



Body Mass Bias in a Combat Fitness Test

Ross R. Vickers, Jr.
John H. Reynolds
Brian J. McGuire

Naval Health Research Center

Report No. 11-20

Naval Health Research Center
140 Sylvester Rd.
San Diego, California 92106-3521

Body Mass Bias in a Combat Fitness Test

Ross R. Vickers, Jr.[1]
John H. Reynolds[2]
Brian McGuire[3]

[1]Naval Health Research Center
140 Sylvester Road
San Diego, CA 92106-3521

[2]Center for Naval Analyses
4825 Mark Center Drive
Alexandria, VA 22311

[3]United States Marine Corps
Training and Education Command
Quantico, VA 22134-5027

Report No. 11-20 was supported by the Office of Naval Research, Arlington, VA, under Work Unit No. 60704. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release; distribution is unlimited. Human subjects participated in this study after giving their free and informed consent. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

Abstract

Allometric theory predicts that, pound for pound, lighter individuals will perform better than heavier individuals on strength and endurance tests. This study evaluated body mass bias as a factor in the U.S. Marine Corps Combat Fitness Test (CFT), which consists of movement to contact (Movement), an ammunition can lift (Lift), and maneuver under fire (Maneuver). Allometric modeling indicated small to moderate biases favoring lighter Marines for each CFT element. The biases were smaller than predicted by theory. Age and gender did not modify the bias. Wearing personal protective equipment eliminated the biases for the movement to contact and maneuver under fire. Despite the biases for individual elements, CFT scores were not related to weight. This difference reflects the fact that CFT scores are based on absolute performance. The better performance of heavier individuals on Lift offset their poorer performance on Movement and Maneuver. The CFT elements are biased measures of physical fitness, but the CFT is an unbiased measure of fitness for duty. These apparently contradictory conclusions derive from the difference between fitness defined as relative pound-for-pound performance and the absolute performance required by combat tasks. CFT scores did not correlate with weight, so the CFT will not adversely affect personnel decisions.

Body Mass Bias in a Combat Fitness Test

Military organizations use physical fitness tests to assess physical readiness. These tests may be biased against larger individuals (Vanderbergh & Crowder, 2006), so bias is an issue when a new test is introduced. This study evaluated the bias of the recently promulgated U.S. Marine Corps Combat Fitness Test (CFT; Commandant of the Marine Corps, 2008).

Allometric theory is the basis for asserting that fitness tests are biased. The general form of an allometric model is $y = AM^b$. When considering fitness tests, y is a performance variable (e.g., oxygen uptake or 3-mi run time), A is a scaling coefficient, M is mass, and b is the mass coefficient. The scaling coefficient determines the general range of test scores. The mass coefficient indicates the presence or absence of body mass bias. If $0 < b < 1$, the test bias favors smaller individuals.

Allometric theories commonly predict biases favoring smaller individuals. Typical theoretical derivations give $b = .75$ or $b = .67$. However, the predicted value can depend on the specific activity (Jaric, Mirkov, & Markovic, 2005; Markovic & Jaric, 2004) and the exertion level for that activity (Darveau, Suarez, Andrews, & Hocachka, 2002; Suarez & Darveau, 2005). Some formulations may be valid only under special circumstances, such as an infinite size range (Savage, Deeds, & Fontana, 2008). Also, models can rely on invalid assumptions (Nevill, Stewart, Olds, & Holder, 2004). As a consequence, it is not surprising that there is no current consensus regarding the mass coefficient (Agutter & Wheatley, 2004; White & Seymour, 2005). The important point for present purposes is that all of the theories yield mass coefficient estimates that fall between 0 and 1. The magnitude of the bias is the subject of the debate; there is a consensus that a bias exists.

One reason for the difficulty in reaching a consensus is that there may be no single mass coefficient. The metabolic rate coefficient increases with exertion (Jones & Lindstedt, 1991; Markovich, Vucetic, & Nevill, 2007; Batterham & Jackson, 2003; Bergh, Sjodin, Forsberg & Svendenhag, 1991; Chamari, Bouchaidi, Hachanga, Kaouech, & Wisloff, 2005). This coefficient may depend on subject characteristics, including age (Weir, Housh, Johnson, Housh, & Ebersole, 1999), stature (Batterham, Tolfrey & George, 1997; Folland, McCauley & Williams, 2008), body composition (Batterham, Vanderbergh, Mahar, & Jackson, 1999; Folland et al., 2008), and habitual physical activity (Batterham et al., 1999; Jensen, Johansen & Secher, 2001; Markovich et al., 2007). Subject and activity factors may have to be considered jointly to predict the mass coefficient. For example, Markovic et al. (2007) found that as exertion increased, the metabolic rate mass coefficient increased in untrained subjects, but decreased in trained subjects.

The uncertainty regarding the mass coefficient is directly relevant to assessing bias in military fitness tests. For example, obstacle course performance involves a complex set of activities. It is reasonable to expect the mass coefficient for overall performance to be some function of the mass coefficients for the component activities. Each activity may have a different mass coefficient (Jaric et al., 2005), so the coefficient for overall performance is uncertain. This complexity might explain why previous studies have found that obstacle course performance has been essentially independent of mass (Bishop, Crowder, Fielitz, Lindsay, & Woods, 2008; Kusano, Vanderbergh, & Bishop, 1997). The CFT is similar to an obstacle course in that it is a composite of several activities. The combination of theoretical uncertainty about the likely magnitude of bias and empirical evidence that the CFT may not be biased at all makes the CFT a prime example of Gould's (1966) observation that theoretical claims for body mass bias always require empirical confirmation. This report provides the requisite empirical evaluation of CFT body mass bias.

Methods

Sample

Data were collected from 2,428 Marines serving at 18 U.S. Marine Corps installations. The sites were selected to ensure that the sample contained a range of Marine Corps occupations, and that tests were performed in different physical environments (e.g., desert, mountains). Participants were volunteers from the rosters of participating units at each installation.

The research team recorded self-reports of age, rank, gender, most recent Physical Fitness Test (PFT) score, and the 3-mi run time for that PFT at each data collection site. Weight was measured on a balance scale at the test site. Most individuals performed the test in a t-shirt, utility trousers, and boots. A subset wore the personal protective equipment (PPE) worn in combat in addition to the basic test attire.

Men and women differed significantly ($p < .001$) on every sample descriptor (see Table 1). By Cohen's (1988) effect size (ES) criteria, the weight, $r_{pb} = -.467$, and run time, $r_{pb} = .371$, differences were moderately large; the PFT score difference, $r_{pb} = -.129$, was small; and the age difference, $r_{pb} = -.074$, was trivial. The correlations among the descriptors were small or moderate except for the large correlation of PFT score with 3-

mi run time. This correlation was expected because points derived from performance on the 3-mi run time are one component of the overall PFT score.

