OEM Online First, published on November 3, 2016 as 10.1136/oemed-2016-103684

Fitness tests and occupational tasks of military interest: a systematic review of correlations

Veronique D Hauschild, David W DeGroot, Shane M Hall, Tyson L Grier, Karen D Deaver, Keith G Hauret, Bruce H Jones

Injury Prevention Program, Army Public Health Center (Provisional), Aberdeen Proving Ground, Maryland, USA

Correspondence to
Veronique D Hauschild, Injury Prevention Program, Army Public Health Center (Provisional), 5158 Blackhawk Rd, Aberdeen Proving Ground, MD 21010-5400, USA; veronique.d.hauschild.civ@mail.mil

Received 8 March 2016
Revised 16 August 2016
Accepted 11 October 2016

ABSTRACT
Physically demanding occupations (ie, military, firefighter, law enforcement) often use fitness tests for job selection or retention. Despite numerous individual studies, the relationship of these tests to job performance is not always clear.
This review examined the relationship by aggregating previously reported correlations between different fitness tests and common occupational tasks.
Search criteria were applied to PUBMED, EBSCO, EMBASE and military sources; scoring yielded 27 original studies providing 533 Pearson correlation coefficients (r) between fitness tests and 12 common physical job task categories. Fitness tests were grouped into predominant health-related fitness components and body regions: cardiorespiratory endurance (CRe); upper body, lower body and trunk muscular strength and muscular endurance (UBs, LBs, TRs, UBe, LBe, TRe) and flexibility (FLX). Meta-analyses provided pooled r's between each fitness component and task category.
The CRe tests had the strongest pooled correlations with most tasks (eight pooled r values 0.80–0.52). Next were LBs (six pooled r values >0.50) and UBe (four pooled r values >0.50). UBs and LBe correlated strongly to three tasks. TRs, TRe and FLX did not strongly correlate to tasks.
Employers can maximise the relevancy of assessing workforce health by using fitness tests with strong correlations between fitness components and job performance, especially those that are also indicators for injury risk. Potentially useful field-expedient tests include timed-runs (CRe), jump tests (LBs) and push-ups (UBe). Impacts of gender and physiological characteristics (eg, lean body mass) should be considered in future study and when implementing tests.

What this paper adds

- Military, firefighting and law enforcement employees are often required to perform physical fitness tests for job selection, placement and/or retention.
- Studies regarding the relationship between common fitness tests and physical job performance have been inconsistent.
- Aggregated correlation data from applicable studies can be used to describe a relationship between core fitness components and 12 common task categories relevant to military and other physically demanding occupations.
- Fitness tests that measure cardiorespiratory endurance, lower body strength and upper body muscular endurance are particularly relevant when assessing employees' health status for these jobs.

INTRODUCTION
Occupations such as the military, firefighting and law enforcement require employees to perform vigorous, physically demanding tasks such as dragging victims to safety, moving quickly and carrying heavy loads. Personnel are often required to demonstrate specified levels of physical capability for job selection, placement and/or retention.[1–8] There are two primary types of physical capability tests. One type, known as, a 'criterion task' or 'content-based' performance test uses standardised job-task simulations, such as lifting and carrying a mannequin a specific distance.[6 9] The second type of test measures general physical 'constructs' or 'components' that are important to successful job performance, such as cardiovascular endurance and muscular strength.[2 9 10] Though criterion task tests can provide stronger associations (predictive value) with performance outcomes than component-based fitness tests,[9] they can be logistically complex and have limited applications since they are designed and tested to reflect specific scenarios, tasks and equipment. Component fitness tests are generally cheaper and more broadly applicable, but they may be challenged if the relationship between them and job performance has not been adequately demonstrated.[11 12]

As an example, the US Army Physical Fitness Test (APFT) is a three-test battery consisting of a 2 min push-up test, a 2 min sit-up test and a two mile run for time.[2] A passing score on the APFT is used as a key determinant for US Army service entry and retention to ensure a base level of physical fitness for every soldier.[2] Instituted in 1980, the APFT has been criticised because of insufficient evidence validating its association with military job performance and combat.[2 12–14] New batteries of fitness tests proposed in 2002 and 2010 were also considered not validated, so were not implemented.[2 13] As military services redefine combat readiness requirements to support full gender integration in combat positions,[9 15 16] the APFT has again been part of a reevaluation effort.[13 17]

Fitness tests used as job selection, promotion or retention criteria should be linked to capabilities critical to the nature of the job in order to address legal and antidiscrimination requirements.[9 11 18] Towards this goal, numerous studies have evaluated the relationship between individual fitness tests and

To cite: Hauschild VD, DeGroot DW, Hall SM, *et al*. *Occup Environ Med* Published Online First: [*please include* Day Month Year] doi:10.1136/oemed-2016-103684

Table 1 Physical fitness component groups and occupational task categories used for correlation meta-analyses

Four health-related physical fitness component groups*		Twelve common occupational task categories†
Cardiorespiratory endurance (CRe)	Ability to sustain low-intensity muscle contractions for extended period of time. Gold-standard physiological measure is the body's maximum rate oxygen (VO_2max). Also known as 'aerobic fitness', 'aerobic capacity' and 'stamina'. Example tests: ▶ Machine tests (treadmill, step) that include measuring or estimating VO_2max ▶ Surrogate measures: distance run tests for time (eg, 1–3 miles), fixed-distance runs (12 min)	▶ *Lift and lower (single)*‡ One time maximum lift of equipment ▶ *Lift and lower (repeated)*‡ Repeated lifting equipment on/off ground or vehicles ▶ *Lift and carry*‡ Carry various equipment items various distances ▶ *Stretcher carry*‡ A specific type of lift and carry task (two person) ▶ *Push–pull*‡ Manual movement of equipment along a surface (not lifting)
Muscular strength (UBs, LBs, TRs)	Ability to exert maximal force against a fairly immovable object for very brief period (seconds). Measurements reflect force; no physiological gold standard exists.Example tests: ▶ One-repetition maximum (1RM) lifts using various free weights or isometric machines (UBs and LBs, depending on machine) As a surrogate: *Explosive strength* is a form of power that refers to ability to use maximum energy to rapidly project object or body in a single maximum effort in a very brief amount of time (seconds) Example tests: ▶ Jumps (eg, vertical, broad jump, squat) (LBs) ▶ Shot put, ball throws (UBs)	▶ *Casualty drag*‡ Life-saving task may include extrication and/or different carrying techniques ▶ *Dig*‡ Establish fighting position, structural support, fill sandbags ▶ *March/walk (with a load)*‡ Move body long distances wearing some form of gear ▶ *Move fast*§ With or without change of direction for short distances ▶ *Climb*§ Includes scale, jump, descend stairs, walls, vehicles, obstacles
Muscular endurance (UBs, LBs, TRs)	Ability to conduct high-intensity muscle contractions repeatedly for relatively short periods (30 s to 2 min). Measurements reflect force/time but no physiological gold standard exists. Example tests: ▶ Push-ups and pull-ups (UBe) ▶ Weight lifting maximum repetitions (UBe) ▶ Endurance squats (LBe) ▶ Sprints and shuttles (LBe) ▶ Sit-ups (TRe)	▶ *Crawl*§ High and low techniques ▶ *Multiactivity*‡,§ Combination of three or more tasks; 'obstacle course'; 'circuit'
Flexibility (FLX)	Ability to flex or lengthen various parts of the body. Includes static (steady hold) or ballistic (repeated, rapid) forms. Example test: ▶ Sit-and-reach (static, back and hamstring stretch)	

