs. 50 kg woman, such that scaled scores were equal, his/her actual fitness test score would be 15-20% lower (7). Because these fitness test scores factor significantly into consideration for promotion and advancement, the penalty against heavier men and women has substantive implications beyond physical fitness.

The use of such scaled scores, however, is not practical. Computation, not feasibly done by hand, also yields strange units that are not easy interpretable. As a result, other scoring methods have been proposed. Correction factors, dimensionless numbers based one's M, that are multiplied by the raw score, have been proposed (8). Mathematically equivalent to scaled scores, they have the advantage of preserving the original units of the raw score and are easily interpretable. They still require tables and are not intuitively attractive to non-exercise-scientists. A backpack run distance run test, in which each service member carries the same weight backpack, has been proposed (9). In this study, actual 2MR data of male military academy cadets was used to mathematically model the influence of experimental alterations in backpack weight on M-bias. Results suggested a range of backpack weights within which M-bias was statistically zero. No empirical validation of this model, however, has been done.

Based on the present authors' experiences, some manuscript reviewers have been highly skeptical of the use of allometric scaling which, by definition, gives a "credit" for heavier BW. The skepticism is based primarily on the assertion that this encourages weight gain for better scores. Of key importance, however, is the fact that a scaled score is comprised of two scores: PU or 2MR, for example, and M raised to a certain exponent. While the heavier M would tend to improve a scaled score, the other key factor that must be considered is the effect of the added weight on the PU and 2MR scores. As Vanderburgh & Laubach modeled mathematically (10) for distance running, allometrically scaled scores should be worse after gaining fat mass because the credit granted for the heavier M is less than the penalty caused by the effect of the added fat mass on the raw scores. This assertion, however, has not been tested empirically for any physical fitness test.

Therefore, the purpose of this study was two-fold: 1. To assess the effects of experimentally adding 30 lbs. of excess mass in the form of a loaded military backpack for the Army's PU and 2MR tests on M-bias; and 2. To test the hypothesis that allometrically scaled scores in the loaded conditions for both events are worse than in the unloaded conditions.

Methods

Subject selection was based on criteria employed by Vanderburgh & Mahar (2) for a body-mass bias study: fit, lean, college-age, male cadets. As such, they had previous experience with backpack load carriage under various conditions and would be familiar not only with military PFTs but also tactics on how to best perform these arduous tests. These criteria would reduce the confounding effects of effort and body fatness. Women were not included because of lack of access to a sufficiently large enough sample. The university's institutional review board approved this study and written informed consent was obtained from every subject prior to data collection.

We chose the standard U.S. Army issue military backpack loaded with issued Army gear to a total weight of 30 lbs. as the mechanism by which subjects carried the extra load for both the PU and the 2MR. This was based on the fact that certain U.S. Army soldiers (e.g., infantrymen) are tasked with "rucksack runs" which require sustained distance running with a loaded backpack. The 30 lbs. was chosen based on previous modeling of a 2MR backpack (9). In that paper, researchers used metabolic equations to estimate the resultant run times under different loaded conditions and determined that the lightest zero-bias condition occurred at approximately 30 lbs. The choice of this lightest possible load that still met the zero-bias standard follows from the documented relationships between load carriage and musculoskeletal injuries (11).

Subjects were 56 male ROTC cadets at a midwestern U.S. university. All had completed a full six weeks of physical training as part of the ROTC program, which included at least two sessions of acclimation to the PU and 2MR events with the loaded backpack. All subjects reported for two days of testing, each separated by one week. On both days, subjects participated in a full Army Physical Fitness Test (12). While both days required the situps test, this event was omitted from any analysis in the present paper since it could not be done in a loaded condition.

