Validity of energy expenditure estimation methods during 10 days of military training

Running header: Activity monitoring in military training

Andrew G. Siddall[1*], Steven D. Powell[1], Sarah C. Needham-Beck[1], Victoria C. Edwards[1], Jane E. S. Thompson[1], Sarah S. Kefyalew[2], Priya A. Singh[2], Elise R. Orford[2], Michelle C. Venables[2], Sarah Jackson[3], Julie P. Greeves[3], Sam D. Blacker[1], Steve D. Myers[1].

[1]Occupational Performance Research Group, University of Chichester, Chichester, UK

[2]Medical Research Council Elsie Widdowson Laboratory, Cambridge, UK

[3]Army Personnel Research Capability, Army Headquarters, Andover, UK

*CORRESPONDING AUTHOR

Email: A.Siddall@chi.ac.uk

ABSTRACT

Wearable physical activity (PA) monitors have improved the ability to estimate free-living total energy expenditure (TEE) but their application during arduous military training alongside more well-established research methods has not been widely documented. This study aimed to assess the validity of two wrist-worn activity monitors and a PA log against doubly-labelled water (DLW) during British Army Officer Cadet (OC) training. For 10 days of training, twenty (10 male and 10 female) OCs (mean ± SD: age 23 ± 2 years, height 1.74 ± 0.09 m, body mass 77.0 ± 9.3 kg) wore one research-grade accelerometer (GENEActiv, Cambridge, UK) on the dominant wrist, wore one commercially-available monitor (Fitbit SURGE, USA) on the non-dominant wrist and completed a self-report PA log. Immediately prior to this 10-day period, participants consumed a bolus of DLW and provided daily urine samples, which were analysed by mass spectrometry to determine TEE. Bivariate correlations and limits of agreement (LoA) were employed to compare TEE from each estimation method to DLW. Average daily TEE from DLW was 4112 ± 652 kcal·day^{-1} against which the GENEActiv showed near identical average TEE (mean bias ± LoA: -15 ± 851 kcal·day^{-1}) while Fitbit tended to underestimate (-656 ± 683 kcal·day^{-1}) and the PA log substantially overestimate (+1946 ± 1637 kcal·day^{-1}). Wearable physical activity monitors provide a cheaper and more practical method for estimating free-living TEE than DLW in military settings. The GENEActiv accelerometer demonstrated exceptional validity and could be useful for assessing TEE in large-scale, longitudinal military studies.

KEY WORDS: Doubly-labelled water; Wearable technology; Physical activity, Army; Accelerometry

INTRODUCTION

In military populations, measurement of the physical activity (PA) profile of personnel is important for monitoring health and training outcomes. Quantifying energy expenditure (EE) can inform evidenced-based interventions to optimise training volume, recovery, management of energy availability and injury risk mitigation strategies. Military training involves highly arduous physical exercise, unusual field-based activities such as heavy load carriage, digging and casualty extraction in addition to types of technical drill and weapons handling. The scope of unique activities performed in a range of environments, sometimes during periods of energy deficit and sleep disruption, mean it is challenging for investigators to employ experimental techniques required to accurately determine physical demand.

The doubly-labelled water (DLW) method is well-established as a 'gold-standard' process for determining free-living total EE (TEE) in humans [1]. The DLW technique has previously been used to quantify TEE in military cohorts (of approximately 19.6-19.8 $MJ.day^{-1}$ per individual (4380-4550 $kcal.day^{-1}$) [2]. However, the DLW method imposes significant challenges to investigators such as high financial cost, requirement for specialist materials, staff and analysis and participant burden which means that it can only be feasibly administered in small group samples over a short time period. Recent advances in wearable technologies have improved the ability to estimate free-living TEE in humans while limiting financial cost and user burden, and may be a solution to objectively assessing TEE in larger military cohorts [2].

Research-grade activity monitors that use movement data alone (i.e. accelerometers) have demonstrated varied success when compared to the DLW method, with TEE prediction models ranging from weak to strong (R=0.13-0.86) [3]. Accelerometers have shown efficacy when distributed to large military cohorts for physical demands monitoring [2,4,5]. However, research in military settings has led some researchers to caution that activities such as loaded marching or weapons handling could be misclassified as other movements or misinterpreted by TEE estimation algorithms as these are derived from typical human movements in the general population [6]. Multi-sensor activity monitors, which attempt to improve TEE estimation by combining accelerometry with physiological monitoring (e.g. heart rate), are available as relatively inexpensive consumer-grade monitors ranging to sophisticated research tools. Research-grade multi-sensor tools have been shown to improve TEE estimation over accelerometry alone [7,8] and demonstrate good agreement with criterion measures of TEE [9]. However, more-affordable consumer-grade monitors have shown varied validity based on the output variables analysed (e.g. steps, active minutes) and activity intensity (e.g. sedentary, moderate, vigorous) [10,11].

The large cohort sizes often studied in the military setting have resulted in researchers adopting relatively low-cost alternatives to DLW and activity monitors such as self-report logging of PA [5,12]. The use of self-report PA can introduce potential error via subjectivity and recall bias [5,13,14]. While objective measurement of activity using wearable activity monitors may seem a viable solution to these barriers, many have been designed specifically for the general population and for use by an individual user. Therefore, the comparative efficacy of using different methods of PA monitoring in a military environment remains unclear. In addition to data validity, a monitor's physical robustness and ability to handle and give easy access to data from large cohorts are vital considerations for suitability in this setting. It was hypothesised that the agreement between TEE estimated from DLW during a 10-day military training period and estimates from a research-grade wrist-worn accelerometer would be superior than estimates from a wrist-worn multi-sensor consumer-grade activity monitor and a self-report PA log.