*The four primary activity-based health-related physical fitness components in conjunction with definition and example tests used to provide measurements.[7 19]
†Categories from review of tasks common to US Army and other military ground forces[13 17 24 25] and found to be similar to tasks in occupational studies.
‡Includes manual movement of equipment, supplies, people; variables include amount of weights, duration, distances, heights, terrain and environmental conditions and other preceding/concurrent activities. Since stretcher carry has been historically studied as a unique task, it is considered separate from the lift and carry task.
§Movement of body activities include variable loads, distances, duration, heights, terrain and environmental conditions and other preceding/concurrent activities.

various job tasks.[17] Many studies use regression models, but the selection of different tasks and test variables has made it difficult to compare study outcomes. Several studies have also examined the relationship by calculating Pearson correlation coefficients (r values) between measures of individual fitness test and occupational task performance. However, findings have not been consistent or robust and have been constrained to the existing studies' specific settings, tasks and fitness tests.

As a result, employers have continued to be confronted with the dilemma of how to select the best and most practical means to promote, monitor and test the physical fitness of employees.[1 4 5 9 13] This systematic review aggregated data from applicable studies to describe the relationship between core components of health-related physical fitness activity (cardiorespiratory endurance (CRe), muscular strength, muscular endurance and flexibility (FLX))[7 19] and the performance of common physically demanding job tasks. The physical fitness components and associated tests that most strongly correlate with common physical job tasks provide a basis for the selection of relevant health-related occupational fitness tests.

METHODS

Study design

Systematic review methodology was used to identify and select applicable quality original studies, extract the desired correlation data and synthesise the outcomes using a meta-analytical technique to describe the collective evidence.[20–22] This type of review did not require human use consent. The comprehensive investigative team included a military physiologist, a public health scientist, a kinesiologist, a physical therapist, a physician-epidemiologist and two statisticians. Because the project was conducted in response to a directed military initiative, the data collected and analysed were archived in a military technical report.[17]

In order to evaluate the data amassed from selected studies, individual fitness tests were organised into the four components of health-related physical fitness activity: cardiovascular endurance (CRe), muscular strength, muscular endurance and FLX (table 1). Since fitness tests for muscular strength and muscular endurance primarily test one region of the body, these two components were also grouped into predominant body regions. These groups included strength and muscular endurance of the upper body (UBs and UBe), lower body (LBs and LBe) and trunk (TRs and TRe)). Skill-related fitness components (eg, agility, coordination, balance, power and speed) were not a focus of this review. However, some common tests of power or speed (eg, jump tests, sprints, shuttle runs) encompass elements of muscular strength and muscular endurance so were included.[9 23]

In addition to grouping the fitness tests, 12 task categories (table 1) were identified as representative of the common physically demanding occupational tasks of interest. These task categories were derived from a review of common military tasks to identify those of greatest relevance to all Army personnel.[13 17 24 25] However, these task categories are also relevant to civilian public safety and emergency response-related occupations such as firefighting and law enforcement.[26–28] While the exact task details can vary among different military services or civilian occupations, the general functional movements and types of physical demands represented by each task category were deemed to be similar. For example, victims or casualties can be different sizes and weights, and they can be carried over varying distances and surfaces. Though individual studies used unique quantitative metrics (eg, specific weights or distances), the variation captured by the aggregate data more appropriately replicates the real-world variability.

Search criteria for studies

Eligible literature included original studies of military and civilian healthy adults, ages 18–65 years, who performed one or more of the identified tasks. The population of interest was intended to reflect the diversity of personnel serving in the US military as well as other occupational sectors (ie, firefighters and law enforcement). Included study populations spanned a wide range of demographic and motivational variables and included extremely fit, athletic personnel and less fit personnel. Selected studies had to provide Pearson correlation coefficients (r values) between physical fitness tests and performance measurements for one or more tasks.

Search strategy

The systematic literature search included English language studies published in peer-reviewed journals and scientific technical reports between 1970 and February 2013. Sources included PubMed (MEDLINE), Biomedical Reference Collection, Academic Search Premier, Nursing & Allied Health Collection: Comprehensive, Cochrane Methodology Register, CINAHL & CINAHL Full Text and EMBASE. In addition, the Defense Technical Information Center (DTIC) was searched for relevant military technical reports and subject matter experts contacted for additional references. Search terms were adapted to address different database systems but included variations of the following: 'test, requirement or standard', 'performance' or 'capability', 'functional ability' and 'work', 'job' or 'occupation', or 'task,' as well as 'physical fitness', 'mobility' and a variation of each of our selected key physical fitness component terms: 'cardiorespiratory', 'aerobic fitness', 'muscle strength', 'muscle endurance' and 'flexibility'.

Study selection and data extraction

Two team investigators (VDH, DWD) conducted separate database searches and merged results into a single Endnote file. Investigators reviewed titles and abstracts, excluding those that did not meet study criteria or provide Pearson correlation coefficients. Eligible full-text studies were reviewed and scored by each investigator to ensure study relevance and quality. The quality scoring tool included 10 appraisal criteria adapted from similar reviews:[21 23] (1) study objective, (2) design and methods, (3) sample size, (4) population and setting characteristics, (5) control for confounders in study design to isolate effect of interest, (6) repeatability, (7) data presentation that demonstrated controls for confounding, (8) analyses techniques, (9) adequacy of results and (10) variability. Scoring differences between investigators were discussed to achieve consensus. Data extracted from selected studies included study and population characteristics, health-related component fitness tests, applicable job tasks with metrics and measurements and correlation values. Individual selection study bias was minimised by including significant and non-significant correlations from all selected studies. When applicable, reported correlations were standardised using absolute r values.

Statistical analyses and interpretation

To synthesise the extracted data, investigators grouped the r values into fitness component and task category combinations and mathematically combined them into single representative effect estimates.[21 29] Meta-analysis techniques were applied using the number of studies, original study r values and sample sizes to generate pooled r values between each fitness component and job task combination.[29] For fitness component and task combinations with only one study, a pooled r was not calculated. Original r values were converted to a common test metric (z-values and corresponding variance) using Fisher's r to z transformation.[29]

$$Z_{r_i} = \frac{1}{2}\ln\left(\frac{1+r_i}{1-r_i}\right) \quad s_z^2 = \frac{1}{n-3} \tag{1}$$

Fisher's z values from the original studies were combined using fixed or random effect models depending on their homogeneity.[21] The presence of heterogeneity was tested using the Cochran Q statistic, with $p<0.05$ as the level of significance. An I^2 test was also performed to quantify heterogeneity. The fixed effect model was used if there was no evidence of heterogeneity. In cases of statistically significant heterogeneity, the pooled effect estimate was determined using the random effect model. This was performed using formulas within Excel file of extracted data. CIs 95% and p values were calculated for each pooled correlation coefficient. Subgroup analyses of separate gender data and specific fitness tests were conducted for task-test combinations with correlations from at least two studies for one or more task for each physical component. The mean weighted correlations were also calculated for each fitness component for all task categories combined.