On the first day, testing was in the unloaded condition and on the second, the PU and 2MR were done in the loaded condition (wearing the 30 lbs. backpack). Prior to the events on the first day, height, weight, and % body fat were assessed via a calibrated balance scale with stadiometer, and three-site skinfolds (chest, abdomen, and thigh skinfolds and the Jackson Pollock (13) equation). Prior to any events on the second testing day, all loaded backpacks were also weighed to ensure that each was within 1 lb. of the 30 lbs testing weight. Subject descriptive data (means ± SD) were M = 182.8 ± 25.0 lbs, height = 70.5 ± 3.0 in, and body fat % = 13.5 ± 5.9%.

Subjects completed the PU event for both days using the standards prescribed by the U.S. Army (12). Half of the subjects performed the maximum number of properly executed PU in two minutes scored by a counterpart then switched. After at least a 15 min break, subjects then ran the 2MR test (unloaded first day, loaded second day) on a flat, asphalt course with the goal of minimal time to complete.

A simple Pearson Product Moment correlation coefficient was used to determine the magnitude of M-bias in both load conditions with M and PU or 2MR as the variable pairs. A log-linear transformation with regression (2) enabled the determination of M exponents with confidence intervals. This transformation allowed for comparison of these empirically determined exponents with others cited in the literature, including those derived theoretically. A paired t-test was calculated to compare the difference in scaled scores between the loaded and unloaded conditions. For the standard of statistical significance and because multiple analyses were done, alpha levels for means comparisons were set at 0.01. For exponent calculations, the 95% confidence intervals were provided in congruence with conventions in allometry research (1).

Results

Table 1 shows subject performance data. Compared to the unloaded conditions, loaded performance decreased by 32.2% and 18.8% for the PU and 2MR, respectively. As shown in Fig.1 and 2, in the unloaded condition, the correlations between M and performance were -0.38 ($p = 0.0039$) and 0.42 ($p < 0.0013$), for PU and 2MR, respectively, suggesting significant M-bias. These correlations are similar to those reported elsewhere for the same tests with military academy male cadets (2,3). As indicated in the same figures, the correlations for the loaded conditions were -0.06 and 0.06 ($p = 0.661$ for both), suggesting that the loaded condition eliminated the M-bias.

Table 1. Subject Performance Mean, (SD)

	PU (Reps)	2MR (sec)
Unloaded	61.3 (13.4)	857.0 (98.2)
Loaded	41.6 (11.9)	1017.8 (108.0)

Using the allometric scaling technique described elsewhere (2), log-linear regression yielded M exponents and 95% CI of 0.66 ± 0.46 and 0.29 ± 0.20 for pushups and 2MR respectively. With the 95% confidence intervals containing 1/3 but not 0, these exponents are congruent with those reported elsewhere (2,3,4,6) for the same population of college-age military cadets for PU and 2MR.

Scaled scores were calculated for each subject by event using the previously referenced conventions: $PU \cdot M^{0.33}$ and $2MR \cdot M^{-0.33}$ in the unloaded conditions but adding 30 lbs. to M in the loaded. Paired t-tests for each event revealed that the loaded scale scores were 38% and 12% worse than the unloaded for PU and 2MR, respectively ($p < 0.001$ for both). Fig. 3, which graphically depicts the scatterplots of loaded vs. unloaded scaled scores, better illustrates the trend that nearly every subject had a worse loaded scaled score.

