Technical Report
1174

Soldier Protection Benchmark Evaluation (SPBE) Physiological Data Collection and Analysis

Fort Greely, Alaska
17 September–5 October 2012

T.B. Hughes
J.R. Williamson
A.R. Hess
W.T. Young
A. Dumas
K.D. Fischl
B.A. Telfer

19 August 2013

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Natick Soldier Research, Development, and Engineering Center under
Air Force Contract FA8721-05-C-0002.

This report is based on studies performed at Lincoln Laboratory, a federally funded research and development center operated by Massachusetts Institute of Technology. This work was sponsored by the Natick Soldier Research, Development, and Engineering Center (NSRDEC) under Air Force Contract FA8721-05-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Enterprise Acquisition Division

Massachusetts Institute of Technology
Lincoln Laboratory

Soldier Protection Benchmark Evaluation (SPBE) Physiological Data Collection and Analysis

Fort Greely, Alaska
17 September–5 October 2012

T.B. Hughes
J.R. Williamson
A.R. Hess
W.T. Young
A. Dumas
K.D. Fischl
B.A. Telfer
Group 48

Technical Report 1174

19 August 2013

Lexington Massachusetts

This page intentionally left blank.

ABSTRACT

The United States Army is continually evaluating and improving the form, fit, and function of protective equipment for the individual Soldier. To improve upon the evaluations, the Product Manager Soldier Protective Equipment (PM SPE) asked Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) to incorporate a physiological status monitoring capability into their tests. During the execution of the Army's Soldier Protection Benchmark Evaluation (SPBE), MIT LL outfitted 34 Soldiers with physiological monitors, accelerometers, and GPS's to collect objective physical performance data. The Soldiers performed a predefined set of activities wearing four different protective gear configurations. Data from the sensors supplied by MIT LL were analyzed to assess the effects of the various configurations on the physical performance of the Soldiers. Accounting for percent effort exerted, statistically significant (p-value < 0.01) differences in the speed and movement of Soldiers wearing the different configurations could be seen for a 5-kilometer road march and an obstacle course. During the 5-kilometer march, skin temperature was significantly higher for the more protective and encapsulating configurations. Normalizing for speed on the 5-kilometer march, significantly higher core temperatures were observed for heavier configurations. The use of on-body sensors to collect physiological data during the SPBE provided the Army with an objective dataset to use in their evaluation of the various configurations of protective gear.

This page intentionally left blank.

ACKNOWLEDGMENTS

The authors are grateful to Cheryl Stewardson, Blake Mitchell, Carolyn Bensel, and Leif Hasselquist from Natick Soldier Research, Development, and Engineering Center and the Program Executive Office Soldier (PEO Soldier) Product Manager Soldier Protective Equipment (PM SPE) for financial support and collaboration for this project. We would also like to thank the Maneuver Center of Excellence (MCoE) Maneuver Battle Laboratories (MBL), U.S. Army Research Laboratory (ARL) Human Research and Engineering (HRED), and Army Physical Fitness School for their collaboration and providing the opportunity to execute this research. Special recognition goes to the test participants, who were enlisted men and woman assigned to Fort Wainwright, AK. Our thanks also go out to all those Warfighters who serve our country.

This page intentionally left blank.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Page

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

LIST OF ILLUSTRATIONS
(Continued)

LIST OF TABLES

This page intentionally left blank.

1. INTRODUCTION

1.1 BACKGROUND

The Soldier Protection System (SPS) Army Personal Protective Equipment program has a goal of designing improved Soldier protective equipment [1], specifically in the areas of modularity, scalability, mission-tailoring, and equipment weight. Typically this equipment has been measured for effects on Soldier performance through timed physical activities (objective) and user surveys (subjective). The times have been used to compare different protective gear configurations for a given activity. The surveys have conveyed the feelings, thoughts, and opinions of the users about a particular configuration after completing an activity. The focus of these evaluations had always been form, fit, and function of the equipment.

To improve upon these evaluations, the Product Manager Soldier Protective Equipment (PM SPE) desires a new test methodology that not only includes the previous standard measures, but adds the ability to measure physiological and cognitive performance during specific Soldier tasks. Physiological parameters have not previously been collected by the Army in field environments for the purpose of evaluating the impact of protective gear on Soldier physical performance.

The new methodology also aims to collect a subjective and objective data set to serve as a possible benchmark for comparisons between current, standard-issue, and future protective gear. This program has brought together the efforts of PM SPE; Natick Soldier Research, Development, and Engineering Center (NSRDEC); Maneuver Battle Lab (MBL); U.S. Army Research Laboratory Human Research and Engineering Directorate (ARL-HRED); and Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) to collect the desired benchmark data.

1.2 OBJECTIVES

The role of MIT LL, as an independent investigator during the data collection, was guided by two objectives. The first was to collect objective physiological data sets (heart rate, respiration rate, core temperature, activity level, and body position) for all consenting Soldiers in the Soldier Protection Benchmark Evaluation (SPBE) fielded environment. The second was to determine the best physiological parameters to use as standard metrics in evaluating the impact of different levels of protection on Soldier performance. These parameters were identified by analyzing the collected data and assessing the performance under the different uniform configurations.

1.3 OVERVIEW

1.3.1 Study Description

A group of 34 Soldiers participated in a four-week protective equipment evaluation in Ft. Greely, Alaska, from September through October 2012. Over the course of the evaluation, each Soldier completed the same set of daily activities in four different uniform configurations, unless he or she was sick or injured. The configurations corresponded to different protection levels.

United States ARL-HRED conducted surveys at the conclusion of several of the activities to gather user feedback on the form, fit, function, and comfort of the different uniform configurations. Natick Soldier Research, Development, and Engineering Center timed the modified Army Combat Readiness Test (mACRT) and administered cognitive evaluations at the conclusion of physically taxing activities. MIT LL collected physiologic data on consenting Soldiers throughout each day and also timed the 5-kilometer march and the Urban Obstacle Lane.

1.3.2 Protection Levels[1]

The Soldiers wore four different configurations of current, standard-issue protective gear, shown in Figure 1.

Configuration A: Army combat uniform (ACU) trousers with Army combat shirt (ACS), Army combat helmet (ACH), weapon with sight, and eye protection

Configuration B: ACU trousers with ACS, ACH, weapon with sight, eye protection, plate carrier with front/rear plates, pelvic protection undergarment (PUG), and tactical assault panel (TAP)

Configuration C: ACU trousers with ACS, ACH, weapon with sight, eye protection, improved outer tactical vest (IOTV) with front/rear/side plates, and TAP with standard load

Configuration D: ACU trousers with ACS, helmet, weapon with sight, eye protection, IOTV with front/rear/side plates, TAP with standard load, deltoid protection system (DAPs), and protective pelvic outer garment (POG)

For the Foot March event, in addition to their assigned uniform configuration, Soldiers wore an assault pack with a 35 lb load of specific Soldier items.

---

[1] For a more complete explanation of each protection level, please see [1].

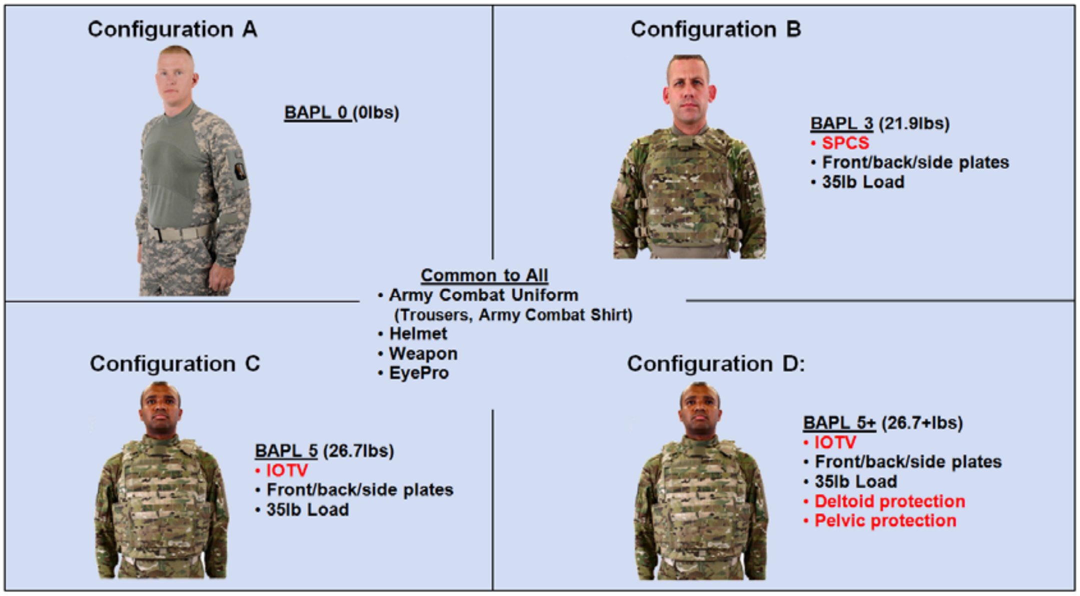


Figure 1. Equipment configurations used during SPBE testing. Upper left, Army combat uniform (ACU). Upper right, Soldier Plate Carrier System (SPCS). Bottom left, right, improved outer tactical vest (IOTV).

1.3.3 Daily Events[2]

On each of the four data collection days, the Soldiers performed the same sequence of events—two events in the morning, lunch, and then four events in the afternoon. Below are summary descriptions of each of the events. For further details, please see the "Soldier Protection Benchmark Evaluation (SPBE) results report, September 17–October 5, 2012." This is available from Natick Soldier Research, Development and Engineering Center.

Modified Army Combat Readiness Test (mACRT)[3] – a timed fitness course that consists of ten continuous events—a 200-meter run, low hurdles, high crawls, under and over, casualty drag, balance beam ammunition can carry, point-aim-move, 40-yard ammunition can shuttle sprint, agility sprint, and a 200-meter run. This was performed both in the morning and the afternoon. A diagram of the mACRT is given in Figure 2.

---

[2] For a more detailed description and images of each event, see [1].

[3] Activities of particular interest to the physiologic monitoring data collected by MIT LL.

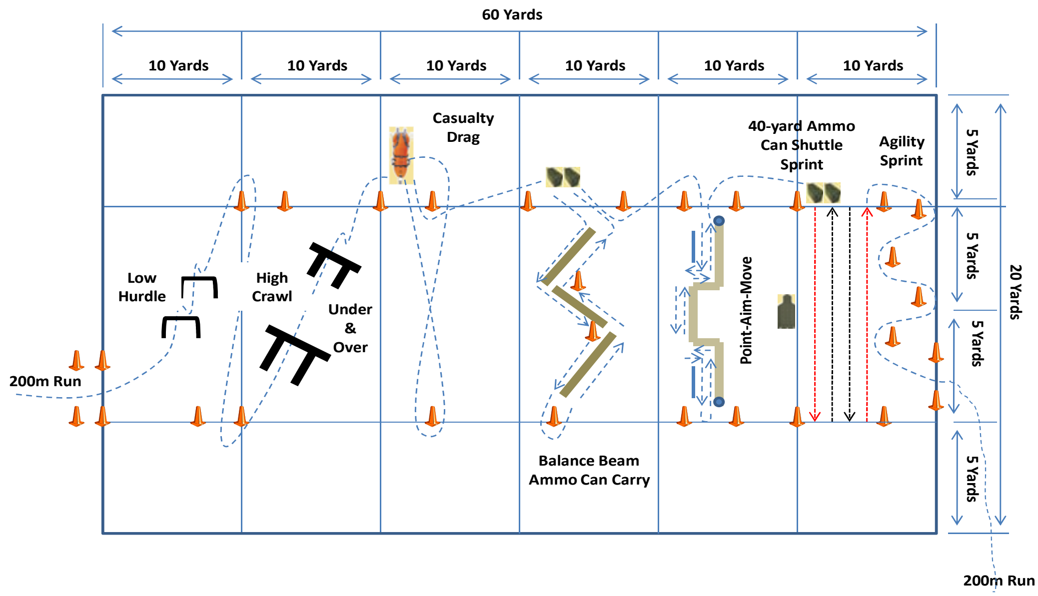


Figure 2. Modified Army Combat Readiness Test.