Marine Corps fitness standards differ by gender and age group, so these factors were predictors in the statistical models tested in this paper (see Data Analysis). Table 2 gives the sample sizes within each age group.

Combat Fitness Testing

The CFT consists of the time to complete an 880-yd run, the number of times a 30-lb (13.6 kg) ammunition can be lifted from chest height, to full arm extension, and returned to chest height in 2 min, and the time to complete a simulated maneuver under fire. These CFT components are referred to as Movement, Lift, and Maneuver in this paper. The CFT was not modified to allow for differences in the physical settings across test sites. Analyses indicated that test site differences had little effect on CFT performance (see Appendix A).

Gender had a strong effect on CFT performances (Table 3). The average woman took 40 s longer for Movement, completed 28 fewer repetitions for Lift, and took 58 s longer for Maneuver. Each difference represented a large effect by Cohen's ES criteria: Run, ES = 1.37; Movement, ES = 1.65; Lift, ES = -1.84; Maneuver, ES = 1.44. The differences were moderate when expressed as point-biserial correlations.

Wearing PPE impaired CFT performances (see Table 3). Movement increased by 11 s, Lift decreased by 3 repetitions, and Maneuver increased by 26 s. The differences generally represented small-to-moderate effects by Cohen's ES criteria: Movement, ES = .47; Lift, ES = -.23; Maneuver, ES = .68. These CFT performance differences occurred even though the PPE and non-PPE groups had previously performed comparably on the PFT, ES = .20, and the 3-mi run, ES = .11, (Table 1). The small PFT score difference was statistically significant, $p < .005$, but the trivial run time difference was not, $p = .101$.

Aggregate CFT Scores

The aggregate CFT score in this study was the usual Marine Corps point scoring system (Commandant of the Marine Corps, 2008). This scoring assigns 60 points on each CFT element to performance equaling the cutoff between the 1st and 2nd percentiles of the score distribution. A score of 100 points is received for any performance that exceeds the 90th percentile of the score distribution. Points for intermediate performances are distributed evenly over the range from the 1st to 90th percentiles. Marines whose performance falls below the 60 point level receive no score. These procedures are applied separately to performances registered in each age/gender group (e.g., 17–26 year-old males). The overall CFT score is the sum of the three component scores for a Marine who received at least 60 points on all three components. Marines who failed one or more components receive no score. Thus, CFT scores range from 180 points to 300 points for Marines who meet the minimum standard for each individual component.

Converting Run Times

Run times were converted to metabolic rate estimates for secondary analyses. Many allometric models, both theoretical and empirical, have estimated mass coefficients for metabolic rate. The run time-to-metabolic rate conversion made it possible to compare the present findings to that earlier work.

Metabolic rate is a function of velocity. When a person runs over flat ground, the energy expenditure rate is $\dot{V}_{O_2} = 0.2v + 3.5$, where $\dot{V}_{O_2}$ is oxygen uptake in $\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$

and v is velocity in $m \cdot min^{-1}$ (American College of Sports Medicine at www.acsm.org). The equation applies when $v \geq 120$ $m \cdot min^{-1}$. This minimum velocity criterion is satisfied when the 3-mi run is completed in less than 40:14 min and when the 880-yd run is completed in less than 6:42 min. All study participants met both criteria.

Data Analysis

All analyses were performed with SPSS-PC, Version 17. Initial bivariate analyses were linear regression and correlation. The primary multivariate analysis procedure was analysis of variance (ANOVA).

A natural logarithm transformation was applied to CFT performances and to weight to convert the allometric model to a linear form. Preliminary analyses demonstrated that this linear variant produced the same results as fitting the nonlinear (see Appendix B). The linear variant was adopted to simplify tests for age and gender effects on the mass coefficient.

The central analyses fitted multivariate allometric models to CFT data. Log-transformed CFT performances were the dependent variables. The initial predictors were age category, gender, and log-transformed weight plus the two-way interactions between these three predictors. The interactions were omitted from the final models even though preliminary analyses indicated that 8 of 9 interactions (3 tests with 3 interactions per test) were statistically significant ($p < .014$); the gender x weight interaction for Movement was the exception ($p = .219$). However, the statistical significance for the interactions was basically a function of the large sample size for the analyses as the variance explained by each interaction was trivial, ES $\leq .004$.

ES was the metric for interpreting the results. Statistical significance was not a useful criterion for identifying important associations because the large sample size meant that even trivial effects could be statistically significant (Rosenthal & Rosnow, 1984). Cohen’s (1988) criteria classify effects as trivial (i.e., too small to be of practical or theoretical importance), small, moderate, or large.

Results

Performance Prediction

Bivariate analyses. CFT performance displayed small to moderate correlations with age, weight, PFT score, and 3-mi run times (Table 4). The gender differences reported in Table 3 represented moderately large effects: Movement, $r_{pb} = .463$; Lift, $r_{pb} = -.479$; Maneuver, $r_{pb} = .417$.

Allometric Model for Movement

Table 5 presents the multivariate allometric model for Movement. The table entries yield group-specific allometric models when the appropriate parameter values are inserted into the generic equation

$$\ln(\mathrm{p}) = A + B_{\mathrm{i}} + C_{\mathrm{j}} + (b^{*}\ln(\mathrm{W})) \qquad (1)$$

The dependent variable is the performance measure that is being modeled. The model coefficients include the intercept, *A*, the gender effect, *B*, the age effect, *C*, and the mass coefficient, *b*. The Equation 1 subscripts refer to gender (*i*) and age group (*j*). The generic equation is converted into a group-specific Movement equation by substituting the appropriate values from Table 5. For example, the group-specific inserting $A = 4.629$, $B =$

-0.227, C = -.101, and D = .191 into Equation 1 yields Equation 2 as the allometric equation for 17–26 year-old men.

$$\ln(\text{Movement}) = 4.629 - .227 - .101 + (.191*\ln(W))$$

$$= 4.301 + .191*\ln(W) \tag{2}$$

Reversing the log-transformation converts the group-specific allometric equation to the original nonlinear allometric form.

$$Movement = 77.774W^{.191} \tag{3}$$

Because age and gender only contribute to the scaling constant, A, Equation 4 is a generic mass bias equation for Movement that applies to all age/gender groups.

$$Movement = A_G W^{.191} \tag{4}$$

The group-specific value of A is obtained by inserting the appropriate B and C values from Table 5 into Equation 1.