Since empirical guidelines for interpreting the magnitude of correlation coefficients are lacking, a scale to interpret the pooled r values was created a priori. The five-tiered scale was based on a review of guidelines for similar types of relationships found in human performance and social sciences.[17 30 31] For example, $r\geq0.70$ has been described as a very strong or excellent linear relationship, $r\geq0.50$ as a demarcation for large or strong relationships and $r\leq0.30$ or <0.10 as a weak relationship.[17 30 31]

RESULTS

The literature search and study selection are summarised in figure 1. Of 273 studies identified for full text review, most were excluded due to the lack of requisite statistical analyses. Of the 33 studies selected for scoring, investigators' scores were generally consistent and brief discussions reconciled differences. Five studies were eliminated because study procedures and statistical methods were not suitable for this analysis. Two studies described data from the same study, so one was eliminated. While the quality scores of the remaining studies varied, each study provided adequate methodological validity to minimise individual study biases for the purpose of this review.

The 27 selected studies (table 2) represented a variety of US and international military and non-military healthy adult populations.

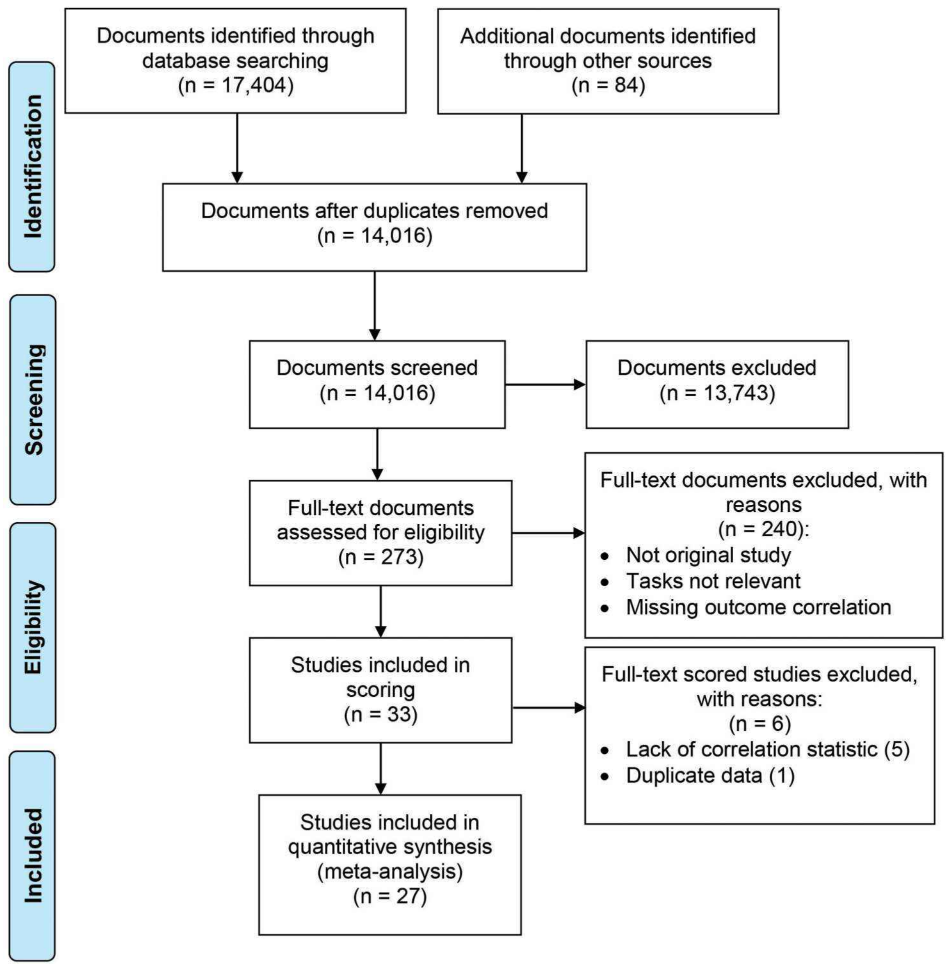

Figure 1 Study selection flow chart.

Studies included 13 military population studies,[32–36 38 39 42 44–46 53 54] 10 firefighter, law enforcement, or peace officer studies,[26–28 37 40 43 47 49 50 52] and 4 other relevant studies of healthy civilian populations.[41 48 49 51] The selected studies provided 533 distinct Pearson r correlation coefficient values between physical fitness tests and job tasks. The most frequently studied fitness components were UBs (18 studies, 122 correlations) and UBe (20 studies, 117 correlations). The least studied fitness component was flexibility (FLX, addressed in only three studies yielding 15 correlations). Of the task categories, multiactivity was the most frequently studied (11 studies, 76 correlations), followed by the lift and carry (6 studies, 64 correlations) and stretcher carry (6 studies, 59 correlations). No correlations were found between the crawl task and LBe, and only one study evaluated the correlation between the loaded march task and CRe fitness; therefore, pooled r values were not calculated for these combinations.

Table 3 presents the pooled correlation values between each physical fitness component and occupational task category. Though the number of correlations for each meta-analysis varied and fixed models were primarily used to address heterogeneity among studies, the vast majority of the pooled r values had narrow CIs ($p<0.05$). Pooled values also did not appear sensitive to outlier correlation values (data not shown).[17] Separate pooled r values were calculated for some specific types of physical fitness tests (footnotes in table 3). For example, from the CRe fitness component group, separate pooled r values were calculated for timed-distance runs (ie, 1.5, 2, or 3.1 miles), fixed-time distance runs (ie, distance in 12 min) and tests that provided maximum oxygen uptake or VO_{2max}. Individual test-specific pooled r values with relatively robust data included hand grip tests for UBs, the standing broad jump (SBJ) and vertical jump (VJ) for LBs, push-ups for UBe and sprint tests for LBe. Though data were too limited to assess differences among males and females for all tasks, separate correlation data for male and female sample populations were adequate for calculating pooled r values between each fitness component and the crawl and stretcher carry tasks. The CRe component had the highest pooled correlations for both genders for the stretcher carry (male pooled r=0.63, female pooled r=0.60) and the crawl task (male pooled r=0.63, female pooled r=0.74).[17]

Table 4 summarises the overall strengths of the pooled r values for each fitness component. The CRe component had the greatest number of very strong (≥0.70) to strong (≥0.50) pooled correlations (8 of 12 task categories). The next most notable fitness component was LBs, followed by UBe and UBs, respectively. When all tasks were combined, CRe was the only fitness component with a strong weighted average (mean) correlation; weighted mean correlations were moderate for UBs, LBs, UBe, LBe and TRs, fair for TRe and weak for FLX (data not shown).[17]

DISCUSSION

Summary of evidence

This review combined correlation data from existing studies to examine the evidence of association between core components of physical fitness and common physically demanding occupational tasks. Numerous individual studies regarding these relationships exist, but findings have been inconsistent and no prior review has