Round Robin – four activities to evaluate the given protective gear configuration—vehicle ingress/egress, weapons/equipment compatibility, don/doff, and casualty evaluation/drag. See reports produced by Army Research Laboratory (ARL) Human Research and Engineering Demand (HRED) and Maneuver Center of Excellence (MCoE) MBL on the Soldier Protection Baseline Evaluation for information on these topics.

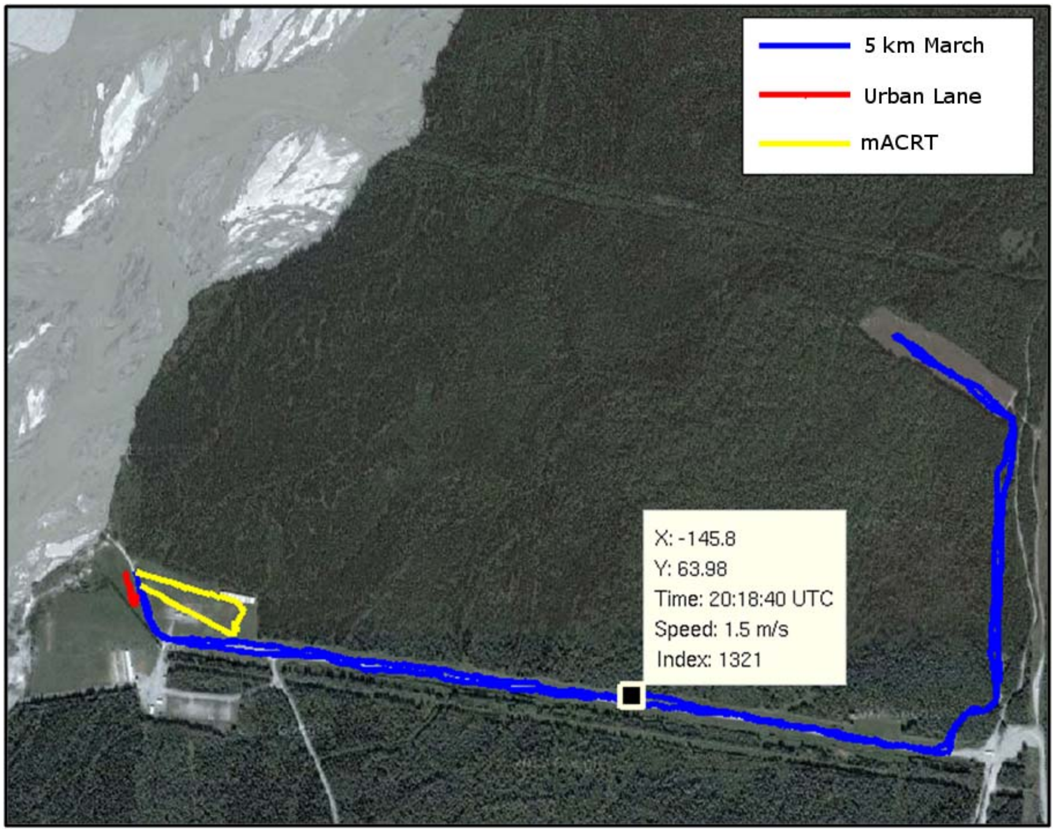


Figure 3. 5 km Foot March Route, Urban Lane Course, and mACRT Course.

5-kilometer Foot March[4] – a self-paced march, with the addition of a 35-pound assault pack to the assigned uniform configuration. Participants were instructed to complete the march as quickly as they were safely capable of doing, and all did so with each load configuration. The participants were grouped based on their configuration, and each group was spaced by 10–15 minutes. Individual start and finish times were recorded.

Urban Obstacle Lane – a timed obstacle course to simulate an urban environment containing six obstacles—4-foot wall, window, stairs, ladders, 2-foot incline/decline ramp, and a tunnel. The emphasis for these obstacles was on specific Soldier mobility tasks that were not covered during the mACRT.

mACRT[4] – a second run of the previously described fitness course.

---

[4] Activities of particular interest to the physiologic monitoring data collected by MIT LL.

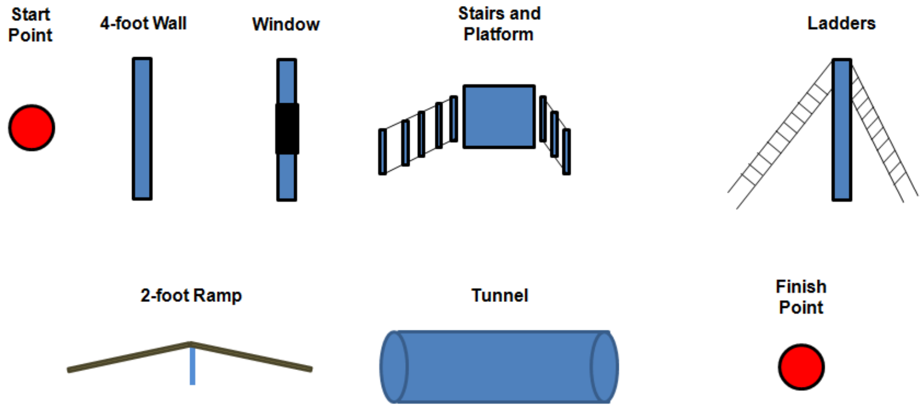


Figure 4. Urban obstacles.

Modified Short-Range Marksmanship (SRM) Course – assessment of shooting ability in assigned configuration from prone, kneeling, and standing (low-ready) positions on a 25-meter range.

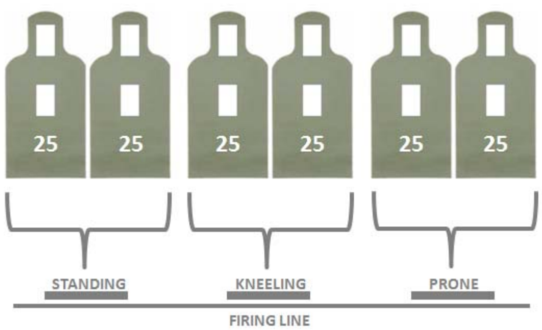


Figure 5. Modified Short-Range Marksmanship Course targets.

In addition to the daily events, the Soldiers also performed a range-of-motion assessment and emergency doff/reassemble activity in each configuration on one of their off days from the physical events.

1.3.4 Data Collection Instrumentation

Several pieces of monitoring equipment were used to gather physiological data on the Soldiers throughout the study. What follows is a brief description of each piece of equipment and what parameters it measures.

Equivital™ EQ02 LifeMonitor [2] – The Hidalgo Equivital™ EQ-02 is a FDA 510(k) certified system consisting of a torso-mounted belt and Sensor Electronics Module (SEM), as shown in Figure 6. The belt in the system measures physiological signals such as electrocardiogram (ECG) and respiratory rate. The SEM measures tri-axial acceleration, skin temperature, ambient temperature, heat flux, and core temperature. Core temperature measurement is facilitated by swallowing a MiniMitter Inc. capsule, as seen below in Figure 7, that wirelessly transmits core temperature data to the SEM. Embedded algorithms in the SEM produce several variables from the raw data, including heart rate (from ECG), body motion, and orientation (both from the accelerometer). The SEM has a battery life of up to 48 hours and has 8 GB of memory which allows approximately 50 days of data logging. At the start of the field exercise, each Soldier was issued a sized EQ-02 belt, and the raw ECG waveform was viewed by study coordinators to assess proper fit. For the duration of the study, SEM data was downloaded daily as well as stored on the SEM.

Equivital™ EQ02 SEM

Equivital™ EQ02 Chest Belt

Figure 6. Equivital™ EQ02 Life Monitor.

MiniMitter, Inc. Core Temperature Capsule [3] – An ingestible capsule (see Figure 7) that measures and transmits (to the SEM) core body temperature. The capsule is approximately 21.9 × 8.5 mm (length × width). In the morning before the day's events began, each consenting Soldier was given a capsule and cup of water. The MIT LL test team supervised the ingestion of the capsule and, using a handheld device, verified that it was transmitting data. At the conclusion of the test day, the same device was used to check if the capsule was still transmitting from each Soldier.

Figure 7. MiniMitter, Inc. core temperature capsule.

ActiGraph GT3X+ Monitor [4] – Each consenting participant wore an ActiGraph GT3X+ accelerometer, as seen in Figure 8, on each foot to record the acceleration experienced by the feet. This particular model is well known and commonly used for assessment of physical activity using accelerometry. The ActiGraph GT3X+ records acceleration along the X, Y, and Z axis and has an adjustable data rate of 30 Hz to 100 Hz. This study sampled at 100 Hz. Each sensor weighs 19 grams and has dimensions of 4.6 cm × 3.3 cm × 1.5 cm. They have a battery life of 30 days when fully charged and can record data continuously for nearly 7 days at 100 Hz before exceeding capacity. Before the start of each day's activities, the ActiGraph sensors were all initialized to start recording data at a predetermined time to facilitate time synchronization. Prior to beginning activity, participants were asked to attach one monitor onto the laces of each boot with Velcro straps (following a standard orientation on the boot), and they wore the monitors throughout the entire day. Study coordinators checked each sensor for proper orientation and secure attachment during equipment setup.

Figure 8. ActiGraph accelerometer.

Garmin Foretrex GPS[5]*/Qstarz BT-Q1000XT GPS* [5] – Soldiers were outfitted with one of two types of GPS data loggers, shown in Figure 9, for recording movement during daily activities. The devices sampled location every 2–4 seconds and placement was at the Soldier's discretion. The most common and recommended location was in an arm pocket. Other mounting locations included on the wrist, or in various pockets on the leg. These devices have a position accuracy of approximately 3 meters. Only 20 total GPS units were available for the 34 Soldiers, so the units were assigned to different Soldiers on different days. In other words, on test days where there were

17 Soldiers participating, each Soldier wore a GPS unit. On test days where there were 34 Soldiers participating, only 20 Soldiers were outfitted. Due to this constraint of a limited number of GPS units, daily downloading and initialization was required.

Garmin Foretrex GPS **Qstarz BT-Q1000XT GPS GPS**

Figure 9. GPS devices.

<u>*Kestrel Weather Station*</u> [6] – A small, easily portable weather station that measures wind direction, wind speed, temperature, wind chill, and relative humidity is shown in Figure 10. The collected weather data was used to compare the environmental conditions on the different test days. Additional weather data was collected via the Cold Regions Test Center weather station located at their Georgia Range. The Kestrel weather station was borrowed from USARIEM to support the SPBE tests.

Figure 10. Kestrel 4400 Weather Station.

This page intentionally left blank.

2. PARTICIPANTS

2.1 PROTOCOL

Due to the use of humans as experimental research subjects and the fact that personal, medically related data was to be collected, MIT LL completed a detailed Institutional Review Board protocol prior to the execution of the SPBE. The protocol outlined (a) the reason/motivation for the data collection; (b) the type of monitoring equipment that would be used, how it works, and what type of data it collects; (c) any risks involved; (d) how the volunteers and their personal information would be protected; and (e) the role MIT LL would be playing in the overall study. This protocol was submitted to and approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES) and the Army Human Research Protection Office (AHRPO).

2.2 BRIEF AND CONSENT

At the beginning of the study, prior to any activities taking place or data being collected, all of the Soldiers participating in the equipment evaluation were briefed about the physiological data collection portion of the study. Any questions they had about the study were answered, and finally they were asked to volunteer. No commanding officers were present while the Soldiers were asked to volunteer, and no forms of coercion were used to gain participants. There were no consequences for not volunteering to participate in the physiologic data collection, and Soldiers were free to withdraw themselves at any time. Soldiers could elect to (a) not participate at all, (b) participate in full, or (c) participate in parts of the physiologic data collection. Those who agreed to any kind of participation signed consent forms.