Table 5 gives the ES for each predictor in the form of a partial ε^2. Each ES is based on the difference between a specific group and the reference group. Women were the gender reference group, and Marines who were 46+ years of age were the age reference group. The three age effects represent the differences between the specified younger age group and the 46+ reference group. Applying Cohen's ES criteria, the partial ε^2 values in the table indicated that gender produced a large effect. The difference between Marines in the 40–45 and 46+ age groups was trivial. All other effects were small.

Allometric Model for Lift

Higher Lift scores indicated better performance, so the positive coefficients in Table 6 indicate better performance than that seen in the reference group. The gender difference favored men and was moderately large. Age had little or no effect. The mass coefficient was <1.00, so lighter individuals performed better on a pound-for-pound basis. The mass effect was at the upper boundary of the range for small effects.

The weight effect on Lift performance was consistent with some theoretical formulations. Specifically, the result was consistent with theoretical arguments that produce $b = .67$ as the mass coefficient. This value fell just within the upper boundary of the 95% confidence interval for the empirical estimate, [.525, .693].

Allometric Model for Maneuver

The structure of the Maneuver model (see Table 7) paralleled the Movement model. The gender effect was substantial and the difference between the 40–45 and 46+ age groups was trivial. The three remaining model parameters were associated with small effects.

One similarity between the Movement and Maneuver models was particularly important. The Maneuver mass coefficient, $b = .221$, was almost identical to the corresponding Movement coefficient, $b = .191$. The substantial overlap of the 95%

confidence intervals for these two parameter estimates (see Tables 5 and 7) implies that the coefficients were not significantly different.

Personal Protective Equipment Effects

Two hundred and fifty-seven (257) men completed the CFT in PPE (see Table 8). The PPE sample was younger and heavier than the non-PPE sample. The PPE sample had higher PFT scores despite slightly slower 3-mi run times.

Wearing PPE impaired CFT performance. Times increased for Movement and Maneuver and fewer Lifts were completed. The Lift and Movement effects were small; the Maneuver effect was moderate.

All three PPE mass coefficients were smaller than the corresponding non-PPE coefficients (see Table 9). The PPE coefficients indicated minimal body mass bias for Movement and Maneuver. The total absence of bias could not be ruled out in either case, as $b = 0$ fell within the 95% confidence intervals for both CFT components. The Lift PPE coefficient was significantly greater than zero, but mass bias was reduced somewhat as the PPE coefficient was significantly less than the non-PPE coefficient.

CFT Component Bias

The analyses presented to this point indicated body mass bias was present in CFT performance. The next question was, “How much are heavier Marines being penalized?” This question was answered by repeating Vanderbergh’s (2007) computation of correction factors that would be needed to eliminate the bias. The correction factors were derived by taking the 10th percentile of the weight distribution as a point of reference. The correction factor for a given weight is $M = (W / W_R)^b$ where W / W_R is the ratio of an individual's weight, W, to the reference weight, W_R, and b is the mass coefficient (Vanderburgh, 2007).

Table 10 compares corrections based on theory to empirically-based corrections. The empirical corrections were less than predicted from theory. The trend was particularly pronounced for Movement and Maneuver. When coupled with the actual weight distribution, the average Movement correction was 3% without PPE and 1% with PPE. The average Maneuver correction was 4% without PPE and 2% with PPE. The average Lift correction was 9% without PPE and 10% with PPE. In each case, the average empirical correction was less than the average of 13% derived from theory.

CFT Score Bias

The aggregate CFT score was not biased even though the components were biased. Weight and age had little effect on CFT scores. The weight–CFT correlation was trivial for all groups non-PPE men, $r = .008$, $n = 1614$, $p = .756$, and non-PPE women, $r = .094$, $n = 194$, $p = .194$. The age–CFT correlation was small for non-PPE men, $r = -.166$, $n = 1610$, $p < .001$, and for non-PPE women, $r = .113$, $n = 194$, $p = .116$.

It is not clear whether allometric arguments apply to the CFT score. This score is not directly related to physiological processes whereas the actual performances on the CFT components. However, an allometric analysis was carried out to verify that the bivariate correlations of weight with CFT score did not conceal a nonlinear relationship corresponding to the allometric model. The analyses employed the general allometric model in its nonlinear form because the results obtained with the logarithm of the CFT score as the dependent variable would be harder to interpret than the results obtained with the actual score as the dependent variable. The mass coefficient for non-PPE men was $b =$

.012 with a 95% confidence interval, [-.017, .040], that included $b = .000$. The mass coefficient for non-PPE women was $b = .072$ with a 95% confidence interval, [-.029, .172], that included $b = .000$.

CFT-PFT Comparison

Although it was not the primary focus of this study, the data provided the opportunity to compare CFT scores with PFT scores. It is not necessary to have two tests if they measure the same performance capabilities. The evidence indicated that the tests were not equivalent. If both tests measured essentially the same construct, each test was moderately reliable (i.e.., $\alpha = .85$), and the tests were administered at approximately the same time, the correlation of CFT scores with PFT scores would be $r \approx .80$. The observed correlations were well below this expectation (Table 11). The correlation for non-PPE men was significantly greater than the correlation for non-PPE women, $z = 3.01$, $p < .002$, while PPE status did not affect the correlation among men, $z = .17$, $p > .430$.

Discussion

An apparent contradiction was the most important finding in this study. CFT scores were not related to weight, even though mass bias was evident for every CFT component. This discussion considers the reasons for the apparent contradiction and the implications for CFT applications.

The contradiction involves a conflict between competing fitness concepts. When considered from a physical fitness perspective, fitness is defined by pound-for-pound performance. When considered from a fitness for duty perspective, fitness focuses on the ability to perform combat tasks with no consideration of size.

Which fitness perspective is more appropriate? The reason for developing and implementing the CFT is the critical consideration in answering this question. The CFT was developed to assess the ability to perform combat tasks. An analysis of combat tasks and their importance for unit effectiveness guided the CFT design (Vickers & Hodgdon, 2000). The individual CFT components were selected to require the physical abilities that would be required for effective performance of the most common, important, physically-demanding combat tasks. The names given to the CFT components indicate that they simulate specific types of combat task. Given these considerations, the fitness-for-duty perspective provides the appropriate framework for assessing the CFT. In this view, the CFT does not display size bias.

The implications of adopting the fitness-for-duty perspective merit comment. Consider what could happen if scoring were corrected to adjust for size bias. Heavier individuals might reduce the intensity of their physical training because they now could receive acceptable test scores despite poorer absolute performance. A combat unit’s performance would be limited to the least common denominator set by those relatively unfit individuals. For example, movement to contact can only be as fast as the slowest member if the unit is to arrive intact. Requiring that all individuals meet minimum standards on a test that reflects the actual performance potential provides a useful basis for gauging for unit combat readiness. This test attribute would be lost if scores were individualized by correcting for weight differences.