METHODS

Study design

During 10 days of military training the DLW technique was used to measure TEE in 20 British Army Officer Cadets (OCs) at the Royal Military Academy Sandhurst (RMAS), UK. During the same 10 days, participants also wore two wrist-mounted physical activity monitors – a research-grade accelerometer (GENEActiv (Original), Activinsights Ltd., Cambridge, UK) and a multi-sensor consumer-grade monitor (Fitbit Surge HR, Fitbit, USA) and completed a daily PA log. The observed training period encompassed a selection of typical military activities, including classroom-based lessons, physical training, technical drills and field-based exercise such as combat training and load carriage. After a written and verbal brief participants provided written consent to take part in the study. The investigation was approved by the Ministry of Defence Research Ethics Committee (MoDREC; 780/MoDREC/16).

Preliminary measures

Body mass (Aria® scales, Fitbit, USA) and stature (Leicester Stadiometer, Seca, Hamburg, Germany) were measured at the beginning of the data collection period. Participants were each given the Fitbit to wear on their non-dominant wrist (as it could also act as a watch) and a GENEActiv to be worn on the dominant wrist. These wrist allocations were performed to reduce participant burden of wearing two devices.

Doubly-labelled water

The DLW method used in the present study has been described previously [15]. Briefly, on the evening prior to the 10-day collection period, participants provided baseline urine samples before consuming a measured bolus of hydrogen (deuterium ^{2}H) and oxygen (^{18}O) stable isotopes as water ($^{2}H_2{}^{18}O$). The dose was calculated to provide 150-180 mg of ^{18}O per kg of body mass and 50-80 mg of ^{2}H per kg of body mass. Post-dose urine samples were obtained for the subsequent 10 days, avoiding the first void of each day. Urine samples were frozen at -20°C to be stored for later analysis by an independent laboratory (Medical Research Centre Elsie Widdowson Laboratory (MRC EWL), Cambridge, UK). Isotope disappearance rates were determined through mass spectrometric analysis and used to calculate TEE using the multi-point method described previously [15] and where respiratory quotient was assumed to be 0.85 for all participants.

Research-grade accelerometer

The GENEActiv (Original) is a wrist-worn tri-axial seismic acceleration sensor, with a sensitivity level of ± 8 g. Accelerometers were configured for each user using GENEActiv software version 3.1 (Activinsights, Cambridge, UK) by inputting age, body mass, height and whether the monitor is worn on the dominant or non-dominant hand. Raw acceleration data were collected at 100 Hz and converted to summarise data over 60-s data epochs. The gravity-subtracted sum of vector magnitudes (SVM) for each minute were analysed using a macro-spreadsheet available from Activinsights to estimate metabolic equivalents (METs) using thresholds (Table 1) previously validated for GENEActiv accelerometers [16]. These were summed for each training day to produce MET minutes (MET·mins). In addition, sum of minutes spent in 'sleep' according to GENEActiv monitors were summed for each day. Minutes per day with zero values were replaced with 0.9 METs to establish a low baseline of estimated metabolism. The summed MET.mins were converted to estimated kilocalories using equation 1:

$$MET.mins \times 3.5 \times (BM/200) \qquad \text{(Equation 1)}$$

Where BM is body mass in kg [17].

Table 1. Activity intensity level thresholds utilised in energy expenditure estimation methods

		TEE estimation tool	
Activity intensity level	MET guidelines	PA log (METs)	GENEActiv (SVM)
Sedentary	0.9 - 3.1	2.05	<386
Light	3.2 - 5.3	4.25	386 – 542
Moderate	5.4 - 7.5	6.45	542 – 1811
Vigorous	7.6 - 12.0	9.80	≥1811

Note: Activity levels and MET guidelines described previously [18]*. TEE is total energy expenditure, SVM is gravity-subtracted Sum of Vector Magnitudes at 100 Hz sampling frequency; METs are Metabolic Equivalents.*

Consumer-grade monitor

The Fitbit Surge HR is a multi-sensor monitor which has a digital clock user-interface and houses a tri-axial accelerometer, gyroscope, compass, ambient light sensor, global positioning system and photoplethysmographic heart rate monitor. In order to extract daily TEE data, Fitbit monitors were synchronised to individual accounts where participant characteristics (age, sex, body mass, height) were inputted to individualise EE and basal metabolic rate (BMR) estimation to each participant. Data were extracted using an online data management platform (Fitabase, San Diego, USA) in order to batch-download daily TEE for all monitors in $kcal \cdot day^{-1}$.

Physical activity log

Each day, participants completed a PA log which asked for amount of time spent per day asleep, sedentary and in light, moderate or vigorous activity. The instructions for how to define these activity thresholds and examples of activities that could fall into these categories were given to participants within the activity log (Table 2). The activity intensity levels were given a MET value at the central point of previously defined ranges (Table 1; [18]) and multiplied by the reported duration of activity to produce MET.mins from the PA log. As with the GENEActiv, equation 1 was used to convert MET.mins to kilocalories.

Table 2. Descriptions of activity intensity levels given in the physical activity log

Activity intensity level	Descriptions	Examples
Vigorous	Activities that require hard physical effort and cause rapid breathing and large increases in HR; too high or too intense to chat/converse.	Running, jogging, hiking/marching/patrolling (heavy load-webbing, weapon, Bergan), obstacle/assault courses, circuit training, cycling uphill, competitive team sports (football, rugby, hockey).
Moderate	Activities that require moderate physical effort and cause a noticeable increase in breathing or HR.	Hiking/marching/patrolling (light load e.g. webbing & weapon), walking briskly/marching/drill, lifting & carrying stores, digging, cycling (level), boxing (punch bag), reactive sports (cricket, tennis).
Light	Activities that involve effort but that do not cause an increase in breathing or HR.	Standing with kit, walking at a slow pace, getting washed – showering, ironing kit.
Sedentary	Activities that involve sitting or reclining on or off duty, getting to and from places via transportation, but does not include time spent sleeping. These activities do not require physical effort.	Sitting, lectures, relaxing, completing paperwork, studying, eating.