2.3 COMPOSITION OF SUBJECTS

Thirty-four active-duty Soldiers, ages 20 to 37, participated in the overall equipment evaluation study—30 males and 4 females. Of these, all but one male Soldier agreed to participate in the physiological data collection study in some capacity. The Soldiers' time in service ranged from 3 to 162 months and 28 to 42 months for the males and females, respectively. All four of the female participants had been deployed to a combat zone, Afghanistan in all cases. The ranks of the females included three E4s and one E5. Twenty-seven of the male participants had been deployed to a combat zone, either in Iraq or Afghanistan. The ranks of the males ranged from E2 to Officer, with the largest group (18 Soldiers) being E4s. Weights ranged from 132 to 255 pounds and 148 to 160 pounds for the males and females, respectively. Heights ranged from 66 to 78 inches and 62 to 67 inches for the males and females, respectively.

This page intentionally left blank.

3. METHODS

3.1 DATA COLLECTION

3.1.1 On-Body Sensors

Soldiers began each day by donning the appropriate equipment configuration for that day, as well as a set of sensing devices. The sensing devices, as outlined in Section 1.3.4, included wearing an Equivital EQ02 LifeMonitor, swallowing a MiniMitter VitalSense Core Temperature Capsule, attaching an ActiGraph Accelerometer to each boot, and wearing a Garmin GPS or QSTARZ GPS. Prior to donning any equipment, the Soldiers were asked to consent to participate and were allowed to opt out of wearing any device. Once outfitted, Soldiers participated in all daily activities as described in Section 1.3.3 while wearing that day's equipment configuration and the sensing devices.

Of the 34 Soldiers recruited to perform in the SPBE, one Soldier (subject 22) did not consent to participate in any of the events. One Soldier (subject 112) was injured after the first day of testing and therefore did not participant in the subsequent three days of testing. One Soldier (subject 117) only participated in three days of testing due to sickness on the fourth day. From the 33 participating Soldiers, 4 (subjects 111, 113, 117, and 23) opted out of taking a core temperature capsule on all four days of testing, and one Soldier (subject 214) opted out of taking a core temperature capsule on two of the four days of testing.

3.1.2 On-Body Sensor Data Quality

The ActiGraph and GPS data from Soldiers, by the nature of the devices' measurements, had very few discernible artifacts and were usable for processing. The core temperature capsule and the heart-rate data, with examples shown in Figure 11, contained artifacts and were graded on a green, yellow, and red scale to indicate artifact severity. Quantitatively, the colors correspond to the following percentages of data containing artifacts: green 0–20%, yellow 20–40%, and red 60–100%. The colors from green to red required increasing amounts of filtering or outlier removal. See Section 3.3.2 for the filtering method. In each of the plots of Figure 11 there are two colored lines. The red lines correspond to the raw collected data, and the blue lines correspond to the filtered version of the data that was used for processing.

The core temperature capsule had three failure modes, as seen in Figure 11 (left). Potential failure modes included the capsule dropping transmissions, the capsule completely terminating transmissions due to electronics failure, or the Soldier expelling the capsule before the end of activities. When hardware failure occurred, there was no potential for harm to the Soldier. Of the Soldiers who took a core temperature capsule, only one Soldier's (subject 215) data for one day was irretrievable. Of the data retrieved, only 4% of the data was missing due to communication issues between the SEM and the core temperature capsules. A summary of this information, as well as the quality of the collected data, is shown in Figure 12.

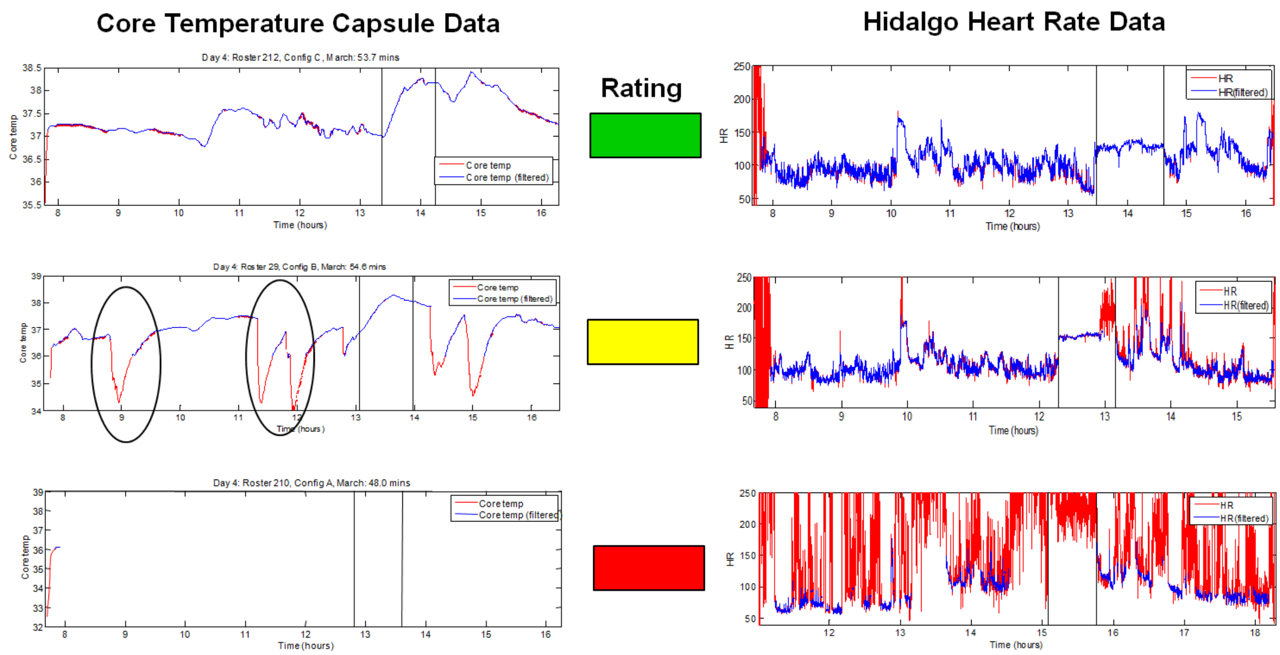


Figure 11. Examples of the data quality rating scale.

Data was retrievable from all Soldiers who wore the Equivital EQ02 LifeMonitor, and 89% of the heart-rate data reported by the LifeMonitor contained no disqualifying artifacts. Disqualifying artifacts potentially resulted from a bad connection between the device and the Soldier's skin, improper fitting of the device, or the general nature of the daily activities (like many of the activities in the mACRT) which caused the harness to be jostled and resulted in a poor connection with the Soldier's skin.

Figure 12. Overview of physiological data collected during the SPBE.

3.1.3 Weather Sensors

Weather data was collected from two weather stations during the main test days (September 24–28, October 1–5). The Kestrel weather station, set up by MIT LL, was located on the mACRT field. Data from the Georgia Range weather station at Ft. Greely was provided by the Cold Regions Test Center. Temperature data from both stations proved to be relatively consistent, while wind speed data was highly variable. The mACRT field was located near tree line, which could have affected wind direction and magnitude.

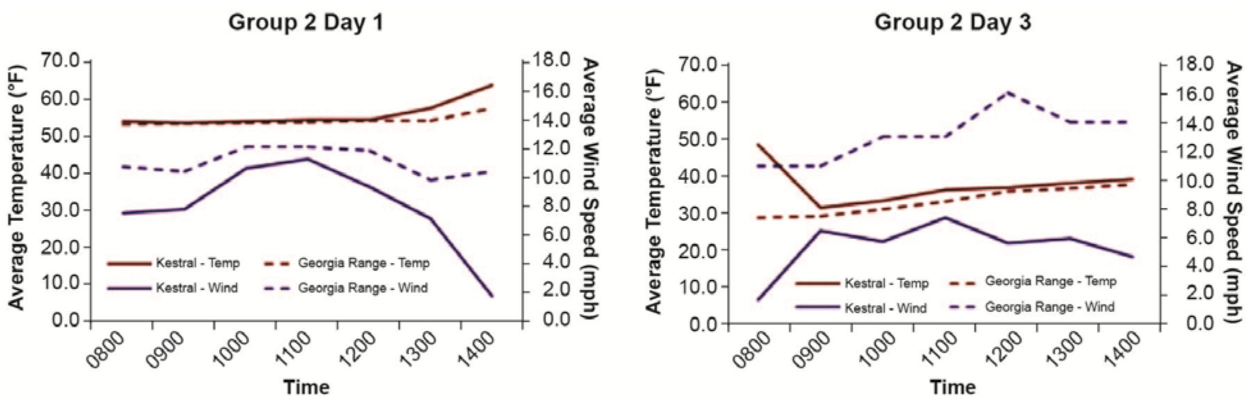


Figure 13. Weather instrument data (temperature and wind speed) comparison.

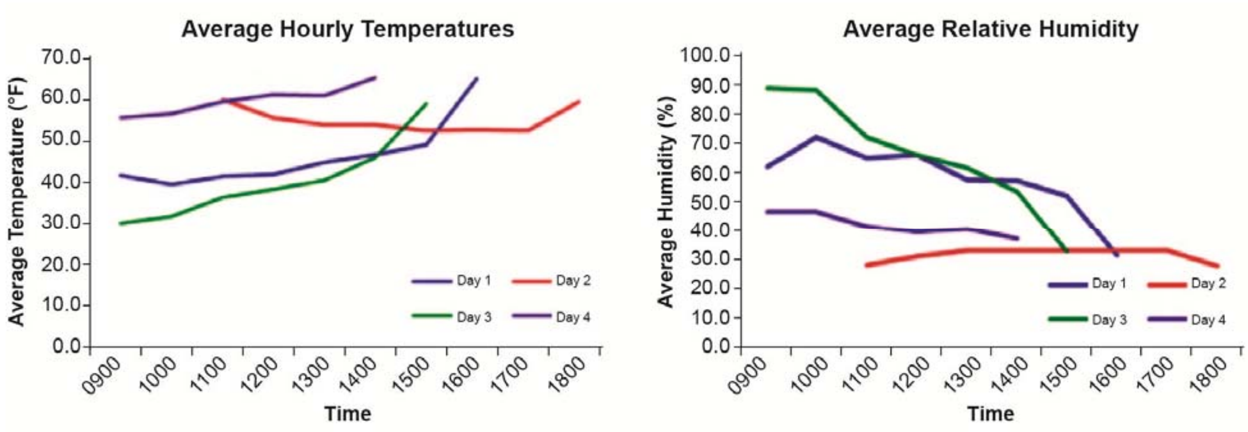


Figure 14. Kestrel weather station temperature and humidity comparisons for Group 1, across data collection days.

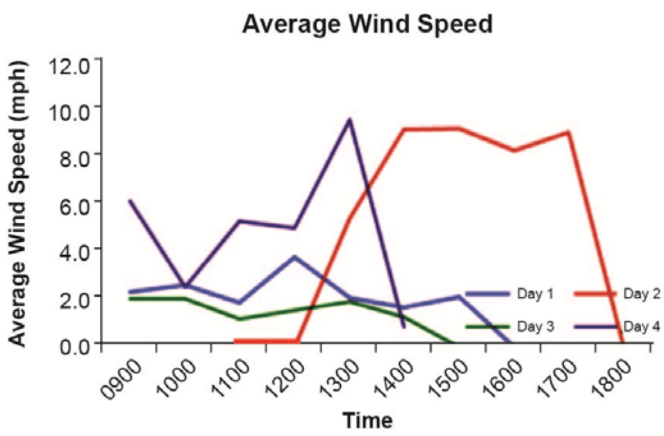
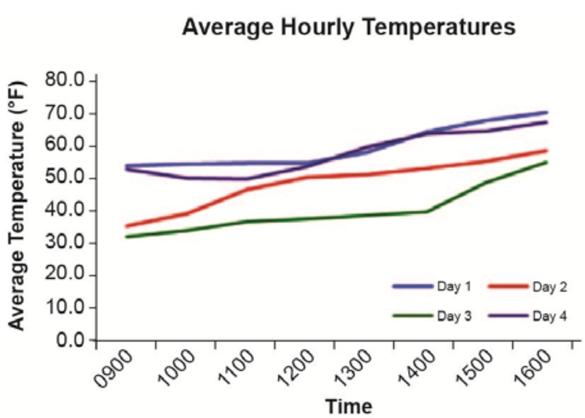


Figure 15. Kestrel weather station wind speed comparison for Group 1, across data collection days.