The fitness-for-duty perspective also has implications for individual Marines. Test scores are part of the input for promotion decisions. A biased test would adversely affect promotion opportunities for some individuals. However, a bona fide link between the test standards and job performance meets the legal standards for avoiding bias (Thompson & Thompson, 1982). The CFT arguably meets this requirement. Beyond this, the study findings provide no evidence that the CFT introduces mass bias. Individual Marines can and do achieve satisfactory performance regardless of weight. There would be no adverse effect on promotion and other opportunities.

The effects of wearing PPE provide further evidence that the CFT scores reflect bona fide occupational requirements. Wearing PPE virtually eliminated the Movement and Maneuver biases and slightly reduced the Lift bias. These effects paralleled Vanderburgh and Flanagan's (2000) report that carrying a backpack reduced body mass bias in a run test. Marines wear PPE in combat, so performance in the gear is closely related to combat requirements. The PPE data were obtained because test designers considered having all Marines wear PPE for the CFT. This course of action could not be implemented because the CFT must be administered to all Marines. PPE is expensive, so its availability is limited to units that need it for everyday operations.

A performance perspective also explains why CFT scores were unbiased despite biases in the test components. Lighter individuals performed better on Movement and Maneuver. Heavier individuals performed better on Lift. When combined, these performance differences produced overall test scores that were independent of weight. This result calls to mind Vanderburgh and Crowder's (2006, p. 756) observation that a fair test can be constructed by combining tests that cancel out body mass penalties. In this case, however, combining tests cancels out performance differences even though all of the component tests are biased in favor of lighter individuals.

It might be argued that the case against bias is based on chance findings. The difference between the observed mass coefficients and theory could indicate that the findings represented were chance results. This argument can be countered by noting several points. The Maneuver and Movement coefficients were consistent with previous empirical estimates for heavy exertion (Jones & Lindstedt, 1991). The competing theories for mass bias include at least one theory that can account for the observed findings (Darveau, Suarez, Andrews, & Hocachka, 2002). Jaric et al.'s (2005) arguments that mass bias varies from one physical activity to another raises the possibility that theory might account for the evidence with proper allowance for the specific activities involved. The degree of correspondence between that theoretical perspective and the current findings cannot be judged with precision because the CFT elements do not correspond precisely to the activity categories in the theoretical model. This point applies with particular force to Maneuver, which is a combination of a number of different activities. Bishop et al.'s (2008) obstacle course findings provide reason to expect minimal bias on Maneuver. Finally, the large sample sizes in this study decrease the plausibility of invoking chance sampling variation to account for differences between the observed mass coefficients and theoretical predictions. On the whole, there is no reason to view the study findings as aberrant in the context of available allometric research.

The impact of an incorrect conclusion must be considered. How would test results be affected if score corrections were introduced to correct for the observed element-by-element bias? Most effects would be minimal because the correction produces large differences only when an individual's weight differs substantially from the reference value. It is also important that few Marines weigh enough to differ markedly from the reference value. For example, the bias for Marines who are 80% heavier than the reference value was ~12%. This bias would apply to men who weighed 227 lb. Only 4.2% of the sample reached or exceeded that weight. For Marines in the 25th to 75th percentiles of the weight distribution, the bias was between 4% and 8%. The average bias was 3% for Movement, 4% for Maneuver, and 9% for Lift. The relatively modest magnitude of these biases can be seen by noting that theory predicted an average 13% bias for each test. The empirical body mass bias is not only less than the theoretical bias, it is modest ifor most Marines. Even small biases are important for individuals near the cutoff for categorical classifications (e.g., failure, first class). However, even in this context, the error introduced by mass bias might be small relative to the random errors resulting from imperfect test reliability.

Whatever its strengths and weaknesses, the CFT will perform comparably for all Marines. Age and gender had only trivial effects on the mass coefficients. The coefficient variation across groups was statistically significant, but the samples were large. Treating the various age and gender groups as equivalent was within the bounds of standard statistical practice.

In summary, the mass biases evident in performance on individual CFT elements did not invalidate the test. CFT scores arguably provide appropriate indications of fitness-for-duty, defined as the ability to perform combat tasks. CFT scores should not be a practical barrier to promotion because test scores were not related to weight. The results are at odds with some theoretical predictions, but the models that yield those predictions are the subject of continuing debate in the allometric research literature. It is more important to recognize that the current findings are consistent with the results of other allometric studies of similar performance tests. The overall conclusion is that mass bias is present, but practically unimportant. The evidence supports the view that the CFT is an unbiased indicator of the performance capacities it was designed to assess.

References

Agutter, P. S., & Wheatley, D. N. (2004). Metabolic scaling: Consensus or controversy? *Theoretical Biology and Medical Modeling, 1*(Nov 16), 13.

Batterham, A. M., & Jackson, A. S. (2003). Validity of the allometric cascade model at submaximal and maximal metabolic rates in exercising men. *Respiratory Physiology & Neurobiology, 135*, 103–106.

Batterham, A. M., Tolfrey, K., & George, K. P. (1997). Nevill's explanation of Kleiber's 0.75 mass exponent: An artifact of collinearity problems in the least squares model. *Journal of Applied Physiology, 82*, 693–697.

Batterham, A. M., Vandenburgh, P. M., Mahar, M. T., & Jackson, A. S. (1999). Modeling the influence of body size on VO2peak: Effects of model choice and body composition. *Journal of Applied Physiology, 87*, 1317–1325.

Bergh, U., Sjodin, G., Forsberg, A., & Svedenhag. (1991). The relationship between body mass and oxygen uptake during running in humans. *Medicine and Science in Sports and Exercise, 23*, 205–211.

Bishop, P. A., Crowder, T. A., Fielitz, L. R., Lindsay, T. R., & Woods, A. K. (2008). Impact of body weight on performance of a weight-supported motor fitness test in men. *Military Medicine, 173*, 1108–1114.

Chamari, K., Moussa-Chamari, I., Boussaidi, L., Hachanga, Y., Kaouech, F., & Wisloff, U. (2005). Appropriate interpretation of aerobic capacity: Allometric scaling in adult and young soccer players. *British Journal of Sports Medicine, 39*, 97–101.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Commandant of the Marine Corps. (2008). *Marine Corps Physical Fitness Program* (MARCORPS Order 6100.13, 1 Aug). Washington, DC: US Marine Corps Headquarters.

Darveau, C.–A., Suarez, R. K., Andrews, R. D., & Hochachka, P. W. (2002). Allometric cascade as a unifying principle of body mass effects on metabolism. *Nature, 417*(9 May), 166–170.