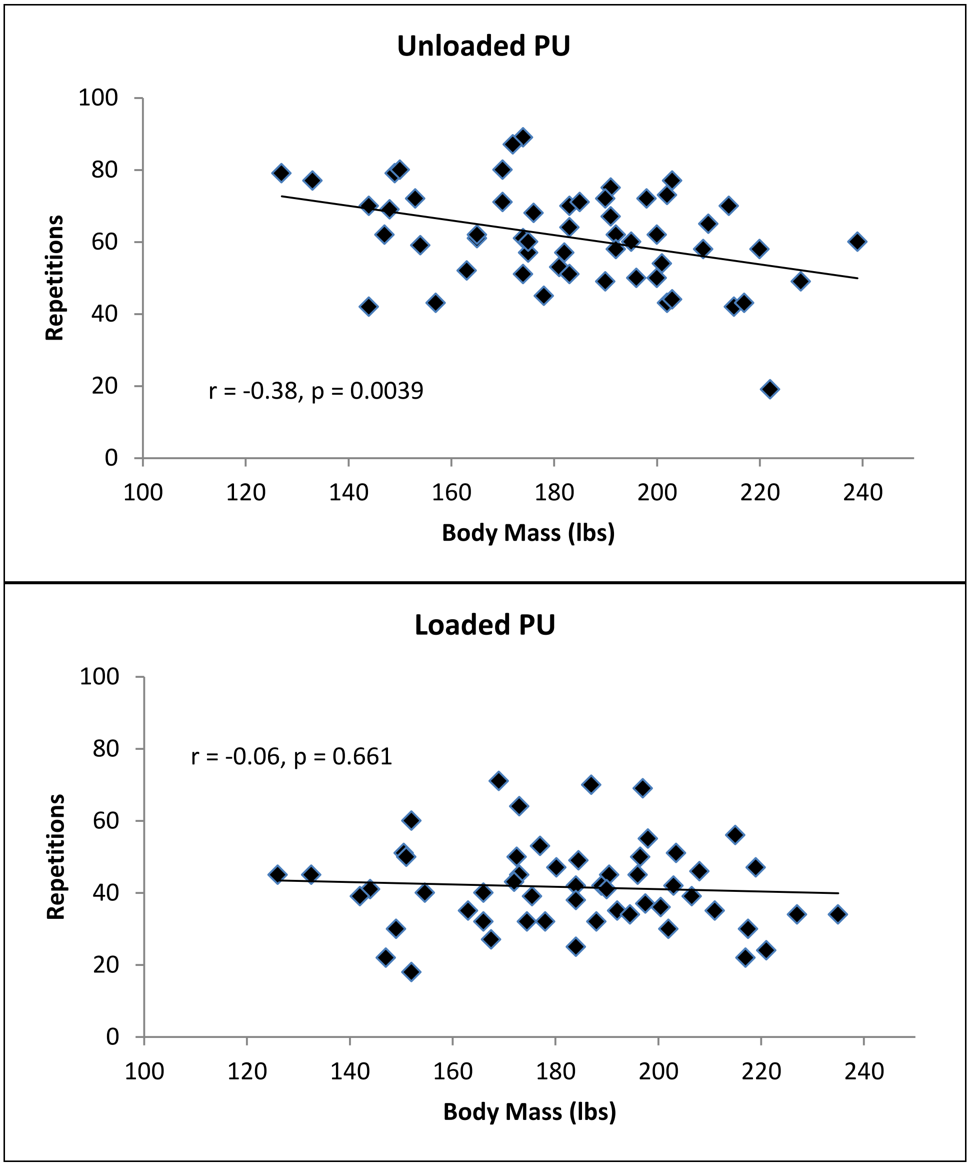

Figure 1. Body Mass vs. Pushups (PU) Repetitions in the unloaded (no backpack) and the loaded (with 30 lbs. backpack) conditions.

Discussion

This is the first study, to our knowledge, that has empirically examined M-bias in a load-carriage test for either a distance run or a pushups test. While Vanderburgh & Flanagan (9) used both theory and empirical data to predict that a 30 lbs backpack would contribute to no M-bias, their assertion was based on metabolic equations and the associated inherent assumptions. It also did not propose a loaded test for pushups nor did it consider the efficacy of such tests. The present finding that the 30 lbs backpack load for the 2MR eliminated the M-bias is then congruent with what theory and empirical data predicted. On the other hand, the finding that M-bias was similarly removed for the loaded PU is somewhat fortuitous as there was no prediction model or theoretical foundation from which such a hypothesis could be tested – other than the likely assertion that some load would yield zero bias.