Table 2 Summary of selected studies based on year of publication*

Source*	Type†	Study country	Population type	Sample size	Gender	Tests groups(# correlations)	Task categories	Study quality rank‡
Wright *et al*[32] 1984	TR	USA	Military (Army)	272	M, F (221, 51)	CRE (3) UB-S (3) LB-S (3) TR-S (3) UB-E (3) TR-E (3)	Lift and lower (single) Lift and lower (repeated)	++
Robertson and Trent[33] 1985	TR	USA	Military (Navy)	45	M, F (24, 21)	UB-E (4) TR-E (6) UB-S (6)	Lift and carry Stretcher carry	++
Beckett and Hodgdon[34] 1988	TR	USA	Military (Navy)	102	M, F (64, 38)	CRE (4) UB-E (16) TR-E (5) UB-S (6) LB-S (10) FLX (4)	Lift and carry Lift and lower (S) Move fast	+++
Mello *et al*[35] 1988	TR	USA	Military (Army)	28	M	LB-E (16) LB-S (16)	Loaded march	++++
Stevenson *et al*[36] 1989	J	Canada	Military	16	M	UB-E (2) UB-S (4)	Lift and lower (S) Lift and lower (R)	++++
Schonfeld *et al*[37] 1990	J	USA	Firefighters	20	M	CRE (6)	Climb Casualty drag Multiactivity	+++
Singh *et al*[38] 1991	TR	Canada	Military	116	M	TR_S (10) UB_S (10)	Casualty drag Lift and lower (R) Dig Stretcher carry crawl	++
Arvey *et al*[28] 1992	J	USA	Police	276	M, F	CRE (6) UB_S (6) LB_S (3) UB_E (6) LB_E (3) TR_E (6)	Casualty drag Move fast Multiactivity	+++
Stevenson *et al*[39] 1992	J	Canada	Military Personnel	132	M, F (99, 33)	UB_S (8) UB_E (8) TR_E (8)	Crawl Dig Lift and carry Stretcher carry	++
Myhre *et al*[40] 1997	TR	USA	Firefighters	279	M, F	UB_S (3) UB_E (1)	Multiactivity	++
Kraemer *et al*[41] 1998	J	USA	Civilian volunteers (military)	123	F	AER (3) UB_S (6) LB_S (6) UB_E (3) LB_E (3)	Lift and lower (R) Lift and lower (S) Loaded march	+
Knapik *et al*[42] 1999	J	USA	Military (Army)	11	M, F (7, 4)	CRE-tr (1) UB-S (4) LB-S (1) UB-E (1) TR-E (1)	Stretcher carry	+++
Williford *et al*[43] 1999	J	USA	Firefighters	91	M	CRE (5) UB_S (5) TR_E (5) UB_E (10) FLX (5)	Casualty drag Climb Push–pull Lift and lower (S) Multiactivity	++
Deakin *et al*[44] 2000	TR	Canada	Military	623	M, F (416, 207)	CRE (5) UB_S (15) LB_S (5) TR_S (5) UB_E (10) LB_E (10) TR_E (5)	Lift and carry Lift and lower (R) Dig Stretcher carry Crawl	++
Pandorf *et al*[45] 2001	J	USA	Military (Army)	12	F	UB-E (4) TR-E (3)	Crawl Multiactivity	+++
Bilzon *et al*[46] 2002	J	UK	Military	93	M, F (52, 41)	CRE (4) UB_S (4) LB_S (2) UB_E (4)	Stretcher carry	+++

Continued

Table 2 Continued

Source*	Type†	Study country	Population type	Sample size	Gender	Tests groups(# correlations)	Task categories	Study quality rank‡
						LB_E (2) TR_E (2)		
Rhea *et al*[47] 2004	J	USA	Firefighters	20	M, F (17, 3)	CRE (5) UB_S (10) LB_S (10) UB_E (25) LB_E (5) TR_E (5)	Casualty drag Climb Lift and carry Multiactivity Push–pull	+++
Barnes *et al*[48] 2007	J	USA	Volleyball players	29	F	LB_S (1)	Move fast	++++
Harman *et al*[49] 2008	J	USA	Civilian volunteers (military)	32	M	CRE (4) UB_E (4) LB_E (4) TR_E (4) LB_S (8)	Casualty drag Move fast Multiactivity	++
Williams-Bell *et al*[50] 2008	J	Canada	Firefighters recruits	41	M, F (32, 14)	UB_S (2) LB_S (1) UB_E (1) LB_E (1)	Multiactivity	+
Michaelides *et al*[49] 2008	J	USA	Firefighters	38	M	UB_S (1) UB_E (1) TR_E (1) LB_S (1) FLX (1)	Multiactivity	++
Hoffman[29] 2009	TR	USA	Peace officers	128	M, F	AER (4) UB_S (4) LB_S (10) UB_E (4) TR_E (4) FLX (4)	Move fast Multiactivity	++
McBride 2009[51]	J	USA	Football players	17	M	LB_S (3)	Move fast	++
Phillips 2010[52]	J	Australia	Firefighters	38	M	UB_S (1) LB_S (1) UB_E (3) LB_E (1)	Loaded March	++
Aandstad 2011[53]	J	USA	Military cadets	42	M	CRE (1)	Move fast	++
Michaelides *et al*[26] 2011	J	USA	Firefighters	67	M	UB_S (21) LB_S (7) TR_S (7) UB_E (7) TR_E (7) FLX (1)	Casualty drag Climb Lift and carry Dig Push–pull Multiactivity	++
Thebault 2011[54]	J	France	Military	19	M	LB_S (2)	Move fast	++

*Study publications order based on year of publication then alphabetically.
†Publication type: J refers to an article in a peer-review journal; TR is a publically available government/military technical report.
‡Quality score descriptor based on consensus of two independent investigators evaluation of maximum of 20 criteria: >15=++++, 15–13=+++, 12–10=++, 9=+; <9 eliminated.
CRe, cardiorespiratory endurance; E, muscle endurance; FLX, flexibility; LB, lower body; S, muscular strength; TR, trunk/core; UB, upper body.

systematically compiled data in a construct evaluation. Though original studies varied in size, population and design, considerable commonalities warranted this rationalised grouping of the data. The resulting pooled correlations between the fitness components and common job task categories provide quantitative evidence to support the selection of occupational fitness tests. The results demonstrate the important contributions of CRe, LBs and UBe and LBe to overall performance of 12 task categories. Not surprisingly, the evaluation demonstrated that some individual tasks are more strongly correlated to certain fitness components than others (eg, UBs is more correlated with the single lift and lower task than any other fitness component). The results also support the concept that no single fitness test represents the overall fitness needed for strenuous jobs. Though much of the research has focused on UBs and UBe, this evaluation indicates that CRe, LBs and LBe deserve particular attention. Data gaps between CRe and LBe for some tasks and are not expected to have led to an overestimation of the importance of these fitness components. Since this review did not identify notable correlations between job tasks and TRs, TRe or FLX, these fitness components appear to be less relevant. This finding may be affected by data gaps (such as between FLX and various tasks, or TRe and TRs and the load carriage walking task) or limited to the singular tests used. For example, though data for TRe was fairly abundant for most task categories, timed sit-ups have essentially been the only test used. There were no strong correlations between the sit-up test and any of the occupational task categories. This finding supports a prior review that questioned the value of the sit-up test and its reliability as an occupational test.[23]