Folland, J. P., McCauley, T. M., & Williams, A. G. (2008). Allometric scaling of strength measurements to body size. *European Journal of Applied Physiology, 102*, 739–745.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Review*, *41*, 587–638.

Jaric, S., Mirkov, D., & Markovic, G. (2005). Normalizing physical performance tests for body size: A proposal for standardization. *Journal of Strength and Conditioning Research, 19*, 467–474.

Jensen, K., Johansen, L., & Secher, N. H. (2001). Influence of body mass on maximal oxygen uptake: Effect of sample size. *European Journal of Applied Physiology, 84*, 201–205.

Jones, J. H., & Lindstedt, S. L. (1991). Limits to maximal performance. *Annual Review of Physiology, 55*, 547–569.

Kusano, M. A., Vandenburgh, P. M., & Bishop, P. (1997). Impact of body size on women's military obstacle course performance. *Biomedical Sciences Instrumentation, 34*, 357–362.

Markovic, G., & Jaric, S. (2004). Movement performance and body size: The relationship for different groups of tests. *European Journal of Applied Physiology, 92*, 139–140.

Markovic, G., Vucetic, V., & Nevill, A. M. (2007). Scaling behaviour of VO2 in athletes and untrained individuals. *Annals of Human Biology, 34*, 315–328.

McArdle, W. D., Katch, F. I., & Katch, V. L. (2001). *Exercise physiology: Energy, nutrition, and human performance* (5th ed.). Philadelphia: Lippincott Williams and Wilkins.

Nevill, A. M., Stewart, A. D., Olds, T., & Holder, R. (2004). Are adult physiques geometrically similar? The dangers of allometric scaling using body mass power laws. *American Journal of Physiological Anthropology, 124*, 177–182.

Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research.* New York: McGraw-Hill.

Savage, V. M., Deeds, E. J., & Fontana, W. (2008). Sizing up allometric theory. *PLoS Computers in Biology, 4*, e1000171.

Suarez, R. K., & Darveau, C. A. (2005). Multi-level regulation and metabolic scaling. *Journal of Experimental Biology, 208*, 1627–1634.

Thompson, D. E., & Thompson, T. A. (1982). Court standards for job analysis in test validation. *Personnel Psychology, 35*, 865–874.

Vanderburgh, P. M. (2007). Correction factors for body mass in military physical fitness tests. *Military Medicine*, *172*, 738-742.

Vanderburgh, P. M., & Crowder, T. A. (2006). Body mass penalties in the physical fitness tests of the Army, Air Force, and Navy. *Military Medicine, 171*, 753–756.

Vanderburgh, P. M., & Flanagan, S. (2000). The Backpack Run Test: A model for a fair and occupationally relevant military fitness test. *Military Medicine, 166*, 418–421.

Vickers, R. R., Jr., & Hodgdon, J. A. (2000). *A summary of findings for initial task identification for military operations in urban terrain (MOUT)* (Technical Report No. 00-04). San Diego, CA: Naval Health Research Center.

Weir, J. P., Housh, T. J., Johnson, G. O., Housh, D. J., & Ebersole, K. T. (1999). Allometric scaling of isokinetic peak torque: The Nebraska Wrestling Study. *European Journal of Applied Physiology, 80*, 240–248.

White, C. R., & Seymour, R. S. (2005). Allometric scaling of mammalian metabolism. *Journal of Experimental Biology*, *208*, 1611-1619.

APPENDIX A

Site Effects

Weather and other conditions varied from site to site. This variation was desirable because the CFT will be performed in different settings and under different conditions once implemented. Site effects were of interest for the present efforts because large differences could obscure the relationship of size with test results.

Results for Non-PPE Men

Variation in average values. All variables differed significantly between test sites, but the variation was not important. Sample size was the primary reason for the statistical significance of the site differences. Site differences typically explained only a small proportion of the variance: weight, $\varepsilon^2 = .057$; PFT, $\varepsilon^2 = .049$; 3-mi run, $\varepsilon^2 = .126$; Movement, $\varepsilon^2 = .105$; Lift, $\varepsilon^2 = .073$; Maneuver, $\varepsilon^2 = .141$; age, $\varepsilon^2 = .486$. Because each analysis involved 15 or 16 groups, the variance explained by site differences translated into small to moderate ES, except for the large effect for age (Cohen, 1988, pp. 273–288).

Variation in allometric coefficients. ANCOVA tested for site effects on the allometric model. The ANCOVA model included the group by covariate interaction to test for site differences in the mass coefficient. For males, the Site x ln(W) interaction was statistically significant for the 3-mi run, $F_{14,1732} = 2.32$, $p = .004$, Movement, $F_{14,1847} = 2.05$, $p = .012$, and Maneuver, $F_{13,1727} = 2.88$, $p < .001$. The interaction approached significance for Lift, $F_{13,1761} = 1.64$, $p = .067$.

The site interactions were trivial even though statistically significant. The site-to-site variation in the mass coefficient accounted for at most 1.1% of test score variance: run, $\varepsilon^2 = .008$; Movement, $\varepsilon^2 = .006$; Lift, $\varepsilon^2 = .004$; Maneuver, $\varepsilon^2 = .011$. Allowing for the number of groups, Cohen (1988) would classify all of these differences as too small to be of theoretical or practical importance. Thus, the large sample size was the primary basis for the statistical significance of site differences in performance.

Results for Non-PPE Women

The analyses for non-PPE women were limited to testing sites with ≥ 9 study participants. Analyses contrasted eight or nine sites for each variable.

Variation in average values. All variables differed significantly between test sites, but the variation was not important. Once again, sample size was the primary reason that the site differences were statistically significant. The typical site difference explained only a small proportion of the variance: weight, $\varepsilon^2 = .059$; PFT, $\varepsilon^2 = .099$; 3-mi Run, $\varepsilon^2 = .161$; Movement, $\varepsilon^2 = .083$; Lift, $\varepsilon^2 = .209$; Maneuver, $\varepsilon^2 = .159$. Age, $\varepsilon^2 = .413$. Because each analysis involved eight or nine groups, the site differences translated into small to moderate ES, except for the large age ES.

Variation in allometric coefficients. No tests for variation in allometric coefficients were conducted for women because the sample sizes at most sites were too small for meaningful comparisons.

APPENDIX B

Comparing Mathematical Variants of the Allometric Model

Linear and nonlinear methods of estimating allometric coefficients were compared. The nonlinear regression fitted the model $y = AM^{b}$. The linear method fitted the model $\ln(y) = \ln(A) + b * [\ln(x)]$. Although these models are mathematically equivalent, they could produce different analytic results because the metric for prediction errors differed. The errors in the first model would be the difference between the raw score and the predicted raw score. The errors in the second model would be the difference between the natural logarithm of the raw score and the predicted value of that natural logarithm. The model coefficients could differ because the two analyses minimized different errors.