Exclusion criteria

Wear-time criteria were used to exclude specific days (per individual) if a monitor did not appear to be worn for sufficient duration on that day. A wear-time criterion of 75% of the 24-day was set for both activity monitors concurrent with previous research [19,20]. In addition, from any tool, if any 10-day mean extended beyond three standard deviations from the population mean, these were treated at outliers and removed from the analysis for that tool.

Statistical analysis

Calculations of energy expenditure from each tool and measures of central tendency and variance (i.e. means, standard deviations) were completed in Excel (Office 2016, Microsoft, USA) and statistical analyses were performed using SPSS version 23.0 (IBM, USA). Bivariate correlations (Pearson's) were performed between average daily TEE from the DLW method and each PA monitoring tool. Bland and Altman plots were constructed to assess the agreement between DLW and each other TEE estimation method, comprising mean bias and 95% limits of agreement (LoA) [21]. To further analyse the comparative agreement of the evaluated estimation tools, 95% equivalence testing was also performed [9,22]. In this analysis, if the 90% confidence intervals (CI) of the tool-measured mean are contained entirely within a given error zone of the criterion mean (in this case, ±10%) those measures are typically considered "significantly" equivalent. Paired t-tests were used to compare mean TEE estimation from each method individually against measurement from DLW. To compare all methods, a repeated-measures analysis of variance (ANOVA) with post-hoc Bonferroni correction was conducted on participants with data across all methods. Statistical significance was set at an alpha value of $p<0.05$.

RESULTS

Participants

Twenty (10 male and 10 female) OCs (mean ± SD: age 23 ± 2 years, height 1.74 ± 0.09 m, body mass 77.0 ± 9.3 kg) participated in the study. Exclusion criteria meant that one participant was removed from the GENEActiv analysis (insufficient wear-time), and eight participants were removed from the PA log (outliers, n=2; insufficient completion of log, n=6). Average daily wear-time was 88 ± 6% for the Fitbit and 87 ± 17 % for the GENEActiv.

Agreement against the doubly-labelled water method

Bland and Altman plots (Figure 1) show the agreement between estimated daily TEE from each estimation method against the criterion standard (DLW). The agreement between tools is illustrated using mean bias and 95% LoA. The research-grade accelerometer showed best agreement but moderate LoA with a mean bias ± 95% LoA of -15 ± 851 kcal·day^{-1}. Agreement with DLW was poorer for the Fitbit (-656 ± 683) but with the narrowest LoA. The PA log performed least well, substantially overestimating TEE in comparison to DLW with large LoA (1946 ± 1637 kcal·day^{-1}). Only the GENEActiv could be deemed statistically equivalent to the criterion measure (DLW), demonstrated by the 90% CI of the measured mean being contained within the recommended equivalence zone of ±10% of the criterion-measured mean (Figure 2).