In addition to the relationship with task performance, other factors should be considered to maximise the value of selected tests, ensure safety and minimise costs. For example, fitness tests can also be used as metrics for other constructs of job success, such as injury risk, attrition, absenteeism and even mortality.[4 7 9 23 55] These constructs can be particularly important for public safety

Table 3 Pooled correlation values (r) between health-based component fitness test groups and occupational task categories

		Muscular strength			Muscular endurance			
Task categories	Cardio respiratory*	Upper body‡	Lower body§	Trunk¶	Upper body**	Lower body††	Trunk‡‡	Flexibility†
Lift and lower (single)	**0.30** [5] (0.15, 0.44)	**0.75** [10] (0.66, 0.81)	**0.60** [7] (0.52, 0.67)	{0.57} [1]	**0.42** [11] (0.31, 0.53)	**0.56** [3] (0.48, 0.63)	**0.16** [4] (0.80, 0.24)	**0.16** [3] (0.05, 0.27)
Lift and lower (repeated)	**0.60** [5] (0.48, 0.70)	**0.61** [11] (0.47, 0.73)	**0.57** [6] (0.37, 0.72)	**0.56** [5] (0.32, 0.73)	**0.62** [6] (0.46, 0.74)	{0.55} [1]	**0.29** [3] (0.05, 0.51)	–
Lift and carry	**0.72** [4] (0.51, 0.85)	**0.43** [19] (0.34, 0.52)	**0.41** [7] (0.20, 0.59)	**0.41** [4] (0.18, 0.60)	**0.50** [17] (0.37, 0.61)	**0.47** [4] (0.35, 0.57)	**0.25** [8] (−0.08, 0.52)	{0.01} [1]
Casualty drag	**0.32** [7] (0.23, 0.40)	**0.38** [9] (0.24, 0.51)	**0.27** [5] (0.14, 0.39)	**0.27** [3] (0.16, 0.37)	**0.33** [11] (0.19, 0.45)	**0.46** [5] (0.20, 0.66)	**0.16** [6] (0.08, 0.25)	{0.06} [1]
Stretcher carry	**0.66** [7] (0.53, 0.76)	**0.65** [22] (0.60, 0.69)	**0.73** [5] (0.62, 0.81)	{0.67} [1]	**0.58** [15] (0.48, 0.66)	–	**0.31** [9] (0.12, 0.48)	–
Push–pull	**0.09** [2] (−0.10, 0.28)	**0.46** [7] (0.28, 0.61)	**0.21** [5] (0.10, 0.32)	**0.42** [2] (0.27, 0.55)	**0.46** [9] (0.29, 0.60)	**0.35** [4] (0.21, 0.48)	**0.20** **[4]** (0.08, 0.32)	{0.06} [1]
Loaded march/walk	{0.60} [1]	**0.28** [5] (0.04, 0.49)	**0.32** [19] (0.25, 0.39)	**0.01** [2] (−0.12, 0.13)	**0.48** [4] (0.25, 0.66)	**0.38** [18] (0.31, 0.45)	–	–
Move fast	**0.59** [8] (0.51, 0.66)	**0.35** [5] (0.20, 0.49)	**0.58** [13] (0.52, 0.63)	–	**0.47** [9] (0.35, 0.57)	**0.69** [2] (0.62, 0.75)	**0.39** [7] (0.33, 0.45)	**0.08** [3] (−0.03, 0.18)
Climb	**0.55** [4] (0.42, 0.66)	**0.22** [5] (−0.04, 0.45)	**−0.09** [3] (−0.24, 0.08)	{0.38} [1]	**0.46** [8] (0.37, 0.54)	**0.44** [3] (0.26, 0.58)	**0.43** [3] (0.30, 0.54)	{0.25} [1]
Crawl	**0.80** [2] (0.72, 0.86)	**0.49** [5] (0.38, 0.59)	**0.65** [2] (0.39, 0.82)	{0.64} [1]	**0.66** [5] (0.50, 0.77)	–	**0.48** [5] (0.22, 0.68)	–
Dig	**0.62** [2] (0.51, 0.71)	**0.44** [9] (0.31, 0.56)	**0.53** [3] (0.37, 0.65)	**0.47** [4] (0.23, 0.65)	**0.38** [5] (0.12, 0.59)	{0.15} [1]	**0.21** [4] (−0.04, 0.44)	–
Multiactivity	**0.52** [9] (0.47, 0.58)	**0.42** [15] (0.33, 0.51)	**0.47** [9] (0.36, 0.58)	{0.53} [1]	**0.46** [17] (0.38, 0.54)	**0.64** [10] (0.62, 0.70)	**0.38** [10] (0.32, 0.44)	**0.08** [5] (−0.02, 0.18)

[ ]=n correlations from original studies; ()=95% CIs; bold numbers are pooled correlation values; { } only a single study found.

*Results reflect pooled correlations for all fitness tests that provided a measurement for CRe including timed runs, fixed-distance runs and tests that measured or estimated VO_{2max}. Though data became increasingly limited when attempting to evaluate these three separate types of CRe tests, they were also evaluated separately when data were adequate for tasks (two or more studies). From these separate analyses, timed runs (1, 1.5, 2 miles) yielded several task for which pooled r>0.50: lift and lower (repeated) r=0.51 (n=3, CI 0.45 to 0.56); move fast r=0.58 (n=7, CI 0.49 to 0.66); multiactivity r=0.52 (n=6, CI 0.46 to 0.59). Tests providing an estimated or measured VO_{2max} included: lift and lower (repeated) r=0.70 (n=2, CI 0.59 to 0.79); stretcher carry r=0.71 (n=4, CI 0.57 to 0.81) and crawl r=0.80 (n=2, CI 0.72 to 0.86).

†The sit-and-reach was the only test used.

‡Results reflect pooled correlations for all fitness tests that provided a measurement for UBs. A common test was the dynamometer grip test. Pooled correlations for grip tests with adequate data yielding pooled r>0.50 included: lift and lower (single) r=0.67 (n=2, CI 0.43 to 0.82); lift and lower (repeated) r=0.59 (n=4, CI 0.27 to 0.80); stretcher carry r=0.61 (n=10, CI 0.52 to 0.70).

§Results reflect pooled correlations for all fitness tests that provided a measurement for LBs including measures of power (single maximum bursts). Common tests included the SBJ and the VJ. Pooled correlations for SBJ and VJ with adequate data yielding pooled r>0.50 included: lift and lower (single) SBJ r=0.71 (n=2, CI 0.63 to 0.77) and VJ r=0.52 (n=2, CI 0.41 to 0.61); stretcher carry SBJ r=0.83 (n=2, CI 0.77 to 0.87); move fast SBJ r=0.52 (n=2, CI 0.31 to 0.68) and VJ r=0.60 (n=6, CI 0.54 to 0.66); multiactivity VJ r=0.52 (n=4, CI 0.44 to 0.60).

¶Results reflect pooled correlations for all fitness tests that provided a measurement for TRs, no single test type had adequate data to pool correlation data.