The model variant choice had little effect on the mass coefficient estimates. The nonlinear estimate was slightly smaller than the linear estimate in five analyses, the two estimates were equal in two analyses, and the nonlinear estimate was larger than the nonlinear in one analysis (see Table B1).

Predictive accuracy is another comparison criterion for the model variants. This comparison would only be meaningful if the dependent variable was expressed in the original units of measurement. An exponential transformation converted the predicted values from the linear model to predicted raw score values. The difference between the predicted raw score and the observed raw score was analyzed for comparison to the errors from the nonlinear model. The nonlinear model errors did not have to be transformed because that model was fitted to the raw data with the difference between the predicted

Table B1
Comparison of Allometric Coefficients

	Model constant				Mass coefficient			
	Nonlinear		Linear		Nonlinear		Linear	
	A	SE	*A*	SE	*b*	SE	*b*	SE
Men								
Lift	4.08	.86	1.27	.21	.53	.04	.55	.04
Move	52.31	5.01	4.00	.10	.26	.02	.25	.02
Man	22.03	3.98	3.37	.16	.41	.04	.35	.03
Run	478.73	40.70	6.20	.08	.20	.02	.19	.02
Women								
Lift	.22	.20	-1.57	.87	1.02	.18	1.02	.18
Move	121.00	34.68	4.82	.28	.14	.06	.14	.06
Man	194.02	108.97	5.30	.06	.05	.11	.03	.10
Run	785.26	224.09	6.64	.29	.13	.06	.14	.06

Note. The nonlinear model was $y = AM^b$; the linear model was $ln(y) = ln(A) = b*ln(M)$.

and observed score as the error. The two models provided equally accurate predictions. None of the differences in Table B2 were large enough to be of practical importance.

The linear variant was used in primary analyses reported in the body of this paper. This variant made it possible to treat group comparisons as ANCOVA problems. The ANCOVA test for parallelism of regression lines was an essential part of gender and age group comparisons.

Table B2

Predictive Accuracy of Nonlinear and Linear Models

	Variance explained			
	Males		Females	
	Nonlinear	Linear	Nonlinear	Linear
Lift	.091	.090	.134	.137
Move	.092	.102	.021	.023
Man	.070	.090	.001	.001
Run	.076	.073	.024	.023

Note. The variance explained for the allometric model was computed for the nonlinear expression of the model. This approach employed the same dependent variable and goodness of fit criterion as the simple linear regression.

APPENDIX C

VO_2 Mass Coefficients Estimated from Run Times

Mass coefficients were computed for oxygen uptake rates estimated from run times. These coefficient estimates provided a basis for comparing the present findings with findings from earlier oxygen uptake rates.

The energy required to run a given distance is the same regardless of how rapidly the run is completed. The energy cost is ~1 $kcal \cdot kg^{-1} \cdot km^{-1}$ (McArdle, Katch, & Katch, 1991, p. 182), so the total energy (E) utilized when a person of mass (M) runs distance (d) is:

$$E = d*M. \tag{C1}$$

If the individual runs at or near his or her maximum sustainable rate, the respiratory quotient (RQ) will be $RQ \approx 1.00$ and 1 L of oxygen will provide ~5 kcal of energy. The total oxygen requirement (O_R) for the run will be:

$$O_R = E/5 = 0.2*(d*M). \tag{C2}$$

Except for the constant, the time in minutes, t, required to complete the run determines the oxygen uptake rate in L per minute (VO2L):

$$VO2L = O_R/t = (0.2*d*M)/t. \tag{C3}$$

The allometric model provides an alternative VO2L formulation:

$$\text{VO2L} = A\text{M}^{b}. \qquad \text{(C4)}$$

Combining Equations C3 and C4 gives:

$$\text{VO2L} = (0.2*\text{d}*\text{M})/\text{t} = A\text{M}^{b}. \qquad \text{(C5)}$$

Solving Equation C5 for t gives:

$$\begin{aligned} \text{t} &= (0.2*\text{d}*\text{M})/A\text{M}^{b} \\ &= (0.2*\text{d}/A)*(\text{M}/\text{M}^{b}) \\ &= (0.2*\text{d}/A)*\text{M}^{(1-b)}. \qquad \text{(C6)} \end{aligned}$$

Equation C6 is the allometric equation for run time. Both the scaling constant and the mass coefficient have simple relationships to the corresponding energy consumption coefficients. The run time scaling constant is (0.2*d)/*A*, but the mass coefficient is the more important value. The run time mass coefficient, $(1 - b)$, is 1 minus the mass coefficient for the associated metabolic rate.

The mathematical relationship of the metabolic rate and run time mass coefficients leads to theoretical predictions for the run time mass coefficient. A theory that predicts $b = .67$ for metabolic rate will predict $b = .33$ for run time. A theory that predicts $b = .75$ for metabolic rate will predict $b = .25$ for run time. This relationship also applies to empirically derived coefficients. Thus, previous research indicating that *b* = .87 for maximal oxygen uptake implies *b* = .13 for run times. Combining theory and empirical evidence, the plausible range for run time mass coefficients was *b* = .13 to *b* = .33. The empirical coefficients in this study fell near the lower end of this range.

The same point can be made by converting run times to metabolic rate estimates. The American College of Sports Medicine's Web site, www.acsm.org, gives $\dot{V}_{O_2} = 0.2v + 3.5$ as the equation for converting running velocity, *v*, to an oxygen uptake rate when running over flat ground. This equation was applied to the Movement and 3-mi run times after converting times to velocity (in meters).

The estimated metabolic rate mass coefficients were consistent with an oxygen uptake generated by moderate to high level of exertion during the two runs. Prior research suggests that the metabolic rate mass coefficient increases from *b* = .67 at rest to *b* = .87 at maximal exertion. The univariate mass coefficients in Table C1 were obtained from separate analyses with weight as the only predictor. The multivariate coefficients were obtained from analyses with age, gender, and weight as predictors.

Table C1

Allometric Model for Metabolic Rate Estimates

	Univariate model				Multivariate model	
	Men		Women			
	b	*SE*	*b*	*SE*	*b*	*SE*
Non-PPE						
3-mi run	.808	.016	.863	.057	.834	.016
Movement	.748	.018	.864	.057	.809	.018
PPE						
3-mi run	.763	.044			.759	.044

Analyses showed that a single coefficient was appropriate for men and women. The multivariate coefficients were consistent with expectation (see Table C1). The 3-mi run, *b* = .834, SE = .016, *t* = 51.89, 95% CI = [.803, .866]; Movement, *b* = .809, SE = .018, *t* = 46.16, 95% CI = [.775, .844]. The overlapping confidence intervals indicated that the two coefficients were not significantly different.