Referring to Figure 14 and Figure 15, Group 1 temperature data from Days 1 and 3 were comparable, humidity on Days 2 and 4 was similar, and wind speed varied significantly, but Days 1 and 3 were the closest.



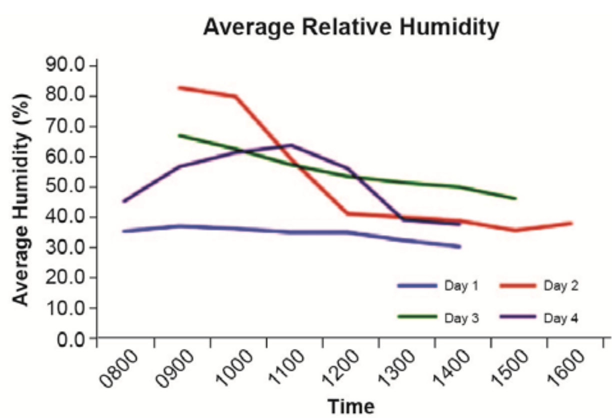


Figure 16. Kestrel weather station temperature and humidity comparisons for Group 2, across data collection days.

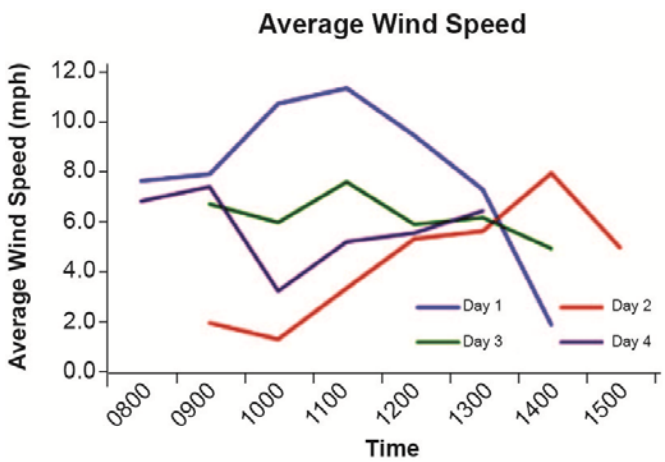


Figure 17. Kestrel weather station wind speed comparison for Group 2, across data collection days.

Referring to Figure 16 and Figure 17, temperature data from Days 1 and 4 were similar. Humidity on Days 2, 3, and 4 were reasonably consistent, and again the wind speed varied significantly, but Days 3 and 4 were the closest.

3.2 DATA SEGMENTATION

3.2.1 Segmentation Using Accelerometry

In order to separate data for whole days into the particular events of interest, the periods of event activity were identified in the accelerometer data. The foot-mounted accelerometers provided reliable indicators of activity due to the sustained periodicity and amplitude of the signal during ambulation versus the noise generated during nonevent activity. Selection and event classification were performed using knowledge of the general time and duration of the events. Identical sequence of events on each day was beneficial to the segmentation process.

3.2.2 Foot-March Segmentation

The 5 km foot march was segmented using the acceleration magnitude signal from the feet. This was performed primarily because the march was a relatively long period (approximately 45 minutes) of increased activity and was easily detectable as the longest stretch of high-magnitude acceleration. First, the acceleration signal was low-pass filtered and a threshold was applied to include segments only from the march. The longest active segments were selected from the filtered signal and were processed to remove any gaps in detection so that the entire portion was within a sensible range of duration for the march.

3.2.3 mACRT Segmentation

mACRT events took place in both the morning and afternoon for each day of testing. The timing of these was measured using timing gates at each obstacle to obtain elapsed times. Study coordinators also used wristwatches to record absolute start and stop times down to the second for each instance of the mACRT. Due to personnel constraints on event coverage, some of the wristwatch times were not recorded, particularly for the afternoon event. Since the mACRT event was much shorter in duration (approximately 4–8 minutes) than the march, a modified approach was taken to isolate the activity using either the absolute watch times or manual segmentation. This segmentation was used as a guide for more precise segmentation guided by the accelerometer and the gate timing.

Segmentation from Absolute Times

Segmentations of the mACRT were further refined from the watch-derived start and stop times by adjusting both times by a temporal bias calculated for each day between the watch times and the accelerometer data. Assuming that the watch times captured an accurate elapsed time, the absolute times were adjusted to maximize acceleration magnitude in the corresponding time window. To eliminate the bias from the watch, the median calculated bias for each day was removed from each recorded start and stop time. Alignment with gate elapsed times was accomplished by segmenting the endpoint of the first two mACRT activities (200 m sprint and low hurdle) within the mACRT using log-likelihood estimates from both feet and the harness accelerometers. This endpoint was then used as an anchor for the gate elapsed times, since the actual start time was a more uncertain measurement due to the running head start before activating the gate.

Manual Segmentation

For instances when absolute recorded times were not available, a coarse segmentation of the mACRT was provided manually. Using a MATLAB GUI, the right-foot accelerometer data was displayed for the entire day’s activity, and boundaries were placed on approximate start and stop times of the mACRT. This was a relatively simple task to perform manually, since the patterns associated with a mACRT in the accelerometry are very apparent. Additionally, the mACRT happened either before or after the 5 km march, which was a rather obvious landmark to locate within the signal and thus made finding the relative location of the morning or afternoon mACRT easier. Manual start and stop times were treated as surrogate watch times in the subsequent segmentation processing.

Time Synchronization

Time synchronization proved to be a common issue when fusing time series data from multiple sensors. Each device had its own internal clock, and despite vigilant daily synchronization with computers and network time, there were discrepancies between absolute times across devices. Between the wristwatches and the ActiGraph accelerometers, there appeared to be a simple bias of ~18 seconds. The SEMs exhibited a temporal drift, which resulted in a linear difference between their data and the

ActiGraphs. This was corrected by comparing foot-march segmentation results between the SEM and ActiGraph accelerometers, and scaling the samples from the SEM linearly (with an offset) to match the segmentation of the ActiGraph activity. Since the SEM accelerometer acquired data at 256 Hz, it was also downsampled to match the sampling rate of the ActiGraph (100 Hz).

3.2.4 Segmentation Limitations

The accelerometers gave reliable data for activity and ambulation detection, but segmentation accuracy was limited because many of the Soldiers started ambulation before crossing the starting cone. In other words, the time at which the Soldiers started walking did not correspond to the start of the event. Thus, the estimation of overall start times for events was made more difficult. Additionally, segmentation of events within the mACRT proved difficult to achieve automatically due to their relative similarity. All events involved some amount of forward locomotion, and the differences between events were generally below a detectable threshold.

3.3 ANALYSIS METHODS

3.3.1 Accounting for Soldier Effort

In analyzing the impact that different levels of protective gear have on Soldier performance, it is critical to account for the fitness and effort exerted by the Soldier(s). For instance, it is possible that a highly motivated Solider in a heavy configuration could have faster times than an unmotivated Soldier in a light configuration. Omitting effort as a parameter in the analysis between Soldiers as well as for an individual Soldier could potentially bias the results and lead to conclusions that the heavy configuration enhances performance or that the light configuration hinders performance.

Motivation and effort levels are likely to vary for a single Soldier over the course of the test for a number of reasons. If a Soldier wearing Configuration C develops blisters during the 5 km march on the first day, this Soldier may not push as hard on the second day when wearing Configuration A. A decrease in times in this case could be a result of the blisters and not the uniform configuration, which would again lead to skewed conclusions. Normalizing the data based on effort exerted in each configuration helps to eliminate external influences on the Soldiers and their performance.

3.3.2 Estimating Heart Rate

A subject’s effort on a particular task is quantified based on the average heart rate (HR) during the task. A valid HR estimate requires the accurate detection of heartbeat intervals. These are called RR intervals because they are derived from the time intervals between the successive R components of the QRST complex, which occurs in ECG waveforms. The initial RR detections are provided by the internal SEM software. Due to movement artifacts and/or a loose-fitting Hidalgo harness, many of these RR estimates can be incorrect. When the ECG waveform is noisy, most of the SEM RR detections tend to have small values (<0.25 seconds) that are physiologically implausible.

An outlier removal algorithm was used to remove invalid RR values. First, RR values are only allowed if, within a local window of the nearest 100 RR values, more than 70% of them have plausible values (>0.25 seconds). Next, additional RR values are removed using an iterative filtering and outlier removal process. Because RR intervals are approximately log-normally distributed, this algorithm uses log(RR) values. In each iteration, the local mean of the remaining log(RR) values is estimated using a second-order Kalman filter operating in the forward (filtering) and backward (smoothing) directions. The log(RR) values that have absolute deviations from the local mean log(RR) estimate greater than 0.25, 0.2, and 0.15, respectively, on iterations 1, 2, and 3, are removed. The removal of outliers on each iteration affects the Kalman filter mean estimates on the next iteration, causing the estimates to gracefully settle into values that are best supported by the data. Heart-rate estimates are derived from the RR values for disjoint 5-second frames, provided there are at least two valid RR values in that frame: HR = mean(60/RR). Figure 18 shows an example of HR values from a single subject during a 5 km march based on the original Hidalgo RR values (red) and the filtered values after outlier removal (blue).

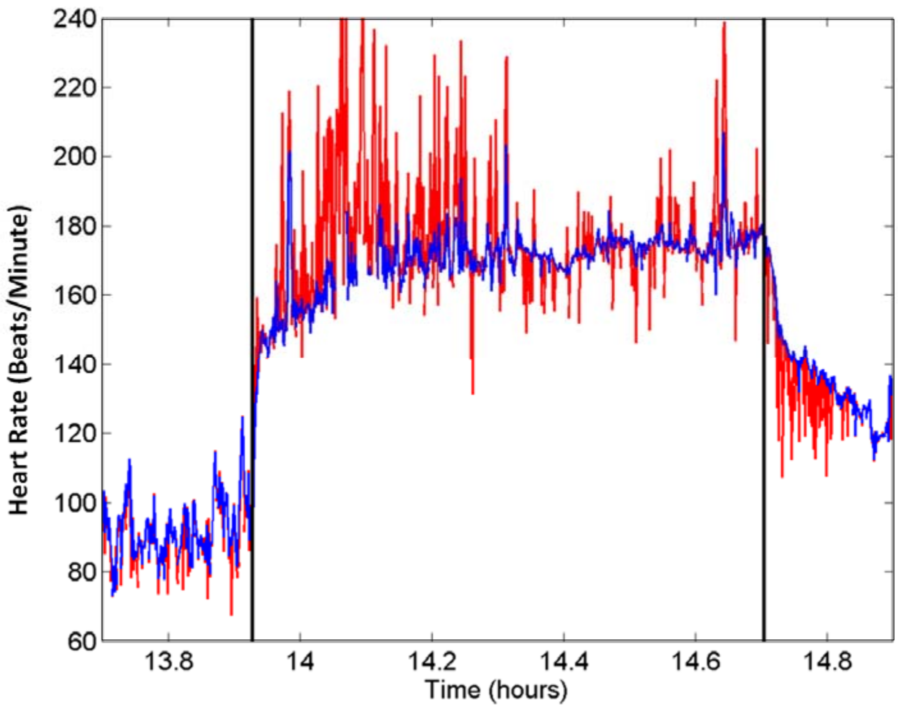


Figure 18. Soldier example heart-rate values. Original Hidalgo HR estimates (red) and MIT LL filtered HR estimates (blue). 5 km march is delineated by bold vertical lines.