**Results reflect pooled correlations for all fitness tests that provided a measurement for UBe including machines and weights, pull-ups and push-up tests. Pooled correlations for push-up tests with adequate data yielding pooled r>0.50 included: lift and lower (repeated) r=0.57 (n=4, CI 0.31 to 0.75); move fast r=0.52 (n=5, CI 0.45 to 0.59); crawl r=0.58 (n=4, CI 0.21).

††Results reflect pooled correlations for all fitness tests that provided a measurement for LBe including sprint tests 100–400 m and shuttle tests, though recognised that the shorter distances completed in <30 s could be considered measures of LB power. Pooled correlations for sprint tests with adequate data yielding pooled r>0.50 included: lift and lower (single) r=0.63 (n=2, CI 0.54 to 0.71); lift and carry r=0.55 (n=2, CI 0.41 to 0.66); casualty drag r=0.53 (n=3, CI 0.44 to 0.61) and multiactivity r=0.71 (n=5, CI 0.66 to 0.75).

‡‡The sit-up test was the test used in all studies except one.

CRe, cardiorespiratory endurance; SBJ, standing broad jump; TR, trunk/core; UB, upper body; VJ, vertical jump.

and life-saving occupations. Employers should also consider the feasibility and reliability of fitness tests. Factors that pertain to specific test selection for each fitness component are described below.

Cardiorespiratory endurance tests

Though included in less than half of the selected studies, run tests are routinely used by armed forces to test CRe and have been a component of US Army fitness test batteries since the early 1900s.[2] This review supports the inclusion of CRe tests in occupational monitoring since this fitness component was strongly correlated with the greatest number of the task categories. This makes physiologic sense given that aerobic metabolism increasingly becomes the dominant source of energy for tasks lasting more than a few minutes.[4 7 23] In fact, because real-life occupational tasks often occur over extended periods of time, the significance of the CRe component is likely underestimated

Table 4 Number of pooled correlations between fitness components and task categories* by correlation strength†

Strength of pooled correlations (pooled r value)†	Cardio respiratory (56 total r values)	Muscular strength (231 total r values): Upper body (122 r values)	Muscular strength: Lower body (84 r values)	Muscular strength: Trunk (25 r values)	Muscular endurance (231 total r values): Upper body (117 r values)	Muscular endurance: Lower body (51 r values)	Muscular endurance: Trunk (63 r values)	Flexibility (15 total r values)
Very strong (r>0.70)	2	1	1	0	0	0	0	0
Strong (0.50≤ r<0.70)	6	2	5	1	4	3	0	0
Moderate (0.40≤r<0.50)	0	5	2	3	6	3	2	0
Fair (0.30≤r<0.40	2	2	1	1	2	2	2	0
Weak (r<0.30)	1	2	3	2	0	0	7	3
Single study or no study; data inadequate to pool for task(s)	1	0	0	5	0	4	1	9
No studies/data found for task(s)	–	–	–	1	–	3	1	5

*For the 12 occupational task categories evaluated in this review, described in table 1.
†Strength of correlation qualified in terms of a priori scale described in methods.[17 30 31]

by this review of short duration tasks. Studies of military and firefighter activities provide evidence for this high demand for CRe during these occupations.[1 3 4 56] An additional value of measuring CRe is that it has consistently been associated with risk of injury and cardiovascular disease in studies of personnel in physically demanding jobs.[1 4 7 17 55 57] Poor CRe has also been linked to higher military attrition or drop-out.[7 23] Though CRe appears to be a critical occupational fitness component for both genders, application of CRe tests should consider potential sex differences in occupational CRe physical ability.[9 13] In addition, body composition may need to be factored into test applications since anthropomorphic measures (ie, lean body mass) took precedence in some of the regression models used in original studies.[27 28 58]

The gold-standard means to monitor CRe is to directly measure VO_{2max} using calibrated calorimeter equipment and trained test personnel.[4 7 23] Since this is impractical for mass routine screening required in many occupational settings, VO_{2max} is often estimated from calculations of more expedient tests or represented by simple surrogate measures such as run time. Run times from timed-distance runs (1.5–26 miles) and fixed-time (12 min) run tests have been validated against VO_{2max} testing with correlation coefficients ranging from 0.70 to 0.90.[23] Their test–retest reliability has also been reported as high, with reliability coefficients ranging from 0.82 to 0.98.[23] Though the strongest correlations between CRe and tasks identified in this review were from tests that provided direct and estimated VO_{2max} measurements, the timed-distance (ie, 1–3 miles) and fixed-time (12 min) runs also yielded strong correlations.

Muscular strength tests

Despite their emphasis in scientific studies, muscular strength tests have not been commonly used in occupational test batteries.[2 12 13] Practical considerations, such as safety and the need for equipment that must be calibrated or standardised, may limit the use of some strength tests in routine mass testing. However, the inclusion of the VJ and hand grip strength tests in the newest Canadian military basic fitness test battery demonstrates potential feasibility.[16] Though UBs has been more frequently studied, this review shows that LBs is strongly correlated with more occupational tasks and therefore may be a more important test consideration.

No physiological gold-standard measurement exists for measuring muscular strength, so tests were quite varied. The most common UBs test was the dynamometer hand grip test, presumably because it is simple, safe and easily administered. A comparison of hand grip test correlations with same-study UBs tests that required activation of larger muscle groups (eg, bench press) suggests some comparative values.[46 47] Though hand grip tests strongly correlated with one-quarter of the task categories, test–retest reliability can be affected by several factors (ie, calibration, use of one or both hands, number of attempts and pre-post maximal exertion).[59] Of the various tests used to measure LBs, field-expedient surrogate tests of power (eg, VJ and SBJ) were especially common. These jump tests were strongly correlated with one-third of the task categories and have had good test–retest reliability coefficients ranging from 0.76 to 0.96.[23 60] So regression analyses also suggest the potential relevance of jump tests. For example, the SBJ was included in the predictive equation identified by Bilzon *et al*[46] and Harman *et al*[49] found that only the VJ (aside from anthropomorphic measures such as lean body mass) consistently entered several predictive regression equations.

Muscular endurance tests

Upper body muscular endurance tests are frequently studied and more commonly included in military and occupational physical fitness tests than LBe tests.[2 6 28] However, LBe strongly correlates to different task categories compared to UBe. Since there is no physiological gold standard, fitness tests used to measure muscular endurance have encompassed a variety of repeated lift tests (free weights or machine) and maximum repetition tests. Push-ups, the most common UBe test, strongly correlated with a quarter of the task categories. A review of the test–retest reliability of push-up tests indicates good reliability coefficients ranging from 0.76 to 0.83.[23] This field-expedient test has particular appeal because it does not rely on equipment (eg, calibrated weights or machines or bars) and is not associated with safety risks (eg, dropped weights, falling from bars). As an added value, low scores on push-up tests have also been found to be a risk indicator for injury in military and law enforcement training populations.[6 55] Measuring LBe, such as with sprint tests (100–400 m), may be of particular value for occupations that require the casualty drag task, since these were the only fitness tests with a strong pooled correlation to this task. Sprint tests require minimal equipment and have reported good test–retest reliability (reliability coefficients ranging from 0.87 to 0.98).[23]