The 3-mi run coefficients could be compared for non-PPE and PPE groups because both groups completed the run without PPE. The coefficient for PPE men was slightly lower than that for non-PPE men, but the 95% confidence intervals overlapped substantially: Non-PPE, CI = [.776, .840]; PPE, CI = [.675, .851]. This overlap indicated that the mass coefficient for men replicated for the two male samples.

Table 1

Descriptive Statistics

				Correlation with:		
	N	*M*	*SD*	Age	Weight	PFT
Non-PPE men						
Age	1883	27.24	8.33			
Weight	1879	182.40	25.04	.298		
PFT	1843	245.24	31.80	-.124	-.280	
3-mi run[a]	1763	22:23	2:12	.284	.284	-.629
Non-PPE women						
Age	278	25.44	6.55			
Weight	277	144.17	17.71	.019		
PFT	269	257.64	29.09	-.063	-.232	
3-mi run[a]	235	25:11	2:45	.053	.149	-.728
PPE Men						
Age	235	26.12	5.42			
Weight	237	192.97	25.14	.141		
PFT	236	251.51	25.58	.154	-.287	
3-mi run[a]	220	22:38	2:07	-.120	.351	-.724

Note. PPE indicates whether the test was performed while wearing PPE.

[a] 3-mi run was scored as total seconds when computing correlations.

Table 2

Age-Gender Distribution for Non-PPE Participants

Age group	Men	Women
17–26	1252	179
27–39	638	98
40–45	157	16
46+	71	2

Note. The distributions differed significantly ($\chi^2 = 8.47$, 3 *df*, $p < .037$) primarily because women were underrepresented in the 46+ age group.

Table 3

CFT Descriptive Statistics

	N	*M*	*SD*	Movement	Lift
Non-PPE men					
Movement	1881	3:24	0:23		
Lift	1793	64.34	15.84	-.187	
Maneuver	1759	3:04	0:38	.555	-.220
Non-PPE women					
Movement	277	4:04	0:29		
Lift	230	35.81	12.22	-.388	
Maneuver	254	4:02	0:53	.602	-.502
PPE men					
Movement	239	3:35	0:27		
Lift	239	60.62	15.32	-.093	
Maneuver	238	3:30	0:47	.572	-.261

Note. The Movement and Maneuver scores were expressed as total seconds in the analyses. The mean and standard deviation have been converted to min:s in the table.

Table 4

Predictor-Performance Correlations

	Men			Women		
	Movement	Lift	Maneuver	Movement	Lift	Maneuver
Non-PPE						
Age	.366	.073	.367	.135	-.106	.161
Weight	.317	.298	.273	.153	.366	.030
PFT	-.525	.321	-.411	-.570	.282	-.316
3-mi run	.641	-.176	.444	.713	-.296	.470
PPE						
Age	.121	.214	.034	-[a]	-[a]	-[a]
Weight	.087	.223	.046	-[a]	-[a]	-[a]
PFT	-.363	.219	-.341	-[a]	-[a]	-[a]
3-mi run	.365	-.138	.321	-[a]	-[a]	-[a]

Note. The cells for women who wore PPE are blank because too few women completed the CFT in PPE to provide reliable estimates of the correlations.

[a] Women wearing PPE were excluded from the analysis because the sample ($n = 17$) was too small to produce reliable findings.

Table 5

Allometric Model Parameters for Movement

					95% Confidence interval		
Parameter	Est	*SE*	*t*	Sig	Lower bound	Upper bound	Partial ε^2
Intercept (A)	4.629	.089	51.93	.000	4.454	4.803	.557
Gender (B)							
Men	-.227	.008	-28.35	.000	-.243	-.211	.272
Women	0[a]						
Age group (C)							
17–26	-.101	.013	-7.69	.000	-.127	-.075	.027
27–39	-.069	.013	-5.17	.000	-.095	-.043	.012
40–45	-.021	.015	-1.40	.160	-.051	.008	.001
46+	0[a]						
Weight (W)							
Mass coefficient (*b*)	.191	.018	10.88	.000	.156	.225	.052

[a] This parameter was set to zero because the indicated group was the reference group to which other Gender or Age groups were compared.

Table 6

Allometric Model for Lift

					95% Confidence interval		
Parameter	Est	*SE*	*t*	Sig	Lower bound	Upper bound	Partial ε^2
Intercept (A)	.490	.214	2.29	.022	.070	.909	.003
Gender (B)							
Men	.476	.020	23.91	.000	.437	.515	.221
Women	0[a]	.	.	.	.	.	.
Age group (C)							
19–26	.006	.031	.21	.833	-.054	.066	.000
27–39	.014	.031	.46	.647	-.047	.076	.000
40–45	-.064	.035	-1.82	.068	-.134	.005	.002
46+	.0[a]	.	.	.	.	.	.
Weight (W)							
Mass coefficient (*b*)	.609	.042	14.51	.000	.527	.691	.095

[a] This parameter was set to zero because the indicated group was the reference group to which other gender or age groups were compared.

Table 7

Allometric Model for Maneuver

Parameter	*b*	*SE*	*t*	Sig	95% Confidence interval Lower bound	Upper bound	Partial ε^2
***Intercept* (*A*)**	4.542	.149	30.47	.000	4.250	4.834	.317
Gender (B)							
Men	-.328	.014	-24.30	.000	-.355	-.302	.228
Women	0[a]	.	.	.	.	.	.
Age group (C)							
19–26	-.202	.021	-9.60	.000	-.243	-.161	.044
27–39	-.136	.021	-6.33	.000	-.178	-.094	.020
40–45	-.065	.024	-2.68	.008	-.112	-.017	.004
46+	0[a]	.	.	.	.	.	.
Weight (W)							
Mass coefficient (*b*)	.221	.029	7.51	.000	.163	.278	.027

[a] This parameter was set to zero because the indicated group was the reference group to which other Gender or Age groups were compared.