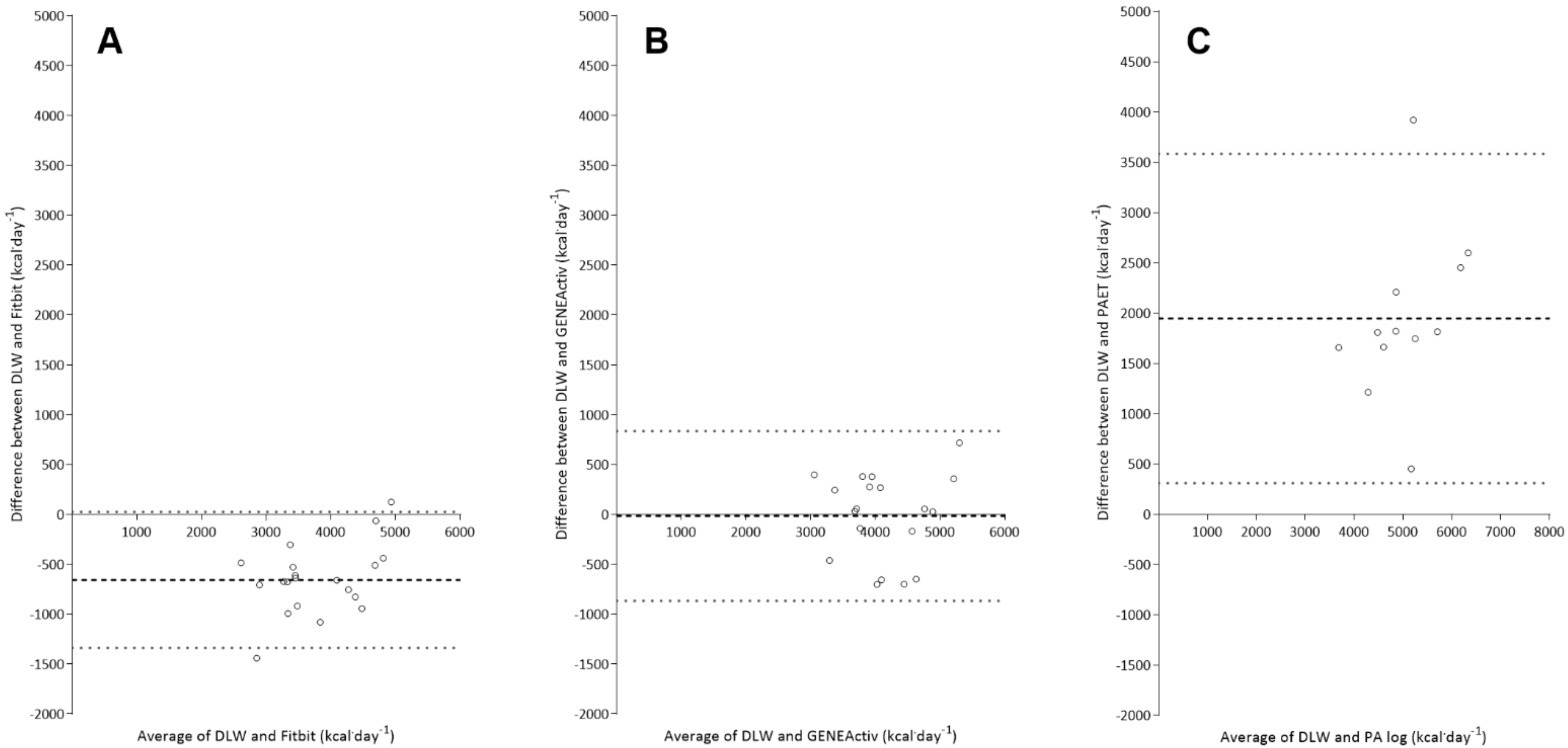

Figure 1. Bland-Altman plots for total energy expenditure estimation. Agreement (mean (black dashed line) ± 95% Limits of Agreement (LoA; grey dotted line)) between 10-day mean daily total energy expenditure (TEE) estimated from doubly-labelled water (DLW) and (A) Fitbit (n=20), (B) GENEActiv (n=19) and (C) PA Log (n=12)

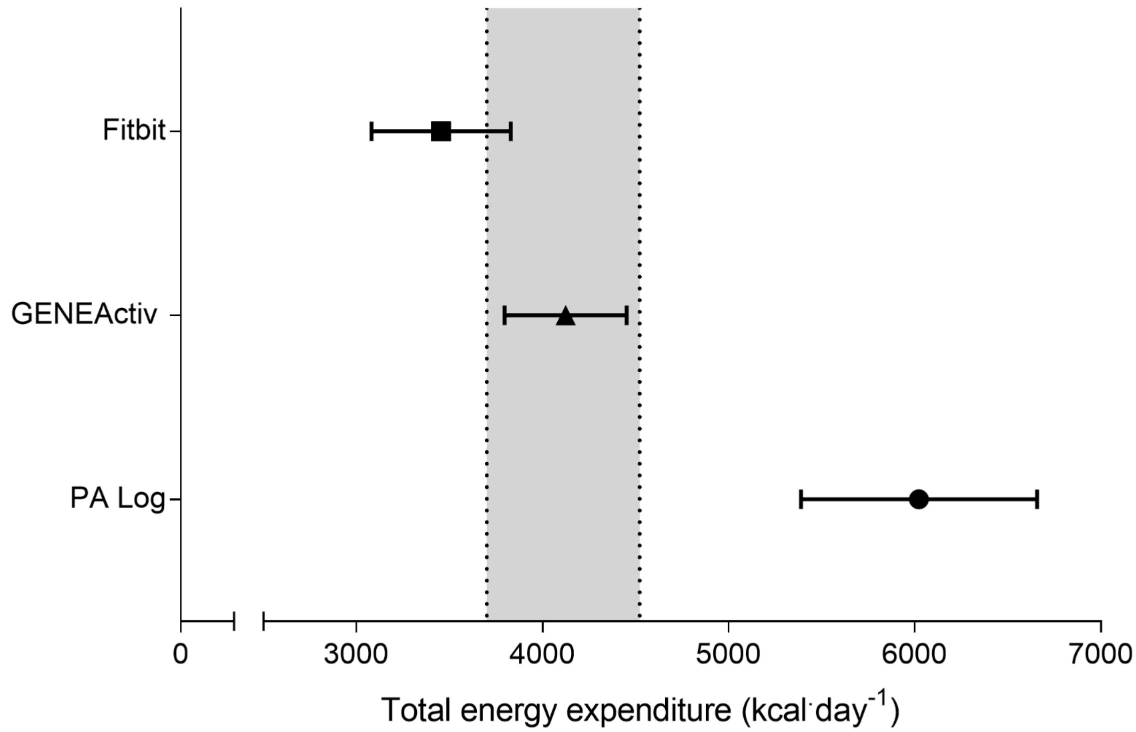

Figure 2. 95% equivalence testing of total energy expenditure. Equivalence test of each TEE estimation with 90% CI from Fitbit (Square), GENEActiv (Triangle) and PA Log (circle) against ±10% of DLW-estimated mean (grey shaded area).

Energy expenditure

The daily energy demand (mean ± SD) of the 10-day period from the DLW method was 4112 ± 652 kcal·day^{-1}. Figure 3 illustrates the average 24-hour EE from each estimation method and individual participant estimated 10-day means. Estimated TEE from both the Fitbit and the PA log differed significantly from DLW on individual comparison ($p<0.05$) and these results were corroborated by repeated-measures comparison between all methods via ANOVA using all participants with full data for each tool (n=11). Linear correlations between TEE from DLW demonstrated that the association between criterion measurement (Figure 4) and both the Fitbit ($r=0.904$, $r^2=0.817$, $p<0.01$) and GENEActiv ($r=0.790$, $r^2=0.624$, $p<0.01$) were stronger than with that of the PA log ($r=0.570$, $r^2=0.325$, $p>0.05$).