3.3.3 Heart Rate–Derived Effort

Valid and reliable HR estimates allow the amount of *effort* given by a subject on an event, such as the 5 km march or the mACRT, to be estimated. This is done using the average HR during the event, provided that at least 25% of the 5-second frames during the event contain valid HR estimates. Effort estimates greatly facilitate the ability to estimate the effect of equipment configurations because direct comparisons between different event trials can be made, even if the subject contributed widely different effort levels during those trials.

We define effort, denoted by x, as the fraction of available HR capacity being used (for a particular configuration),

$$x = \frac{\overline{HR} - HR_{\min}}{HR_{\max} - HR_{\min}},$$

where the individualized HR min and max values are estimated from the entire SPBE data collection for each subject. This happens to be the same equation as for heart-rate reserve, which is recommended by the American College of Sports Medicine for assessing exercise intensity. However, we are using this equation for a different purpose and are estimating minimum and maximum individual heart rates from field data, rather than under controlled conditions as is done for heart-rate reserve. The minimum and maximum heart rates were derived for each individual by manually reviewing the heart-rate data and finding candidate blocks of min and max heart rates. The candidates were pruned to single min and max heart rates by removing outliers in the blocks and finding heart rates that were consistent for approximately 60 seconds.

An example of data generated by using the effort equation is shown in Figure 19. The scatter plot depicts speed (km per hour) as a function of effort among 30 subjects using Configurations B (blue) and C (red) on the 5 km march. No trend between the two configurations is apparent, as illustrated by the linear regression fits (blue and red dotted lines).

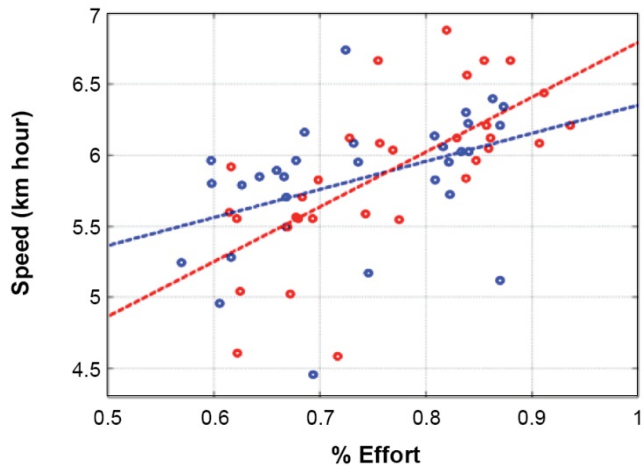


Figure 19. Speed in 5 km march as a function of effort for Configurations B (blue) and C (red). Linear fits shown with dotted lines.

3.3.4 Effort Normalization

Individual within-subject normalization was performed by dividing each subject's effort levels on Configuration B and C by the average effort on both configurations,

$$\tilde{x}^B = \frac{x^B}{.5\left(x^B + x^C\right)}, \qquad \tilde{x}^C = \frac{x^C}{.5\left(x^B + x^C\right)},$$

and also divide each subject's speed by their average speed with both configurations,

$$\tilde{y}^B = \frac{y^B}{.5\left(y^B + y^C\right)}, \qquad \tilde{y}^C = \frac{y^C}{.5\left(y^B + y^C\right)}\ .$$

Using these normalized measures, a trend between the two configurations becomes readily apparent, as shown in Figure 20. It is apparent, given the same effort level, that Configuration C causes a reduction in speed. This configuration effect is quantified using a constrained linear regression fit technique. The maximum likelihood regression fit is made to the normalized values, for a given configuration pair, based on the assumption of a multiplicative effect on speed of one configuration relative to another, given the same effort level. In other words, we find the maximum likelihood percent increase or decrease of one configuration relative to another, using a linear regression fit. The regression fit to the two configurations is shown with dotted red and blue lines. This regression fit corresponds to a fixed decrement of 1.7% in speed given Configuration C compared to Configuration B at the same effort level.

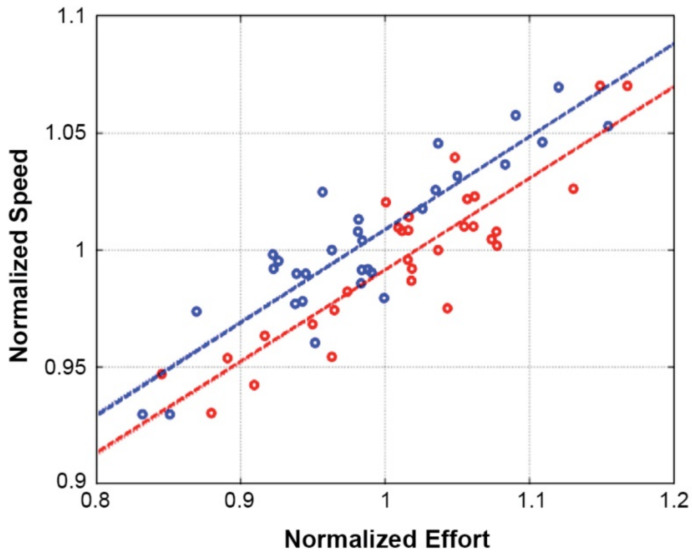


Figure 20. Normalized speed in 5 km march as a function of normalized effort for Configurations B (blue) and C (red). Constrained linear fits shown with dotted lines.

3.3.5 Statistical Tests

The *p*-value, used in this report, is the probability that the null hypothesis is true. In other words, it is the probability that a performance difference related to equipment configurations could be due to chance. To compute p-values, a nonparametric test (i.e., no Gaussian assumption) called the Mann-Whitney U test (also known as the Wilcoxon rank sum test) was used. For the results shown in Figure 20, this test yields $p = 0.0001$, therefore allowing for confident rejection of the null hypothesis. Note that, without access to physiological recordings and only using the normalized speed values, statistically significant differences between the two equipment configurations are not found ($p = 0.69$). Therefore, the use of physiological recordings, such as heart rate, allows for more precise estimates of the relative effects of the equipment configurations to be made and for finding subtle effects that are not obtainable from timing data alone of the 5 km march.

In addition to *p*-values, 95% confidence intervals (based on a Gaussian distribution assumption) were also computed.

3.3.6 Accelerometers and the Effect of Gravity

The earth's gravity vector (direction gravity points) always points to the center of the earth and has a magnitude of 1G (acceleration due to gravity). Left sitting by itself on a surface that is not accelerating, an accelerometer used during this test would measure this gravity vector projected along its three possible dimensions. If one were to pick up the accelerometer and rotate it around, one would see the gravitational vector "moving" with respect to the accelerometer's three dimensions. This effect is illustrated in Figure 21. The 1G is being projected or divided into the three-dimension constituent parts.

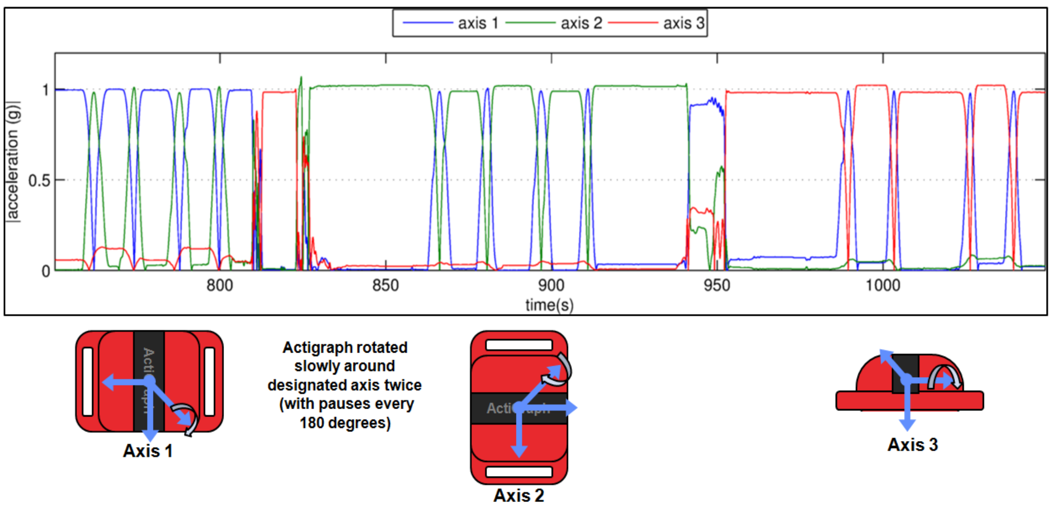


Figure 21. Example of moving an accelerometer around various axes to demonstrate the effect of gravity on accelerometer measurements.

If the accelerations were solely due to gravity, one would know which direction the gravity vector is pointing while the sensor is being moved. But since moving the sensor also induces accelerations, and these accelerations can add and subtract from the gravity vector, one does not know for sure which direction the gravity vector is pointing. As a result, it is very difficult to isolate and remove this "gravity noise" from the accelerometer measurements.

Figure 22 shows the average harness acceleration magnitude during the high crawl averaged across Soldiers for each equipment configuration, with the 95% confidence interval. The average harness acceleration for each equipment configuration is similar, and close to the acceleration due to gravity

(1G), presumably meaning that the Soldiers' torsos are moving at a relatively slow and steady pace (a small fraction of the acceleration due to gravity) during the high crawl. Generally, the magnitude of accelerations is slightly less than 1G, indicating that the average net vertical acceleration due to locomotion is in the direction opposite of gravity.

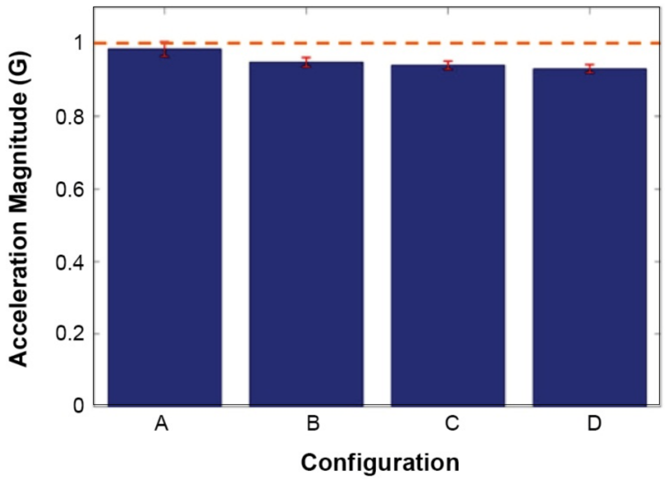


Figure 22. Average harness acceleration magnitude during high crawl (morning mACRT), averaged across all Soldiers for each equipment configuration.

Table 1 shows the p-values for the distributions of harness acceleration magnitude for Soldiers in each uniform configuration for the high crawl. The data in the table indicate that, for the high crawl, Configuration A is separable from Configurations C and D on the basis of average harness acceleration magnitude ($p < 0.01$). However, the other configurations are less separable from one another.

TABLE 1

High Crawl Average Harness Acceleration Magnitude (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	0.0111	A: 30, B: 29
A–C	7.42E-04	A: 30, C: 32
A–D	5.76E-05	A: 30, D: 28
B–C	0.406	B: 29, C: 32
B–D	0.0469	B: 29, D: 28
C–D	0.227	C: 32, D: 28

Figure 23 shows the average harness acceleration magnitude during the balance beam, averaged across Soldiers for each equipment configuration. As with the high crawl, the acceleration is dominated by gravity. This is the case because the acceleration due to gravity is 1G, and as can be seen in the plot, each of the bars are quite close to 1G.