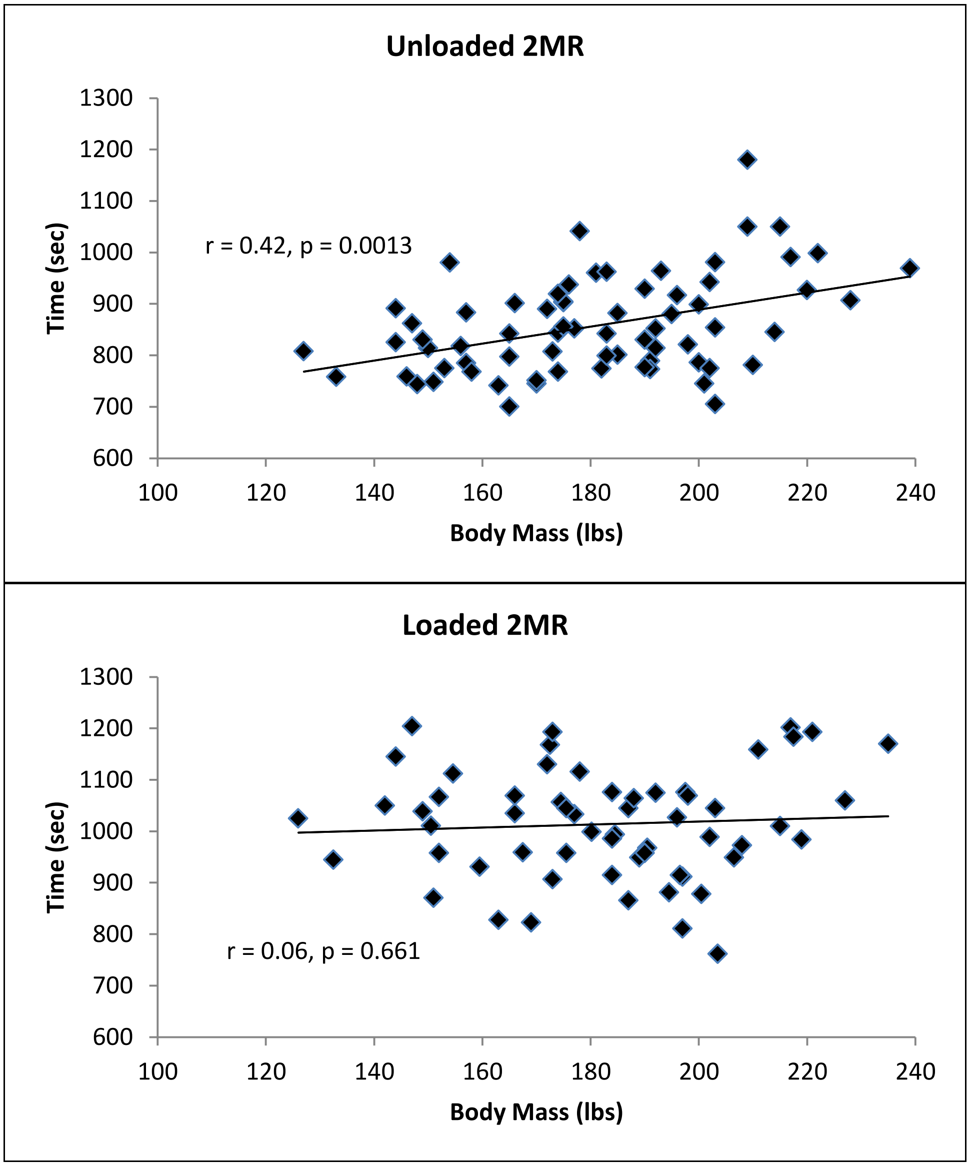

Figure 2. Body Mass vs. Two Mile Run time (2MR) performance in the unloaded (no backpack) and the loaded (with 30 lbs. backpack) conditions.