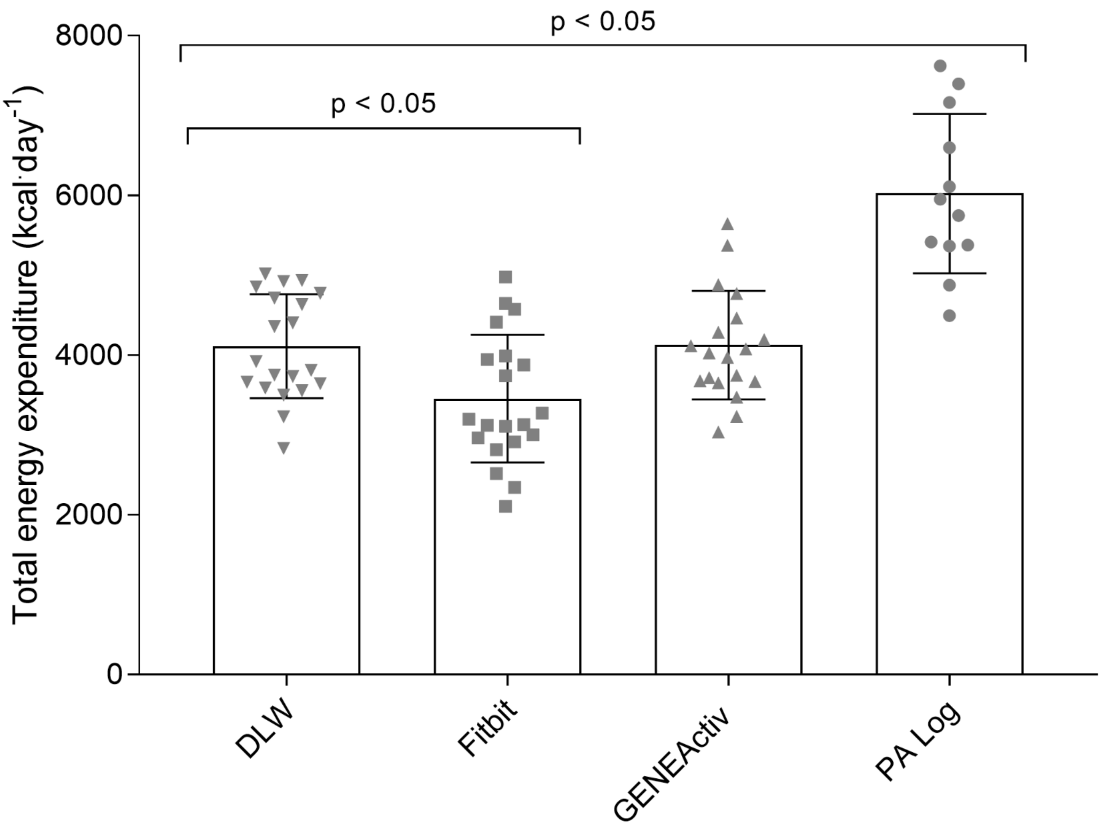

Figure 3. Average daily energy expenditure for each estimation method. Bars are means across the 10-day period computed from all participants for each tool, with error bars representing SD, and data points for each individual. Horizontal parentheses denote significant difference from criterion measurement (DLW; $p<0.05$).