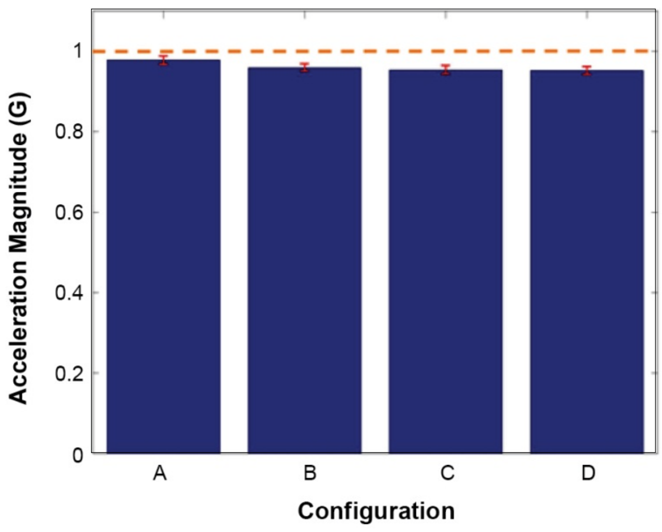


Figure 23. Average harness acceleration magnitude during balance beam (morning and afternoon mACRT), averaged across all Soldiers for each equipment configuration.

Table 2 shows the p-values for the distributions of harness acceleration magnitude for Soldiers in each uniform configuration for the balance beam. The data in the table indicate that, for the high crawl, Configuration A is separable from Configurations C and D on the basis of average harness acceleration magnitude ($p < 0.01$). However, the other configurations are less separable from one another.

TABLE 2

Balance Beam Average Harness Acceleration Magnitude (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	0.0192	A: 24, B: 27
A–C	0.00340	A: 24, C: 27
A–D	0.00140	A: 24, D: 24
B–C	0.4547	B: 27, C: 27
B–D	0.2240	B: 27, D: 24
C–D	0.7932	C: 27, D: 24

The equipment configurations for most of the events are not easily separable on the basis of mean acceleration, mainly due to the confounding effect of gravity on the magnitudes. The mean contains major gravity components that overwhelm the smaller event-driven accelerations. These complications are mitigated later in the following results section by examining the variability of the accelerometer data, in the form of the standard deviation of the magnitude over the course of events. The standard deviation also contains gravity components, but to a lesser degree, allowing a more pronounced event signal.

4. RESULTS SUMMARY

4.1 OVERVIEW

For all of the events listed in Sections 4.2 through 4.4, the differences between equipment configurations were statistically significant ($p < 0.01$) when percent effort was taken into account.

4.2 5 km MARCH

We determined changes in Soldier speed for the different configurations on the 5 km march, by normalizing for Soldier effort using the max-likelihood regression fitting method described in Section 3.3.4. For instance, Figure 24 (left) shows the results of the analysis for multiple equipment pairings: A–B, B–C, and C–D. As can be seen in the graph, going from Configuration A to B, to D, the Soldiers experience reduced speed for a given amount of effort.

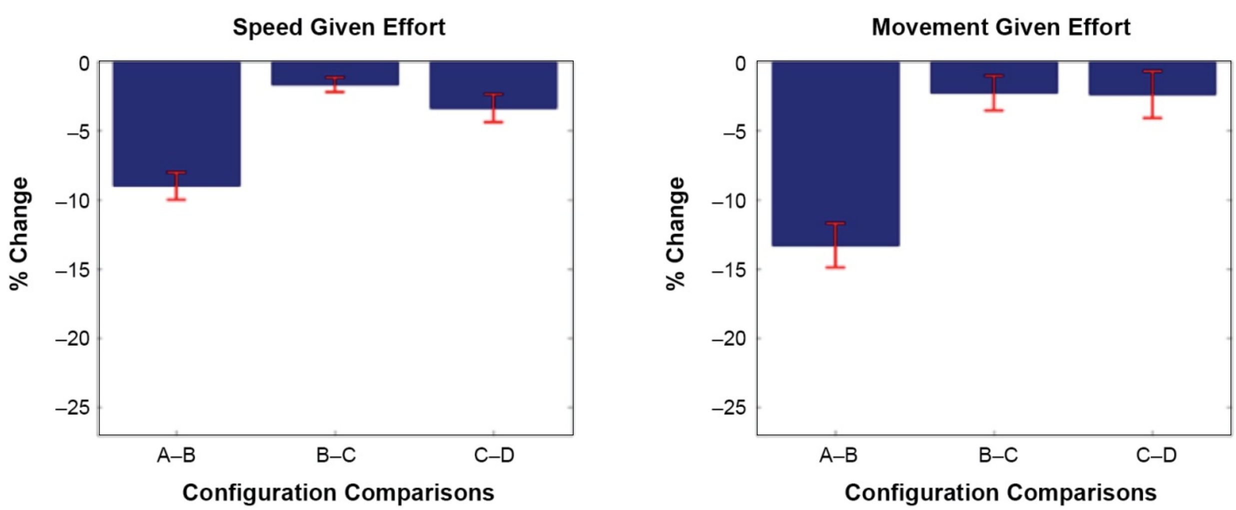


Figure 24. Configuration comparisons for 5 km march. Left: Change in speed (based on timings) given same effort for different configuration pairs. Right: Change in movement (based on accelerometry) given same effort for different configuration pairs.

Figure 24 (right) shows similar effects, which are instead obtained by comparing subject *effort* with *movement (*rather than with *speed*). Movement is defined as the average amplitude standard deviation within 10-second frames (with 5-second overlap between successive frames),

$$m_i = std_i\left(\|\vec{a}\|\right)$$

where ***a*** is the three-axis acceleration vector (recorded at 25.6 Hz) from an accelerometer in the Hidalgo harness worn on the left side of a subject's torso.

4.3 MORNING mACRT

Figure 25 shows the reduction in speed and movement not only occurred during the march, but also during the morning mACRT, which included two sprints and the obstacle course.

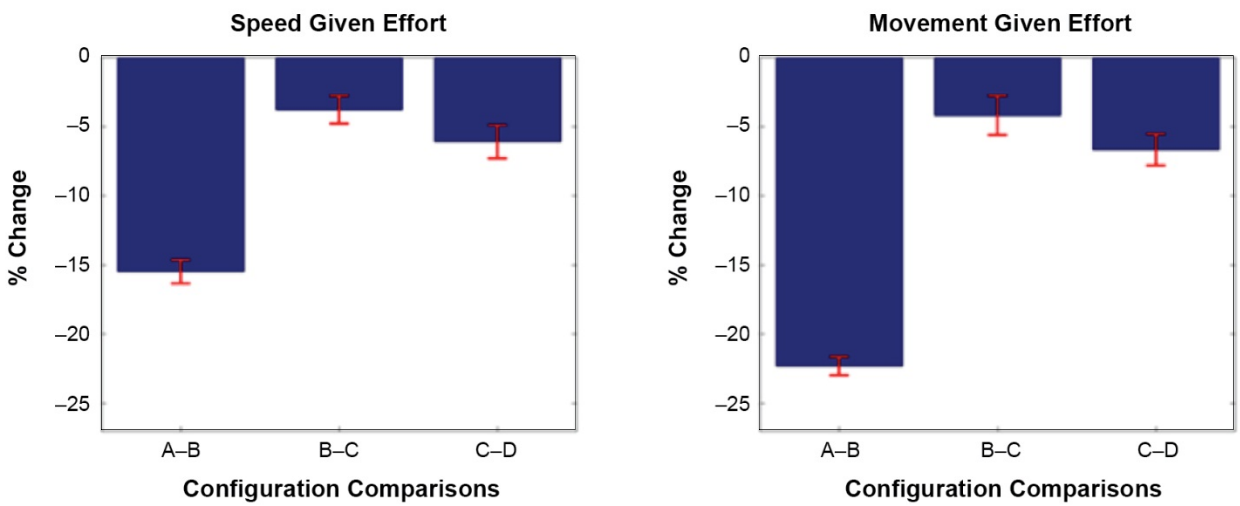


Figure 25. Configuration comparisons for morning mACRT. Left: Change in speed (based on timings) given same effort for different configuration pairs. Right: Change in movement (based on accelerometry) given same effort for different configuration pairs.

4.4 MORNING mACRT OBSTACLES AND SPRINTS

Figure 26 shows results broken out for the mACRT course alone, and Figure 27 shows the results for the two sprints alone. The sprint times were derived by summing the two sprint times together, forming a single cumulative time. As suggested by the confidence intervals, all the configuration differences shown in Figure 24 through Figure 27 are statistically significant ($p < 0.01$).

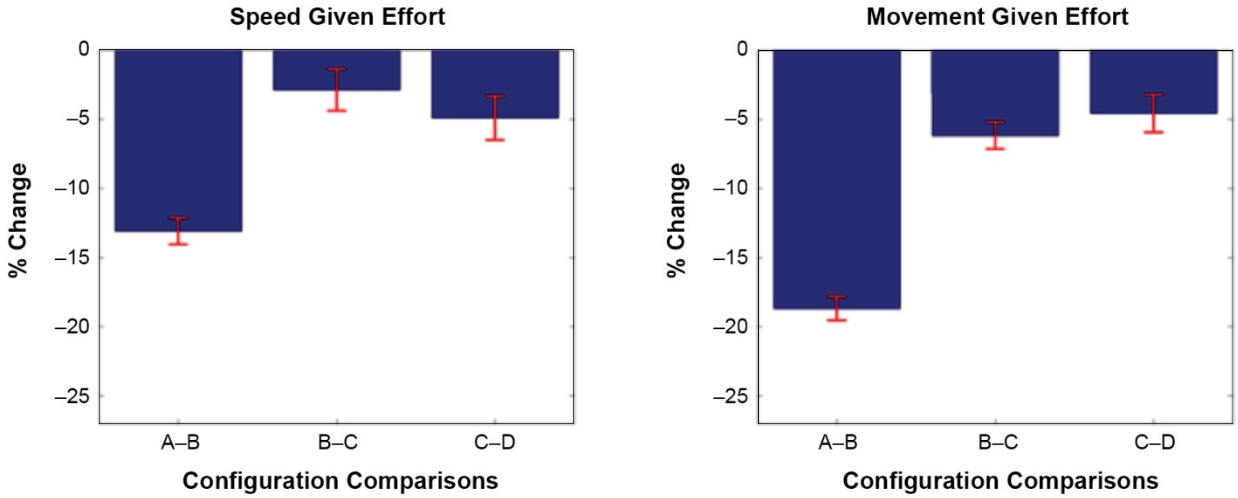


Figure 26. Configuration comparisons for obstacle course in morning mACRT. Left: Change in speed (based on timings) given same effort for different configuration pairs. Right: Change in movement (based on accelerometry) given same effort for different configuration pairs.