Evaluation of the present results includes an examination of how the loaded conditions for the PU and 2MR compare with those predicted from theory or metabolic equations. The predictive technique of Vanderburgh & Flanagan (9) is an appropriate start. Using the same simplified (14) metabolic equations from the American College of Sports Medicine (with VO_{2max} in $mL \cdot kg^{-1} \cdot min^{-1}$):

$$VO_{2max} = 3.5 + 5.36 \text{ (mph)} \quad \text{Eq. 1}$$

We determined the predicted loaded 2MR, based on adding 30 lbs of excess weight and assuming no change in maximal absolute oxygen consumption (VO_{2max}, in $L \cdot min^{-1}$). A paired t-test of actual vs. predicted loaded 2MR indicated no significant difference ($p = 0.38$). Perhaps more importantly, and as graphically depicted in Fig. 4, the Pearson Product Moment correlation between these two variables was 0.752 ($p < 0.001$), suggesting not only no difference but a high correlation. This suggests that the loaded 2MR predicted results match the actual with a reasonably high degree of accuracy.

Examination of the same for pushups cannot be done with metabolic equations which apply only to events limited by maximal oxygen consumption. We can use, however, the theoretical reasoning of Jaric and colleagues (6) which states that the ability to do pushups is directly proportional to upper body strength, a two-dimensional variable, and inversely proportional to mass, which is three dimensional. Since the loaded condition is analogous to adding metabolically inert fat mass, then the decrement in performance should be greater than that expected by the laws of scaling, which are based on the scale model assumption. In the present sample and assuming the scale model, the loaded PU performance should be 95.1% of unloaded. This is calculated as the ratio of loaded vs. unloaded M raised to the -1/3 power, or $(212/182 \text{ lbs})^{-1/3}$. Actual loaded PU performance, however, was 67.8% of loaded, calculated as the ratio of loaded vs. unloaded PU, or 41.6/61.3. This larger decrement would clearly be expected since the added weight in this study contained no muscle mass and would therefore violate the scale model assumption. Furthermore, empirical evidence involving adding external weights in a similarly designed study for the pull-up exercise corroborates this large performance decrement (15).

The empirical examination of the change in scaled scores from the unloaded to the loaded condition is important because only theoretical calculations have been used to make the case that adding fat, or other external mass as in a backpack, would lead to worse scaled scores (10). The present finding represents the first empirical evidence that experimentally adding dead mass, analogous to adding fat mass, leads to worse scaled scores, even though scaled scores offer more credit for more M. Fig. 3 illustrates the clear trend that nearly every loaded scaled score was worse than its unloaded pair.

Observation of the loaded condition suggested that the loaded test was efficacious and logistically practical. The latter was based on the fact that all cadets were issued the standard military issue backpacks as well as the gear that would yield the 30 lbs. total weight. Though running with a backpack is an activity that has precedent in the armed services, performing PU wearing a backpack has not, to our knowledge, been a widely used military physical training event. Nonetheless, subjects reported few, if any, problems with executing proper repetitions. Anecdotal observations of the loaded PU event were that the backpack neither inhibited nor disrupted the pushup movement but, of course, contributed to earlier fatigue. Furthermore, the importance of the ability to carry external loads beyond BW has been well established for service members (1,16,17,18), including substantial evidence that current push-ups, sit-ups and distance run tests are not good correlates of physically demanding occupational tasks of the military (16). Lastly, the loaded test requires no special calculations, correction factors, or non-intuitive interpretations. The case can be made, then, that a two-event test of a loaded PU and 2MR is fairer and more occupationally relevant than the current three-event tests in the unloaded conditions.