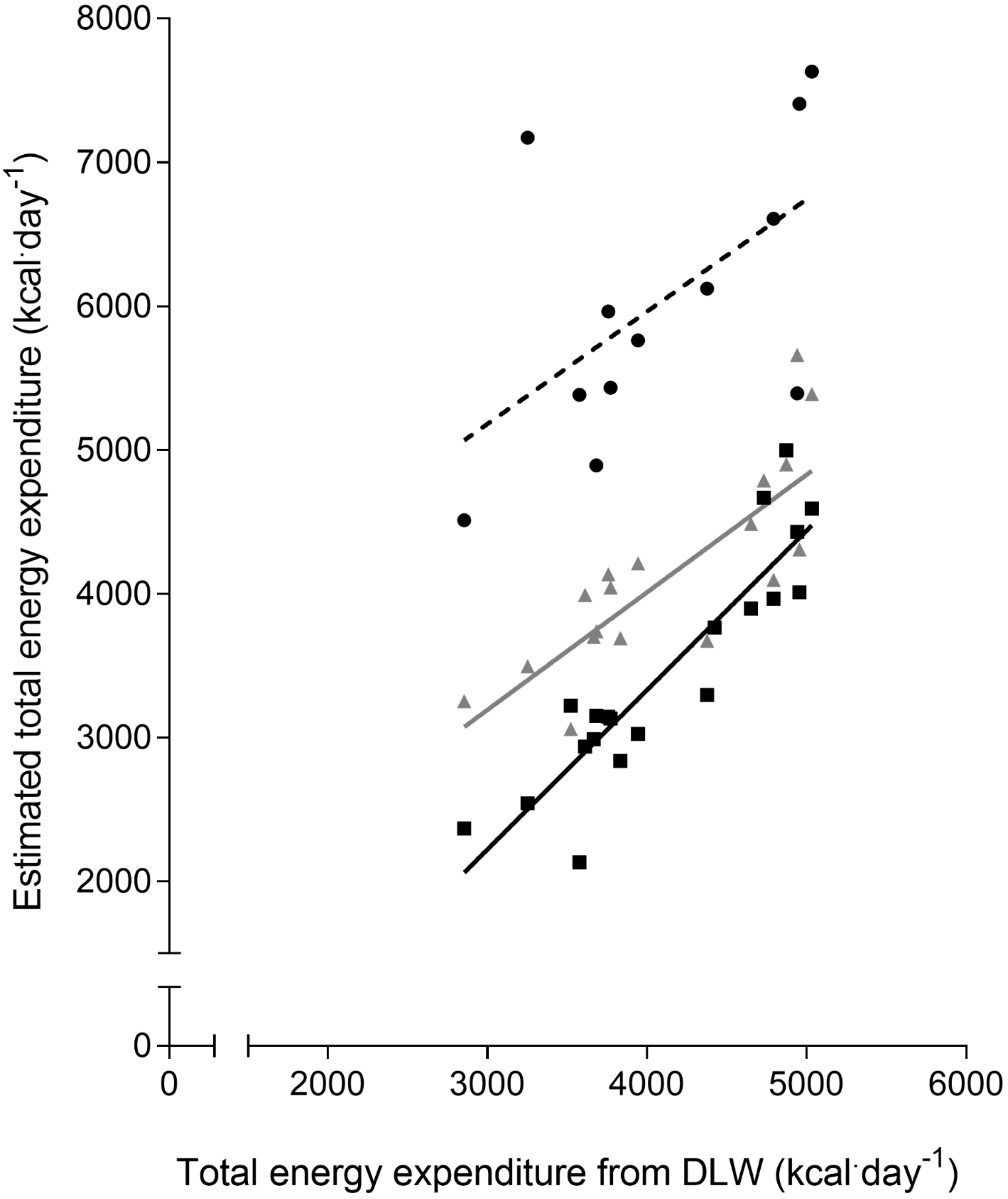

Figure 4. Correlational analysis between estimation methods. Average daily energy expenditure (kcal·day^{-1}) assessed by DLW against estimations by Fitbit (Black, squares; r=0.904, p<0.01), GENEActiv (Grey, upward triangles; r=0.790, p<0.01) and PA log (Black, circles; r=0.570, p>0.05) with lines of best fit.

DISCUSSION

This study examined the validity of three different methods to estimate TEE during military training by comparison with the 'gold-standard' DLW technique. The research-grade accelerometer was the most valid tool examined, exhibiting near identical group average TEE to DLW and with reasonable absolute agreement. In comparison to DLW, the consumer-grade activity monitor exhibited acceptable LoA but significantly underestimated TEE while the self-report PA activity log substantially overestimated TEE. These findings suggest that the research-grade activity monitor is sufficiently accurate for use during military training and a suitable alternative to DLW to measure TEE in this setting.

Accurately measuring the physical activity profile of military personnel in training or on operations is valuable for informing evidenced-based interventions to optimise training, quantify energy availability, and strategies to enhance recovery and mitigate injury risk. The present study is the first published use of the GENEActiv in a military population and supports previous findings of excellent validity of accelerometry-based TEE prediction algorithms in laboratory-controlled [23,24] and free-living conditions in civilian populations [25–27]. Results from this wrist-worn monitor are consistent with previous physical activity monitoring studies in the military using hip-mounted accelerometers, demonstrating practical suitability for large military cohorts [2,4,28] while capturing their activity with sufficient accuracy. Other research-grade accelerometers have been used in free-living conditions opposite DLW to successfully build EE predictions models [3], with several examples in military populations [2,6]. Our data suggest that the GENEActiv can also be used to provide objective measurement of the TEE of unique and arduous physical activity in military settings.

Within research-grade monitors, a multi-sensor approach typically improves TEE estimation over accelerometry alone. In laboratory trials, several models of the Fitbit have underperformed when compared to research tools, either by underestimation of HR and EE [29] or high inter-individual variation among similar tasks [9]. In free-living trials, Fitbits have demonstrated strong correlations with accelerometers but typically when analysing steps alone, and less accurately with absolute EE [10,30]. In the present investigation, the Fitbit was highly correlated with the criterion measurement with moderate limits of agreement, suggesting that subsequent corrections could be employed to make reparations for poor estimation of TEE but would require further investigation. Some of the inaccuracy in TEE could be explained by the heart rate detection technology (photoplethysmography) employed by the Fitbit which uses a light emitting diode positioned at the back of the wrist measuring tissue light propagation changes to detect blood flow. At moderate intensities of activity this has shown good agreement with chest-mounted heart rate monitors in laboratory trials, but validity is poorer at higher intensities of exercise [31]. In addition, accuracy, reliability and detection itself depends on wear tightness, position and other factors that could be violated in other testing environments [32,33]. Justifiably, the algorithms used by Fitbit or other large-scale device manufacturers are not freely available and so it is not possible to determine to what extent heart rate detection influenced overall TEE estimation. All participants in the present study were instructed on how best to position and wear the monitors. However, these participants are a realistic and representative sample of military personnel who would, notionally, wear the monitor in this manner. Therefore, any loss of estimation accuracy and data fidelity that did occur would likely be carried over into a larger-scale cohort.

Concurrent with several previous studies, the self-report methods for TEE estimation demonstrated low user-compliance, high inter-individual variability and overestimation of activity which has been observed in both civilian [34] and military populations [5]. Unfortunately, self-report methods inherently introduce subjectivity and can have a tendency to overestimate activity and underestimate sedentary time [34,35]. Previously, this has been explained as being a product of poor user compliance and recall bias [14] and of floor and ceiling effects, where responses cluster near the top or bottom of a particular variable (such as many hours sedentary and few minutes of vigorous activity), reducing variability in the data [13]. Participant burden can also cause boredom and inaccurate reporting in addition to participants wanting to give socially acceptable answers (i.e. reporting a high volume of highly intense training). While every effort was made for participants in the current study to complete the log daily and honestly, each of the above limitations to subjective profiling of physical activity may occur in these free-living settings. The current study used a relatively small sample size but also experienced low user adherence, resulting in data reduction of eight participants. Participants cited lack of time and difficulty remembering to complete paperwork during field-based training operations as reasons for lack of completion. Understandably, in comparison to wearing activity monitors, the completion of questionnaires represents a burden additional to the busy work and training schedule of OCs. If PA logging is required in future military studies, housing the questions on an electronic device with a notification service for questionnaire completion at specific, suitable times may improve compliance, but would not necessarily improve the overestimation of TEE observed.