Limitations

There are various limiting factors inherent to the selection and statistical analyses of original studies in this review. Project timelines constrained the study selection to English language documents published between January 1970 and February 2013. Publication bias may have further limited the acquisition of relevant data. Original studies were of varied design, quality, limited sample sizes and did not provide adequate participant data. Consequently, gender, age, body size and composition and other characteristics such as health status and motivation levels could not be quantified. The correlation values described the linear relationship as a metric of association, so potential non-linear or multivariable relationships are not described. Although studies were reviewed and scored to ensure relevance and quality, original studies describe different confounding variables and addressed them in different ways. Therefore, the impact of potentially significant confounders like body composition (lean body mass and lean body mass to dead mass ratio)[28 58] or gender[9] on the calculated pooled r values could not be evaluated. It has also been shown that although meta-analyses with Fisher's z value and the Q statistic is a scientifically accepted technique to aggregate comparable data, a limited number of studies can underestimate heterogeneity.[22]

CONCLUSIONS

Occupational health practitioners charged with ensuring a physically capable workforce can use physical fitness tests as indicators of employees' physical health status. The most beneficial tests are those that are strongly related to the common tasks critical to overall job performance. For example, this evaluation of 12 task categories collectively relevant to military and other physical occupations demonstrated physically demanding jobs should consider CRe, LBs and UBe tests. To maximise the value of testing, selection of specific tests should also consider the association of tests to other indicators of job success, such as injury risk and attrition, as well as test reliability, feasibility and cost. Valid and reliable field-expedient tests to consider include timed runs (1.5–3 miles) for CRe, jump tests (VJ or SBJ) for LBs, push-ups for UBe and sprints for LBe. Grip tests for measuring UBs may be useful if reliability can be established. Though frequently used, sit-up and sit-and-reach tests may not provide relevant occupational health metrics. Future studies should evaluate the impacts of physiological characteristics such as lean body mass and gender on the application of these tests and the establishment of appropriately scaled fitness test standards.

Acknowledgements The authors acknowledge editorial contributions of Ms Mimi Eng, statistical consult of Angela Mound and Bonnie Taylor, and reviews by Jessica Sharkey, MPH, and Anna Schuh, PhD.

Contributors VDH, DWD, TLG and KGH conceived the study. VDH and DWD conducted the literature search, assessed quality and selected studies and extracted data. TLG, KGH and BHJ reviewed data and advised on meta-analytical grouping. SMH and KDD gave methodological advice and conducted statistical analyses. VDH wrote the draft of the paper. All authors contributed to writing, reviewing and revising paper and approved final manuscript. All authors are employed by the Army Public Health Center (Provisional), Aberdeen Proving Ground, MD.

Disclaimer The views expressed are those of the authors and do not necessarily reflect official policy of the Department of Defense, Department of the Army, US Army Medical Department or the USA.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Storer TW, Dolezal BA, Abrazado ML, *et al*. Firefighter health and fitness assessment: a call to action. *J Strength Cond Res* 2014;28:661–71.
2 Knapik JJ, East WB. History of the United States army physical fitness and physical readiness testing. *US Army Med Dep J* 2014:5–19.
3 Carlson MJ, Jaenen SP. The development of a pre-selection physical fitness training program for Canadian special operations regiment applicants. *J Strength Cond Res* 2012;26(Suppl 2):S2–S14.
4 Smith DL. Firefighter fitness: improving performance and preventing injuries and fatalities. *Curr Sports Med Rep* 2011;10:167–72.
5 Plat MJ, Frings-Dresen MH, Sluiter JK. A systematic review of job-specific workers' health surveillance activities for fire-fighting, ambulance, police and military personnel. *Int Arch Occup Environ Health* 2011;84:839–57.
6 Payne W, Harvey J. A framework for the design and development of physical employment tests and standards. *Ergonomics* 2010;53:858–71.
7 National Research Council (NRC). *Assessing fitness for military enlistment*. Washington (DC): National Academy Press, 2006:211.
8 Harbin G, Olson J. Post-offer, pre-placement testing in industry. *Am J Ind Med* 2005;47:296–307.
9 Courtright SH, McCormick BW, Postlethwaite BE, *et al*. A meta-analysis of sex differences in physical ability: revised estimates and strategies for reducing differences in selection contexts. *J Appl Psychol* 2013;98:623–41.
10 Sothmann MS, Gebhardt DL, Baker TA, *et al*. Performance requirements of physically strenuous occupations: validating minimum standards for muscular strength and endurance. *Ergonomics* 2004;47:864–75.
11 Jamnik V, Gumienak R, Gledhill N. Developing legally defensible physiological employment standards for prominent physically demanding public safety occupations: a Canadian perspective. *Eur J Appl Physiol* 2012;113:2447–57.
12 Department of Defense Joint Technical Coordinating Group. Summary Report: Research Workshop on Physical Fitness Standards and Measurements within Military Services; August 31 to 2 September. Herndon (VA): Department of Defense, 1999:3–9.
13 U.S. Army Public Health Command (USAPHC). Public Health Report No. 12-01-0614. Prepared by Jones BH, Nindl B, Hauret K, *et al*. Development of a New Army Standardized Physical Readiness Test: 2012 through 2013. Aberdeen Proving Ground (MD): USAPHC, 2015:1–8.
14 General Accounting Office. *Gender issues: improved guidance and oversight are needed to ensure validity and equity of fitness standards*. Washington (DC): Subcommittee on Readiness, Committee on Armed Services, 1998:4–6, 27, 29.
15 Epstein Y, Yanovich R, Moran DS, *et al*. Physiological employment standards IV: integration of women in combat units physiological and medical considerations. *Eur J Appl Physiol* 2013;113:2673–90.
16 Reilly T. Canada's physical fitness standard for the land force: a global comparison. *Can Army* 2010;13:59–69.
17 US Army Public Health Command. Public Health Report No. 12-02-0614. Correlations Between Physical Fitness Tests and Performance of Military Tasks: A Systematic Review and Meta-Analyses. Prepared by Hauschild VD, DeGroot D, Hall S, *et al*. Aberdeen Proving Ground (MD): USAPHC, 2014. Available from DTIC: http://www.dtic.mil/docs/citations/ADA607688
18 Equal Employment Opportunity Commission. Uniform guidelines on employee selection procedures (1978 as revised 2003). *Fed Reg* 2003;43:38289–315.
19 Caspersen CJ, Powell KE, Christenson GM. Physical activity and physical fitness: definitions and distinctions for health-research. *Public Health Rep* 1985;100:126–31.
20 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
21 Garg AX, Hackman D, Tonelli M. Systematic reviews and meta-analysis: when one study is just not enough. *Clin J Am Soc Nephrol* 2008;3:253–60.
22 Von Elm E, Altman DG, Egger M, *et al*., Strobe Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573–7.
23 US Army Center for Health Promotion and Preventive Medicine (USACHPPM). Technical Report No 12-HF-01Q9D-04. The case for pre-enlistment physical fitness testing: research and recommendations. Prepared by Knapik JJ, Jones BH, Sharp MA, *et al*. Aberdeen Proving Ground (MD): USACHPPM, 2004: 100 p. Available from DTIC. http://www.dtic.mil/get-tr-doc/pdf?AD=ADA426848
24 Department of the Army. *Soldier's manual of common tasks warrior skills level 1*. Soldier Training Publication (STP) 21-1. Chapter 3. Washington (DC): Headquarters Department of the Army, 2011:3–233.
25 North Atlantic Treaty Organization. *Optimizing operational physical fitness*. North Atlantic Treaty Organization Research and Technology Organization, 2009. http://www.dtic.mil/dtic/tr/fulltext/u2/a502544.pdf (accessed online October 2016).
26 Michaelides MA, Parpa KM, Henry LJ, *et al*. Assessment of physical fitness aspects and their relationship to firefighters' job abilities. *J Strength Cond Res* 2011;25:956–65.
27 Hoffman and Associates. Physical Readiness Validation for the Nevada POST Category III, Final Report. 6 March 2009. P A1–17; p E4.
28 Arvey RD, Landon TE, Nutting SM. Development of physical ability tests for police officers: a construct validation approach. *J Appl Psychol* 1992;77:996–1009.
29 Cooper H. *Research synthesis and meta-analysis: a step by step approach*. 4th edn. Washington (DC): Duke University, SAGE Publications, 2009:1–22, 145–197.