Any recommendation for the use of a loaded 2MR should be accompanied by research evidence of efficacy and injury risk along with recommendations to mitigate this risk. Because the loaded PU test involves no similar ground reaction forces, only the 2MR is discussed in this context. Well documented is the fact that with repetitive load bearing marches, bone resorption in the lower extremities occurs before the bone remodels as a result of the training stimulus (11,19,20,21). The bone, then, is temporarily more susceptible to injury. This illustrates the necessity for not only an acclimation period for those new to load carriage but adequate recovery periods as well. An increase in rucksack load may contribute to the potential for injury due to the increased vertical loading rates rather than just the ground reaction forces (20). This underscored our desire to use the lightest load possible for testing that would still elicit zero M-bias. Another notable risk factor is gender, given that women are more susceptible to stress fractures than men (11,19,20). For this reason, a standard backpack weight of 30 lbs for women, which is a greater percentage of body weight as compared to men, would perhaps be ill-advised. Further research is warranted to evaluate lighter backpack weights for women.

An extensive review of load carriage by Knapik et al. (11) suggests that training for backpack load carriage events can reduce injury risks associated with the actual tests. Noteworthy is that

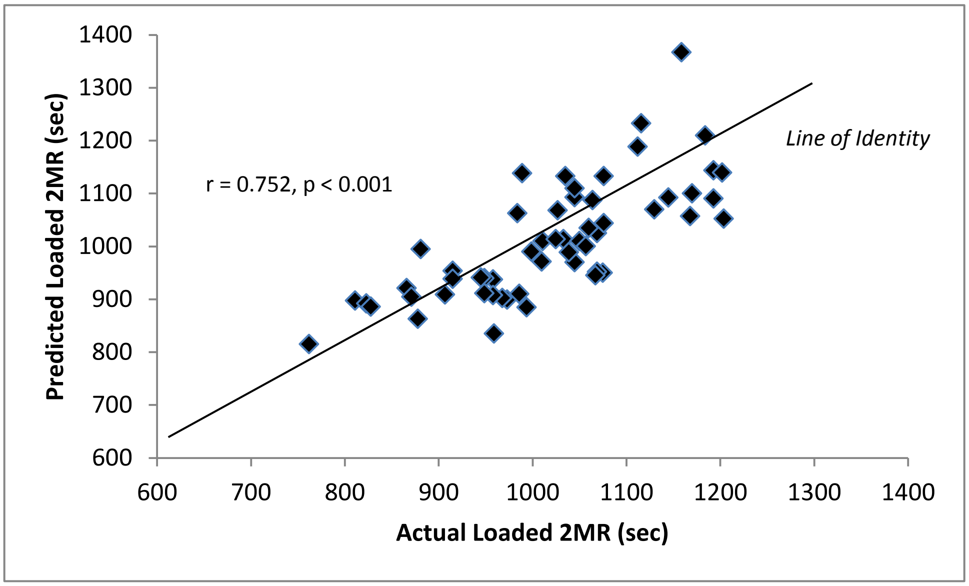

Figure 3. Actual vs. Predicted Loaded Two-Mile Run (2MR) performance. Prediction is based on simplified metabolic equations from the American College of Sports Medicine (14).

nearly all empirical data were based on loads above 20 kg (40% heavier than the present study at 30 lbs. load). Furthermore, the incidence of injury was reduced with preconditioning, abdominal strengthening and/or lower-extremity strengthening exercises. These researchers concluded that load carriage training, if part of one's essential tasks, should be done at least twice monthly with occupationally appropriate loads and distances. New unit members should be given adequate time to adapt to the maintenance level. Other strategies to reduce stress fracture risk include use of specialized insoles, cessation of cigarette smoking, and improved nutrition (11,22). The present study, however, involves running with backpacks, an activity that has not been nearly as extensively studied as forced marching and further epidemiological examination of the effects of loaded running speeds on injury risk is warranted.