Table 8

Comparison of PPE and Non-PPE Samples of Men

	Non-PPE men			PPE men					
	N	*M*	*SD*	*N*	*M*	*SD*	ES	t^a	Sig
Background									
Age	1883	27.24	8.33	235	26.12	5.42	-.13	2.79	.006
Weight	1879	182.40	25.04	237	192.97	25.14	.42	-6.13	.000
PFT	1843	245.24	31.80	236	251.51	25.58	.20	-3.44	.001
3-mi run	1763	22:23	2:12	220	22:38	2:07	.12	-1.64	.101
CFT									
Movement	1881	3:24	0:23	239	3:35	0:27	.47	-6.04	.000
Lift	1793	64.34	15.84	239	60.62	15.32	-.23	3.43	.001
Maneuver	1759	3:04	0:38	238	3:30	0:47	.68	-8.05	.000

[a] ES = $(M_{\text{non-PPE}} - M_{\text{PPE}})/SD_{\text{non-PPE}}$.

Table 9

Allometric Models for PPE Men

	PPE *b*	*SE*	t^a	Sig^a	Lower Bound	Upper Bound	Non-PPE *b*	t^b	Sig^b
Movement	.052	.063	0.82	.412	-.072	.176	.191	2.19	.015
Lift	.413	.129	3.20	.002	.159	.667	.609	1.97	.025
Maneuver	.096	.097	.99	.323	-.095	.288	.221	2.28	.012

[a] Test for $b_{PPE} = 0$. [b] Test for $b_{PPE} \neq b_{Non\text{-}PPE}$

Table 10

Bias Distributions

		Non-PPE Men			Non-PPE Women			PPE Men		
	Theory[a]	Move	Lift	Man	Move	Lift	Man	Move	Lift	Man
Descriptives										
N	2410	1879	1879	1879	277	277	277	237	237	237
Mean	1.13	1.03	1.09	1.04	1.03	1.09	1.04	1.01	1.10	1.02
Std Dev	.10	.02	.07	.03	.02	.06	.03	.01	.06	.01
Normality Test										
K-S Z[b]	--[c]	.90	1.03	.87	.66	.63	.65	.74	.82	.73
Sig	--[c]	.394	.244	.431	.778	.823	.797	.643	.507	.727
Distribution										
Minimum[d]	.81	.94	.86	.93	.95	.88	.94	.99	.91	.98
Percentiles										
1%	.90	.97	.93	.97	.98	.94	.97	.99	.94	.99
5%	.97	.99	.97	.99	.99	.98	.99	1.00	1.00	1.00
10%	1.00	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.02	1.01
25%	1.06	1.02	1.04	1.02	1.02	1.04	1.02	1.01	1.06	1.01
50%	1.13	1.03	1.09	1.04	1.03	1.08	1.04	1.01	1.10	1.02
75%	1.20	1.05	1.14	1.06	1.05	1.13	1.06	1.02	1.14	1.03
90%	1.27	1.07	1.18	1.08	1.07	1.17	1.08	1.02	1.18	1.04
95%	1.30	1.08	1.21	1.09	1.07	1.19	1.08	1.02	1.20	1.04

		Non-PPE Men			Non-PPE Women			PPE Men		
	Theory[a]	Move	Lift	Man	Move	Lift	Man	Move	Lift	Man
99%	1.38	1.09	1.25	1.11	1.08	1.22	1.10	1.03	1.24	1.05
Maximum[d]	1.51	1.12	1.34	1.14	1.10	1.26	1.11	1.03	1.25	1.05

Note. The column headings indicate the CFT components of movement to contact (Move), ammunition can lift (Lift), and maneuver under fire (Man).

[a] The theoretical estimates were based on $b = .67$, the mass coefficient used by Vanderbergh (2007).

[b] Kolmogorov-Smirnoff test for a normal distribution. [c] The Kolmogorov-Smirnoff test was $Z = 1.32$ ($p = .061$) for men and $Z = .66$ ($p = .783$) for women. [d] The minimum and maximum values were the bias estimates for the lightest and heaviest Marine, respectively.

Table 11

CFT–PFT Comparison

		PFT		CFT				
	N	*M*	*SD*	*M*	*SD*	*r*	*t*	Sig
Non-PPE								
Men	1845	245.24	31.80	224.12	37.67	.425	24.23	.000
Women	269	257.64	29.09	215.75	41.51	.251	15.50	.000
PPE								
Men	236	251.51	25.58	230.05	32.71	.415	10.23	.000

Note. Statistics may differ from those reported earlier. These statistics describe individuals who had both PFT and CFT scores.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD MM YY)
07 02 11

2. REPORT TYPE
Technical Report

3. DATES COVERED (from – to)
Jan 2009 – Dec 2010

4. TITLE
Body Mass Bias in a Combat Fitness Test

5a. Contract Number:
5b. Grant Number:
5c. Program Element Number:
5d. Project Number:
5e. Task Number:
5f. Work Unit Number: 60704

6. AUTHORS
Vickers, Ross R., Jr.; Reynolds, John H.; McGuire, Brian

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Commanding Officer
Naval Health Research Center
140 Sylvester Rd
San Diego, CA 92106-3521

8. PERFORMING ORGANIZATION REPORT NUMBER
Report No. 11-20

8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Commanding Officer
Naval Medical Research Center
503 Robert Grant Ave
Silver Spring, MD 20910-7500

Chief, Bureau of Medicine and Surgery
(MED 00), Navy Dept
2300 E Street NW
Washington, DC 20372-5300

10. SPONSOR/MONITOR'S ACRONYM(S)
NMRC/BUMED

11. SPONSOR/MONITOR'S REPORT NUMBER(s)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Allometric theory predicts that, pound for pound, lighter individuals will perform better than heavier individuals on strength and endurance tests. This study evaluated body mass bias as a factor in the U.S. Marine Corps Combat Fitness Test (CFT), a test consisting of movement to contact (Movement), an ammunition can lift (Lift), and maneuver under fire (Maneuver). Allometric modeling indicated small to moderate biases, favoring lighter Marines for each CFT element. The biases were smaller than predicted by theory. Age and gender did not modify the bias. Wearing personal protective equipment eliminated the biases for the movement to contact and maneuver under fire. Despite the biases for individual elements, CFT scores were not related to weight. This difference reflects the fact that CFT scores are based on absolute performance. The better performance of heavier individuals on Lift offset their poorer performance on Movement and Maneuver. The CFT elements are biased measures of physical fitness, but the CFT is an unbiased measure of fitness for duty. These apparently contradictory conclusions derive from the difference between fitness defined as relative pound-for-pound performance and the absolute performance required by combat tasks. Also, the fact that CFT scores did not correlate with weight means the CFT will not adversely affect personnel decisions.

15. SUBJECT TERMS
Combat Fitness Test, mass bias, allometrics

16. SECURITY CLASSIFICATION OF:

a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
UNCL	UNCL	UNCL	UNCL	44

18a. NAME OF RESPONSIBLE PERSON
Commanding Officer

18b. TELEPHONE NUMBER (INCLUDING AREA CODE)
COMM/DSN: (619) 553-8429

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18