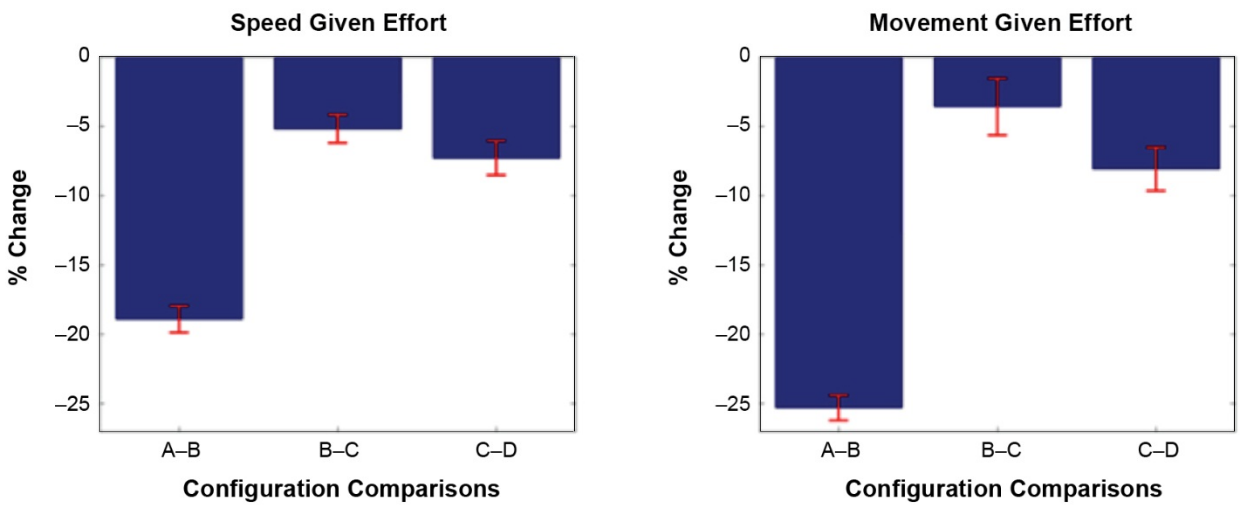


Figure 27. Configuration comparisons for combined sprints in morning mACRT. Left: Change in speed (based on timings) given same effort for different configuration pairs. Right: Change in movement (based on accelerometry) given same effort for a different configuration.

4.5 MORNING mACRT HIGH CRAWL

The analysis of the mACRT obstacle sub-events focuses on the accelerometer data, particularly that of the SEM accelerometer in the torso harness. The results are not normalized for effort. This is due to the fact that, as compared to other events like the march or the mACRT sprints, the other individual sub-events of the mACRT are short in duration. Therefore, Soldier heart rate, during one of these sub-events, is primarily driven by the details of the previous sub-event, and not necessarily attributable to the Soldier effort during the current sub-event.

Movement in this section is defined slightly differently than movement in previous sections. Here it is defined as the standard deviation of the acceleration magnitude over the *entire duration of these sub-events*, as opposed to the 10-second frames of the longer events like the 5 km march or mACRT sprints. Statistics are gathered for the foot and harness accelerometers based on available timing information for the mACRT sub-events. Due to the initial absence of timing gate information, the times for the high crawl are estimated based on a “prone” body position reading from the SEM. As timing gate information became available, it was discovered that the SEM body position reading was a good indicator for the start and end of the high crawl, and this enabled use of data sets where the timing gate information could not be used.

Table 3 shows the p-values for the distributions of harness movement values for Soldiers in each uniform configuration. These results are for all Soldiers in each configuration and are not individualized for each Soldier. Therefore, they do not account for differences in movement between Soldiers that do not arise from uniform configuration differences, such as one Soldier having more overall movement in all configurations than another.

TABLE 3

High Crawl Harness Movement (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	5.60E-06	A: 30, B: 29
A–C	1.55E-07	A: 30, C: 32
A–D	5.61E-10	A: 30, D: 28
B–C	0.0699	B: 29, C: 32
B–D	3.95E-05	B: 29, D: 28
C–D	0.0401	C: 32, D: 28

The data in Table 3 indicate that, for the high crawl, Configuration A is easily separable from the other configurations on the basis of movement ($p < 0.01$). Configurations B and D are also separable ($p < 0.01$). However, the other configurations are less separable from one another. The movement in Configuration C is similar enough to that in Configuration B and D that $p > 0.01$ for each.

In order to elucidate the effects of the different uniform configurations without the confounding factor of variations of movement patterns between different Soldiers, it is necessary to individualize the results by normalizing them for each Soldier with respect to Configuration A, using it as a baseline.

Figure 28 shows the differences in movement computed over the duration of the high crawl for each Soldier in each equipment configuration. The results are individualized for each Soldier by computing the percent change for Configurations B, C, and D from Configuration A, and then averaged across Soldiers. The 95% confidence intervals are also computed.

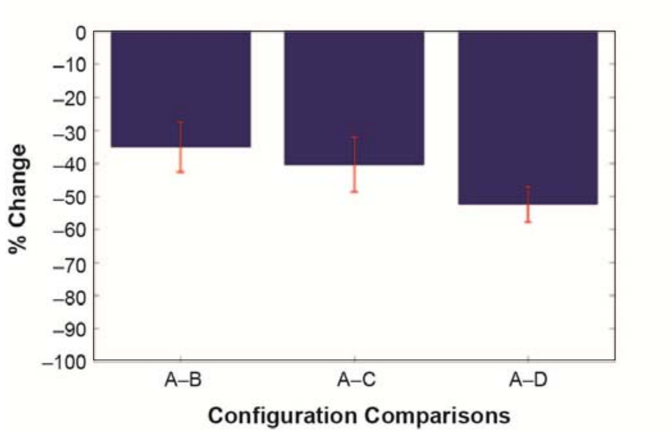


Figure 28. Configuration comparisons for high crawl (morning mACRT). Percent change in the standard deviation of the harness acceleration magnitude during the high crawl (as compared to Configuration A), averaged across Soldiers for each equipment configuration.

Table 4 shows the p-values for determining if the percent change for A–B can be separated from percent change of A–C, e.g. they come from different distributions. These results represent the differences in the individualized movement for every combination of the comparisons shown in Figure 28. This allows differences to be seen between Configurations B, C, and D, when each is individualized with respect to Configuration A, which is used as a baseline. The data in Table 4 shows that the A–B configuration comparison differs from the A–D comparison in a statistically significant way, but the other combinations do not.

TABLE 4

High Crawl Harness Movement Percent Change (with Respect to Configuration A) p-Values (Individualized) and Number of Soldiers for Each Combination of Uniform Configuration Comparisons

Comparisons	p-Value	Number of Soldiers
A–B and A–C	0.0922	A–B: 26, A–C: 30
A–B and A–D	9.65E-04	A–B: 26, A–D: 27
A–C and A–D	0.0118	A–C: 30, A–D: 27

There is less movement in heavier equipment configurations, with the average Soldier having less than half of the movement in Configuration D than in Configuration A. The acceleration due to motion is attenuated or dampened in heavier configurations. This indicates that the heavier configurations, in addition to making the high crawl take more time, are inhibiting the Soldiers' motions during it. Similarly, there is less foot movement in heavier configurations. However, the effect is less pronounced for the feet than the torso. This is possibly due to the fact that the feet have a wider range of motion that is less inhibited by increasingly heavier configurations than that of the torso. The wider range of motion for the feet is demonstrated in Figure 29. It can be seen in the plot that the amplitude of acceleration for the feet is much higher than that of the torso, likely affected by the impact with the ground.

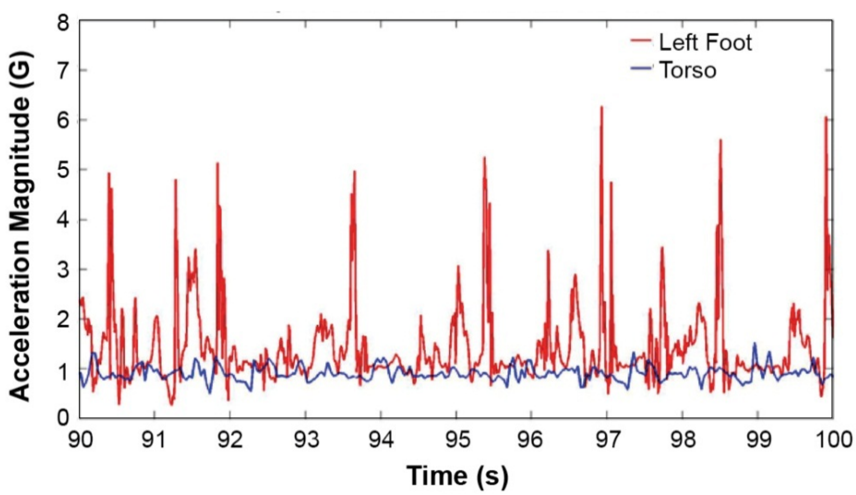


Figure 29. Example acceleration measurements from Hidalgo torso harness-mounted accelerometer (blue) and the left-foot accelerometer (red).

Table 5 shows the p-values for the distributions of left-foot movement values for Soldiers in each uniform configuration. As with the torso data, these results are for all Soldiers in each configuration, and are not individualized for each Soldier.

TABLE 5

High Crawl Left-Foot Movement (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	0.435	A: 30, B: 29
A–C	0.00520	A: 30, C: 32
A–D	0.00300	A: 30, D: 28
B–C	0.0521	B: 29, C: 32
B–D	0.0185	B: 29, D: 28
C–D	0.651	C: 32, D: 28

The data in Table 5 indicate that, for the high crawl, Configuration A is separable from Configurations C and D on the basis of foot movement ($p < 0.01$). However, the other configurations are less separable from one another.

4.6 MORNING mACRT BALANCE BEAM

Table 6 shows the p-values for the distributions of harness movement values for Soldiers in each uniform configuration during the balance beam. As with the high crawl, these results are for all Soldiers in each configuration and are not individualized for each Soldier.

TABLE 6

Balance Beam Harness Movement (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	0.0111	A: 24, B: 27
A–C	2.80E-04	A: 24, C: 27
A–D	0.00100	A: 24, D: 24
B–C	0.0548	B: 27, C: 27
B–D	0.0809	B: 27, D: 24
C–D	0.6989	C: 27, D: 24

Soldier movements in all the configurations during the balance-beam event were similar. For the nonindividualized data, the only configuration pairs that are separable in a statistically significant manner are A–C and A–D ($p < 0.01$).

Figure 30 shows the differences in movement during the balance beam, averaged across Soldiers for each equipment configuration.

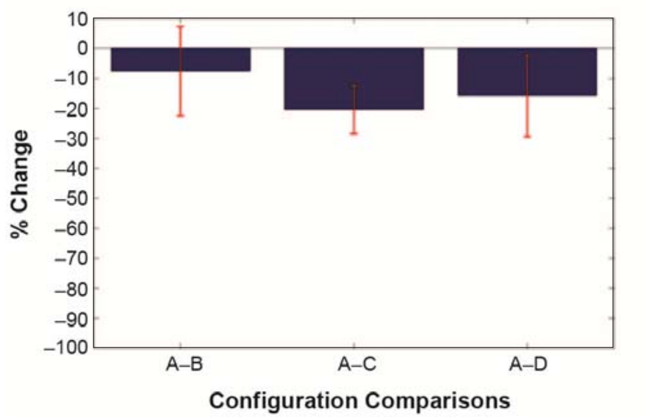


Figure 30. Configuration comparisons for balance beam (morning and afternoon mACRT). Percent change in movement, measured by torso harness accelerometer (as compared to Configuration A), averaged across Soldiers for each configuration.

Table 7 shows the p-values for the percent change results for each uniform configuration with respect to Configuration A. These results represent the differences in the individualized movement for every combination of the comparisons shown in Figure 30. This comparison allows differences to be seen between Configurations B, C, and D, when each is individualized with respect to Configuration A, which is used as a baseline. The data in Table 5 show that there are no statistically significant differences between the individualized configuration comparisons in Figure 30.

TABLE 7

Balance Beam Harness Movement Percent Change (with Respect to Configuration A) p-Values (Individualized) and Number of Soldiers for Each Combination of Uniform Configuration Comparisons

Comparisons	p-Value	Number of Soldiers
A–B and A–C	0.252	A–B: 17, A–C: 21
A–B and A–D	0.402	A–B: 17, A–D: 19
A–C and A–D	0.742	A–C: 21, A–D: 19

Balance-beam data from the accelerometers attached to the boots were also analyzed. Table 8 shows the p-values for the distributions of left-boot movement values for Soldiers in each uniform configuration. As with the high crawl, these results are for all Soldiers in each configuration and are not individualized for each Soldier. The data in Table 8 show that there is no statistically significant difference in boot movement between any pair of uniform configurations.