Physical activity profiles from research-grade accelerometers are computed from raw acceleration data, from a combination of a) multiple, ranked thresholds where the summed magnitude of accelerations in a specific time-frame denote different intensities of movement and b) movement classification algorithms, which identify types of movement or action to either filter or retain for TEE estimation. Despite exhibiting equivalence to criterion-measurement in the current study, researchers have raised concerns that wrist-worn accelerometers may present barriers to accurate estimation of TEE when studying military populations and could explain why the LoA were not narrower [6]. Specifically, hand movements such as weapons handling or drill and the action of carrying a rifle while running may be misinterpreted by activity monitors and limit their validity in military settings [6]. In addition, computation of physical activity data from GENEActiv raw acceleration files is based on activity thresholds derived from a civilian population with a range of habitual activity levels [16]. Despite these possible concerns, in the present study the GENEActiv remained suitably accurate in estimating TEE for the cohort over the 10-day time-frame but individualised outcomes may require further precision. Given that current TEE algorithms for GENEActiv are based on a non-military population but are freely accessible to researchers, with further data collection and further processing in military cohorts, it could be possible to adjust these algorithms specific to military activities and improve limits of agreement. This could be performed by adding an adjustment factor to TEE or by altering the activity intensity thresholds (sedentary, light, moderate, vigorous) based on the military group being monitored.

The military training environment has the advantage of being a free-living setting with some elements that are fixed (to some extent) across the population sample such as training routines, diet and working hours. While this could result in lower inter-individual variation in EE in comparison to civilian free-living studies, this also places importance on the estimation accuracy of the other factors that comprise TEE. Specifically, the thermal effect of feeding (dietary-induced thermogenesis; DIT) and BMR were not directly measurable in the present study, but DIT would be encompassed exclusively within the DLW method. Several previous activity monitoring studies, particularly laboratory-trials, have measured BMR via indirect calorimetry and estimated DIT via documenting caloric intake [3]. For both the GENEActiv and self-report methods, BMR estimation was not required since METs already account for resting metabolism by applying 1 MET per minute of sedentary behaviour. It is possible, though, that this may introduce errors by a lack of participant-specific individualisation. Similarly, applying METs to various activity thresholds does not account for differences in relative fitness which would be a prudent addition to a military-specific EE estimation algorithm in future. With this in mind though, the GENEActiv and Fitbit software use anthropometric data at the outset to personalise TEE estimation, and may have mitigated some of these issues.

From a practical perspective, research-specific tools are typically not designed to withstand heavy use in harsh, uncontrolled environments but more physically robust, affordable consumer-grade monitors may not achieve comparative accuracy [9]. Inspection by study researchers and participant feedback revealed that both wrist-worn monitors were generally robust in the military training environment but are not small enough or possess a low-enough profile from the wrist to avoid damage. Specifically, the brackets fastening the GENEActiv accelerometer to its wrist-strap are easily damaged and the Fitbit loses waterproofing when the screen is broken. In the current study, wear-comfort was not a concern for the majority of participants, but certain advantages become evident if monitors were worn individually, where the GENEActiv would allow an individual to wear their own watch on the alternate wrist, and the Fitbit has an interactive interface giving feedback to participants. The GENEActiv allows open access to raw data and handling with adaptable spreadsheets programmable by users, which allows researchers to interrogate data, data quality and customise analyses. However, without sufficient programming capability, data processing and handling would represent a significant undertaking in a larger, longer-term study. Despite the Fitbit housing a ‘black box’, commercially sensitive algorithm, access to the data management platform Fitabase does allow efficient on-mass download from multiple devices but only of computed daily summary data rather than raw data at the device’s sampling frequency.

The present study used the criterion measurement of TEE via DLW to assess the validity of three measurement tools to estimate TEE during 10 days of military training. The research-grade activity monitor demonstrated exceptional validity and practical suitability for use in the military setting and outperformed the consumer-grade activity monitor and PA log assessed. Therefore, the GENEActiv could be used in large-scale longitudinal studies in the military setting to quantify TEE to inform evidenced-based interventions to optimise training, quantify energy availability, and strategies to enhance recovery and mitigate injury risk.

PERSPECTIVE

While there has been substantial improvement in wearable physical activity monitors in recent years, their validity for estimating energy expenditure in unique and arduous training is under-researched, particularly in comparison to more well-established research techniques and in military populations. Previous activity monitoring in military settings have cautioned that movement patterns unique to the military may render data from accelerometry, and particularly wrist-worn devices, challenging to interpret[6], not comparable to direct observation[5] or in need of correction[2]. The current study directly compares multiple methods of energy expenditure estimation that could be applied in a field-setting to a criterion gold-standard and is also the first study to use the GENEActiv in a military context. The findings suggest this research-grade wrist-worn accelerometer is a valid and practical monitoring tool for this nature of training. This forms a basis for physical demands analyses and training load study in larger cohorts as well as the potential to define military-specific activity intensity levels, previously derived from sample from the general population[16], to improve limits of agreement against criterion measures.

ACKNOWLEDGEMENTS

This research was funded by the Army Personnel Research Capability (UK Ministry of Defence: Army) through the Defence Human Capability Science and Technology Centre (DHCSTC). The authors would like to acknowledge the staff at the Royal Military Academy Sandhurst, and the study volunteers.