30 Mood DR, Morrow JR. *Introduction to statistics in human performance*. Scottsdale (AZ): Holcomb Hathaway Publishers, 2015:92–9.
31 Hemphill JF. Interpreting magnitudes of correlation coefficients. *Am Psychol* 2003;28:78–80.
32 Wright JE, Sharp DS, Vogel JA, *et al*. *Assessment of muscle strength and prediction of lifting capacity in US army personnel*. Natick (MA): US Army Research Institute of Environmental Medicine (available from DTIC), 1984:46.
33 Robertson DW, Trent T. *Documentation of muscularly demanding job tasks and validation of an occupational strength test battery (STB)*. Natick (MA): US Army Research Institute of Environmental Medicine (available in DTIC), 1985:101.
34 Beckett MB, Hodgdon JA. *Lifting and carrying capacities relative to physical fitness measures*. San Diego (CA): Naval Health Research Center, 1987:49.
35 Mello RP, Damokosj AI, Reynolds KL, *et al*. The physiological determinants of load bearing performance at different march distances. Technical report T15–88. DTIC. Natick (MA): Army Research Institute of Environmental Medicine, April 1988:1–25. http://www.dtic.mil/docs/citations/ADA197733
36 Stevenson JM, Andrew GM, Bryant JT, *et al*. Isoinertial tests to predict lifting performance. *Ergonomics* 1989;32:157–66.
37 Schonfeld BR, Doerr DF, Convertino VA. An occupational performance test validation program for fire fighters at the Kennedy Space Center. *J Occup Med* 1990;32:638–43.
38 Singh M, Lee SW, Wheeler G, *et al*. *Task related physical fitness and performance standards for the Canadian army*. Canada: University of Alberta, 1991:569.
39 Stevenson JM, Bryant JT, Andrew GM, *et al*. Development of physical fitness standards for Canadian Armed Forces younger personnel. *Can J Sports Sci* 1992;17:214–21.
40 Myhre LG, Tucker DM, Bauer DH, *et al*. *Relationship between selected measures of physical fitness and performance of a simulated fire fighting emergency Task*. Brooks Air Force Base (TX): Armstrong Lab (available in DTIC), 1997:1–26.
41 Kraemer WJ, Nindl BS, Gotshalk LA, *et al*. *Prediction of military relevant occupational tasks in women from physical performance components. Advances in Occupational Ergonomics and Safety*. Washington (DC): IOS Press, 1998:719–22.
42 Knapik JJ, Harper W, Crowell HP. Physiological factors in stretcher carriage performance. *Eur J Appl Physiol Occ Physiol* 1999;79:409–13.
43 Williford HN, Duey WJ, Olson MS, *et al*. Relationship between firefighting suppression tasks and physical fitness. *Ergonomics.* 1999;42:1179–86.
44 Deakin JM, Pelot R, Smith JT, *et al*. *Development and validation of Canadian forces Minimum Physical Fitness Standard (MPFS 2000)*. Kingston (ON): Ergonomics Research Group, Queen's University, 2000. Chapters 3 and 4, 9–36; Chapter 5, 45–9.
45 Pandorf CE, Harman EA, Frykman PN, *et al*. Correlates of load carriage and obstacle course performance among women. *Work* 2002;18:179–89.
46 Bilzon JL, Scarpello EG, Bilzon E, *et al*. Generic task-related occupational requirements for Royal Naval personnel. *Occup Med (Lond)* 2002;52:503–10.
47 Rhea MR, Alvar BA, Gray R. Physical fitness and job performance of firefighters. *J Strength Cond Res* 2004;18:348–52.
48 Barnes JL, Schilling BK, Falvo MJ, *et al*. Relationship of jumping and agility performance in female volleyball athletes. *J Strength Cond Res* 2007;21:1192–6.
49 Harman EA, Gutekunst DJ, Frykman PN, *et al*. Prediction of simulated battlefield physical performance from field-expedient tests. *Mil Med* 2008;173:36–41.
50 Williams-Bell FM, Villar R, Sharratt MT, *et al*. Physiological demands of the firefighter Candidate Physical Ability Test. *Med Sci Sports Exerc* 2009;41:653–62.
51 McBride JM, Blow D, Kirby TJ, *et al*. Relationship between maximal squat strength and five, ten, and forty yard sprint times. *J Strength Cond Res* 2009;23:1633–6.
52 Phillips M, Petersen A, Abbiss CR, *et al*. Pack hike test finishing time for Australian firefighters: pass rates and correlates of performance. *Appl Ergon* 2011;42:411–8.
53 Aandstad A, Holme I, Berntsen S, *et al*. Validity and reliability of the 20 meter shuttle run test in military personnel. *Mil Med* 2011;176:513–8.
54 Thebault N, Leger LA, Passelergue P. Repeated-sprint ability and aerobic fitness. *J Strength Cond Res* 2011;25:2857–65.
55 Jones BH, Hauschild VD. Physical training, fitness, and injuries: lessons learned from military studies. *J Strength Cond Res* 2015;29(Suppl 11):S57–64.
56 Tharion WJ, Lieberman HR, Montain SJ, *et al*. Energy requirements of military personnel. *Appetite* 2005;44:47–65.
57 Blacker SD, Wilkinson DM, Bilzon JL, *et al*. Risk factors for training injuries among British Army recruits. *Mil Med* 2008;173:278–86.
58 Vanderburgh PM. Occupational relevance and body mass bias in military physical fitness tests. *Med Sci Sports Exerc* 2008;40:1538–45.
59 Reuter SE, Massy-Westropp N, Evans AM. Reliability and validity of indices of hand-grip strength and endurance. *Austr Occupational Ther* 2010;58:82–7.
60 Vizcaya FJ, Viana O, Del Olmo MF, *et al*. Could the deep squat jump predict weightlifting performance? *J Strength Cond Res* 2009;23:729–34.

OEM

Fitness tests and occupational tasks of military interest: a systematic review of correlations

Veronique D Hauschild, David W DeGroot, Shane M Hall, Tyson L Grier, Karen D Deaver, Keith G Hauret and Bruce H Jones

Occup Environ Med published online November 3, 2016