In summary, these results suggest that both a PU and 2MR test while wearing a 30 lbs. backpack yields results free of M-bias in lean, college-age men. The tests appear practical and efficacious and do not require special calculations or correction factors. Carrying a uniform external load mimics actual field conditions for many service members and is more occupationally relevant than conditions in which BW is the only resistance. Regarding our second hypothesis, the addition of such external weight leads to significantly worse allometrically scaled scores, a hypothesis that had not been empirically tested. This validates the contention that gaining fat mass has a detrimental effect on scaled scores. The complexities of identifying simple tests of occupational relevance for the profession of arms, the physical demands of which are so varied, suggest that better occupational fitness assessments require more task-specific tests (16-18). Furthermore, unit missions and tasks evolve with changes in technology and threat. Therefore, continued research is needed to develop and validate new and existing tests of occupational and health-related fitness. The present results are not meant to provide a one-size-fits-all solution but may be combined with commanders' prerogatives to incorporate very mission-specific assessments as appropriate for each unit.

References

1. Vanderburgh PM: Occupational relevance and body mass bias in military physical fitness tests, Med Sci Sports Exerc 2008; 40:1538-45.
2. 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; 9:67-70.
3. 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, Med Sci Sport Exerc 1996; 28:S183.
4. Markovic G, Jaric S. Movement performance and body size: The relationship for different groups of tests, Eur J Appl Physiol 2004; 92:139-49.
5. Astrand PO, Rodahl K: Textbook of Work Physiology, pp 399-405, New York, McGraw Hill, 1986.
6. Jaric S, Mirkov D., Markovic G: Normalizing physical performance tests for body size: a proposal for standardization, J Strength Cond Res 2005; 19:467-74.
7. Vanderburgh PM, Crowder TA: Body weight penalties in the physical fitness tests of the Army, Air Force, and Navy, Mil Med 2006;171:753-56.
8. Vanderburgh PM: Correction factors for body mass bias in military physical fitness tests, Mil Med 2007; 172:738-42.
9. 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.
   Vanderburgh PM, Laubach LL: Derivation of an age and weight handicap for the 5K run, Meas Phys Ed Exerc Sci 2007; 11:49-59.
10. Knapik JJ, Reynolds KL, Harman E: Soldier load carriage: Historical, physiological, biomechanical, and medical aspects, Mil Med 2004; 169:45-56.
11. U.S. Army Field Manual 21-20. Physical Fitness Training 1998; p.14-4 – 14.6.
12. Jackson AS, Pollock ML: Generalized equations for predicting body density of men, Brit J Nutr 1978; 40:497-504.
13. Swain DP, Leutholtz BC: Metabolic Equations Simplified, p. xi. Baltimore, Williams & Wilkins, 1997.
14. Vanderburgh PM, Edmonds: The effect of experimental alterations in excess mass on pull-up performance in fit young men, J Str Cond Res 1997; 11:230-33.
15. Bilzon JL, Scarpello EG, Bilzon E, Allsop J: Generic task-related occupational requirements for Royal Navy personnel, Occup Med 2002; 52:503-10.
16. Bilzon JL, Allsopp AJ, Tipton MJ: Assessment of physical fitness for occupations encompassing load-carriage tasks, Occup Med 2001; 51:357-61.
17. 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; 43:73-105.
18. Kiuru MJ, Pihlajamaki K, Ahovuo JA: Bone stress injuries, Acta Radiologica 2004; 45:317-26.
19. Zadpoor AA, Nikooyan AA: The relationship between lower-extremity stress fractures and ground reaction force: A systematic review, Clin Biomech 2010; 26:23-28.
20. Chatzipapas CN, Drosos GI, Kazakos KI, Tripsianis G, Iatrou C, Verettas DAJ: Stress fractures in military men and bone quality related factors, Int J Sports Med 2008; 29:922-26.
21. Hinz P, Henningen A, Matthes G, Jager B, Ekkernkamp A, Rosenbaum D. Analysis of pressure distribution below the metatarsals with different insoles in combat boots of the German Army for prevention of march fractures, Gait and Posture 2008; 27:535-38.