REFERENCES

1. Shephard RJ, Aoyagi Y. Measurement of human energy expenditure, with particular reference to field studies: an historical perspective. Eur J Appl Physiol 2012;112:2785-2815.

2. Horner F, Bilzon JL, Rayson M, et al. Development of an accelerometer-based multivariate model to predict free-living energy expenditure in a large military cohort. J Sports Sci 2013;31:354-360.

3. Jeran S, Steinbrecher A, Pischon T. Prediction of activity-related energy expenditure using accelerometer-derived physical activity under free-living conditions: a systematic review. Int J Obes 2005 2016;40:1187-1197.

4. Wilkinson DM, Blacker SD, Richmond VL, et al. Injuries and injury risk factors among British Army infantry soldiers during predeployment training. Inj Prev J Int Soc Child Adolesc Inj Prev 2011;17:381-387.

5. Redmond JE, Cohen BS, Simpson K, et al. Measuring physical activity during US Army Basic Combat Training: a comparison of 3 methods. US Army Med Dep J December 2013:48-54.

6. Kinnunen H, Tanskanen M, Kyröläinen H, et al. Wrist-worn accelerometers in assessment of energy expenditure during intensive training. Physiol Meas 2012;33:1841-1854.

7. Brage S, Brage N, Franks PW, et al. Reliability and validity of the combined heart rate and movement sensor Actiheart. Eur J Clin Nutr 2005;59:561-570.

8. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: new insights and validation studies. Obes Rev Off J Int Assoc Study Obes 2013;14:451-462.

9. Chowdhury EA, Western MJ, Nightingale TE, et al. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PloS One 2017;12:e0171720.

10. Gomersall SR, Ng N, Burton NW, et al. Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry. J Med Internet Res 2016;18:e239.

11. Dominick GM, Winfree KN, Pohlig RT, et al. Physical Activity Assessment Between Consumer- and Research-Grade Accelerometers: A Comparative Study in Free-Living Conditions. JMIR MHealth UHealth 2016;4:e110.

12. Roy TC, Knapik JJ, Ritland BM, et al. Risk factors for musculoskeletal injuries for soldiers deployed to Afghanistan. Aviat Space Environ Med 2012;83:1060-1066.

13. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000;71 Suppl 2:1-14.

14. Ward DS, Evenson KR, Vaughn A, et al. Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 2005;37:S582-588.

15. Coward WA. Stable isotopic methods for measuring energy expenditure. The doubly-labelled-water (2H2(18)O) method: principles and practice. Proc Nutr Soc 1988;47:209-218.

16. Esliger DW, Rowlands AV, Hurst TL, et al. Validation of the GENEA Accelerometer. Med Sci Sports Exerc 2011;43:1085-1093.

17. Bushman B. How Can I Use METs to Quantify the Amount of Aerobic Exercise. ACSM Health Fit 2012;16:5-7.

18. Howley ET. Type of activity: resistance, aerobic and leisure versus occupational physical activity. Med Sci Sports Exerc 2001;33:S364-369; discussion S419-420.

19. Chinapaw MJM, Slootmaker SM, Schuit AJ, et al. Reliability and validity of the Activity Questionnaire for Adults and Adolescents (AQuAA). BMC Med Res Methodol 2009;9:58.

20. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015;47:839-842.

21. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-160.

22. Lee J-M, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014;46:1840-1848.

23. Kelly LA, McMillan DG, Anderson A, et al. Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. BMC Med Phys 2013;13:5.

24. Nightingale TE, Walhin J-P, Thompson D, et al. Influence of accelerometer type and placement on physical activity energy expenditure prediction in manual wheelchair users. PloS One 2015;10:e0126086.

25. Rowlands AV, Olds TS, Hillsdon M, et al. Assessing sedentary behavior with the GENEActiv: introducing the sedentary sphere. Med Sci Sports Exerc 2014;46:1235-1247.

26. Van Loo CMT, Okely AD, Batterham MJ, et al. Wrist Accelerometer Cut Points for Classifying Sedentary Behavior in Children. Med Sci Sports Exerc 2017;49:813-822.

27. Pavey TG, Gomersall SR, Clark BK, et al. The validity of the GENEActiv wrist-worn accelerometer for measuring adult sedentary time in free living. J Sci Med Sport 2016;19:395-399.

28. Ojanen T, Häkkinen K, Vasankari T, et al. Changes in Physical Performance During 21 d of Military Field Training in Warfighters. Mil Med 2018;183:e174-e181.

29. Wallen MP, Gomersall SR, Keating SE, et al. Accuracy of Heart Rate Watches: Implications for Weight Management. PloS One 2016;11:e0154420.

30. Gusmer RJ, Bosch TA, Watkins AN, et al. Comparison of FitBit® Ultra to ActiGraph™ GT1M for Assessment of Physical Activity in Young Adults During Treadmill Walking. Open Sports Med J 2014;8.

31. Jo E, Lewis K, Directo D, et al. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. J Sports Sci Med 2016;15:540-547.

32. Sazonov E, Neuman MR. Wearable Sensors: Fundamentals, Implementation and Applications. Elsevier, 2014.

33. Spierer DK, Rosen Z, Litman LL, et al. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39:264-271.

34. Wanner M, Probst-Hensch N, Kriemler S, et al. Validation of the long international physical activity questionnaire: Influence of age and language region. Prev Med Rep 2016;3:250-256.

35. Macfarlane DJ, Lee CCY, Ho EYK, et al. Convergent validity of six methods to assess physical activity in daily life. J Appl Physiol Bethesda Md 1985 2006;101:1328-1334.