TABLE 8

Balance Beam Left-Boot Movement (Nonindividualized) p-Values and Number of Soldiers for Every Possible Uniform Configuration Comparison

Comparison	p-Value	Number of Soldiers
A–B	0.993	A: 24, B: 27
A–C	0.591	A: 24, C: 27
A–D	0.658	A: 24, D: 24
B–C	0.556	B: 27, C: 27
B–D	0.540	B: 27, D: 24
C–D	0.902	C: 27, D: 24

Due to issues with gate-timing information, a reliable segmentation of the balance-beam accelerometer data is difficult to obtain at this time. Consequently, less data can currently be used for the balance beam analysis than was used for the high crawl. This is evidenced by the larger confidence intervals in Figure 30. The equipment configurations are not as separable for the balance beam as they are for the high crawl data. It is possible that the added inertia of heavier configurations slightly dampens the Soldiers' accelerations to some extent, but it is not yet clear that this is the case. The Soldiers' movements are comparable across configurations, and it is possible that the configurations have little effect on the accelerations of the Soldiers in their necessarily subdued motion as they attempt to maintain balance.

4.7 BODY TEMPERATURE

4.7.1 Skin Temperature

As seen in Figure 31 by applying a time-scaling or warping to each participant's doad march skin temperature data, such that the start and ends of the march match, a very repeatable pattern emerges. Early in the march, there is a drop in temperature, hypothesized to be caused by increased sweat-attributed evaporative cooling and/or exposure to a low outside ambient temperature. It may be due to the fact that many test participants had been inside heated buildings or vehicles until shortly before beginning the road march.

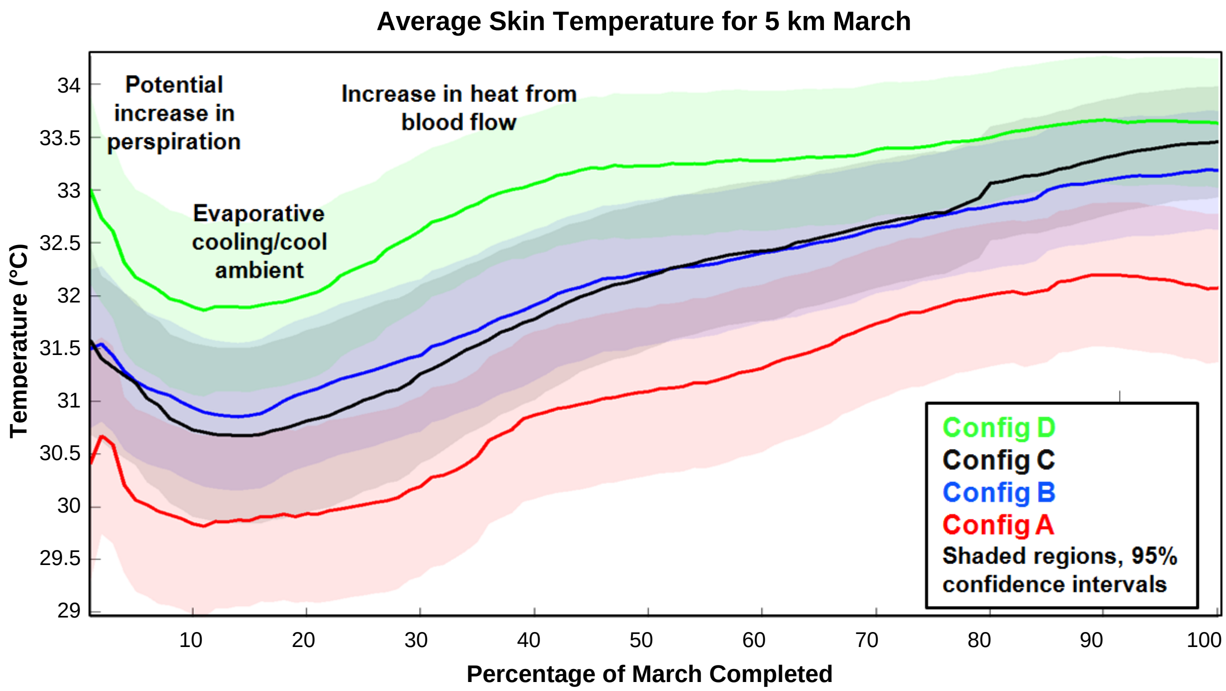


Figure 31. Soldier average skin temperature for 5 km march averaged by scaling march duration to match start and end times. The plot used the following number of Soldiers for computation: A: 31, B: 33, C: 32, and D: 29.

The slow and steady increase, following the sharp decrease, is hypothesized to be due to the body's shunting of generated metabolic heat to the outer layer of the skin via increased vasodilation and blood flow. Configuration D, C, and B maintained approximately 5%, 4%, and 2% respectively higher average skin temperatures than Configuration A. This is likely due to sustained encapsulation of the body armor.

4.7.2 Core Temperature

The core temperatures of Figure 32 show that core temperatures are reasonably close for different configurations at the same percentage of the total march time. Surprisingly, test participants wearing Configuration D had the lowest core temperatures (on average) in the middle portions of the march. However, as Figure 33 shows, test participants wearing Configuration D also had the lowest marching speeds. This raises a key question: Is it possible to estimate the relative effect of two configurations on a test participant's core temperature at the same marching speed?

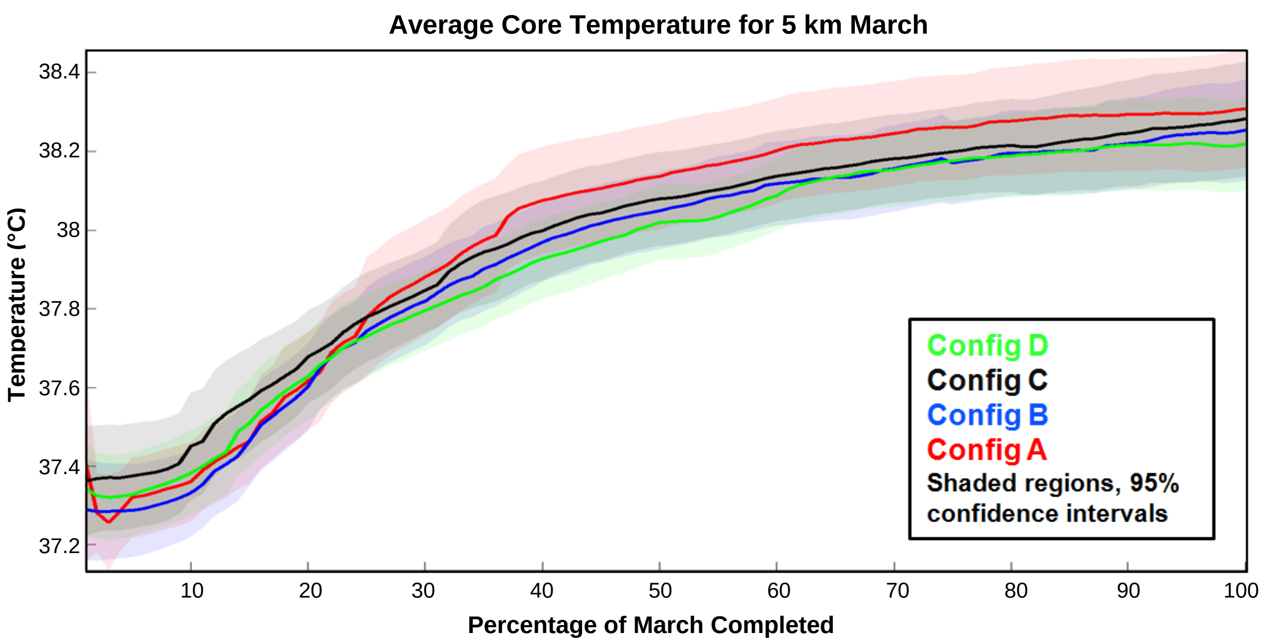


Figure 32. Soldier average core temperature for 5 km march averaged by scaling march duration to match start and end times. The plot used the following number of Soldiers for computation: A: 25, B: 27, C: 26, and D: 27.

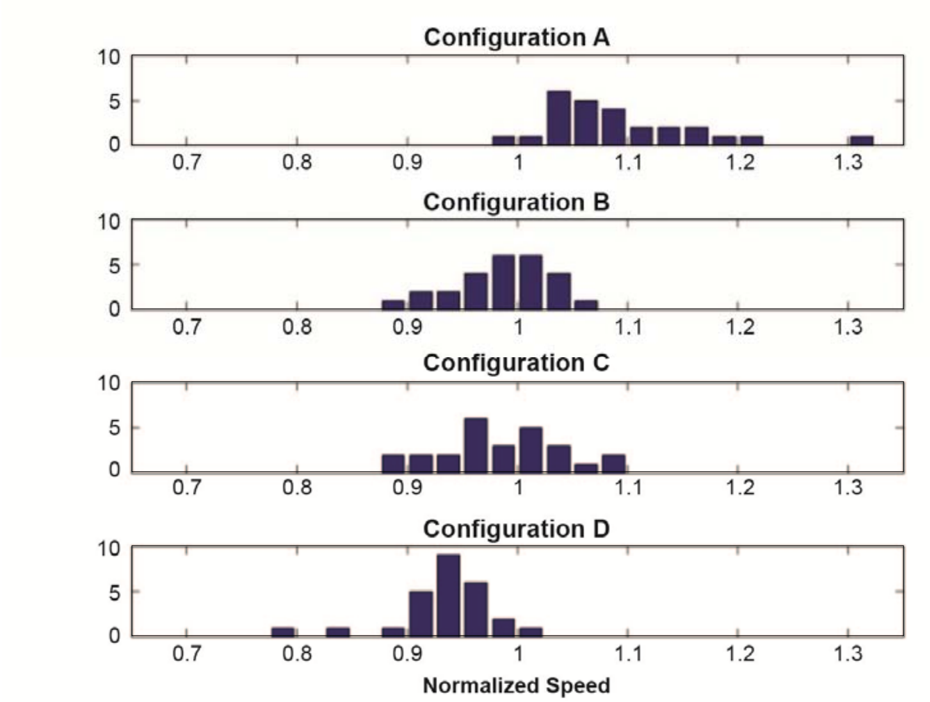


Figure 33. Histogram of test participant normalized speed for the 5 km march (n = 26). The speed of each test participant for each configuration was divided by the average speed of the test participant across all configurations, e.g., a normalized speed of one corresponds to moving at the average speed (across all configurations).

To answer this question, the same type of analysis that was used earlier was applied to estimate the relative effect of two configurations on a test participant's speed, given the same effort. This time, instead of estimating the change in *speed given effort*, the change *in core temperature given speed* was estimated. Note that, similar to what was done to convert *heart rate* into *effort*, the marginal core temperatures, denoted by z, above a core temperature baseline of 36.5°C were used,

$$z = T_{core} - 36.5.$$

Temperature remapping to a minimum baseline temperature of 36.5°C has been used previously by [7]. After subtracting the minimum temperature, within-subject normalization was applied for the two configurations that were being compared (in this example C and D),

$$\tilde{z}^C = \frac{z^C}{.5\left(z^C + z^D\right)}.$$

Figure 34 shows a scatter plot of the normalized speed and normalized core temperature values for two configurations: C (blue) and D (red). The blue and red dashed lines show the maximum likelihood regression fit for the increase in core temperature for Configuration D relative to Configuration C at the same speed. According to this estimate, Configuration D produces about a 15% increase in core temperature relative to Configuration C at the same marching speed. Figure 35 shows this result alongside the result comparing Configurations B and D using the same method, which produces about a 28% increase in core temperature at the same marching speed. So, by performing a fit to the data, one can estimate the increase in core temperature as if the Soldier were moving faster in Configuration D. Therefore, we can see that the core temperature of test participants in Configuration D is predicted to be higher than in Configurations B and C, for a given speed.

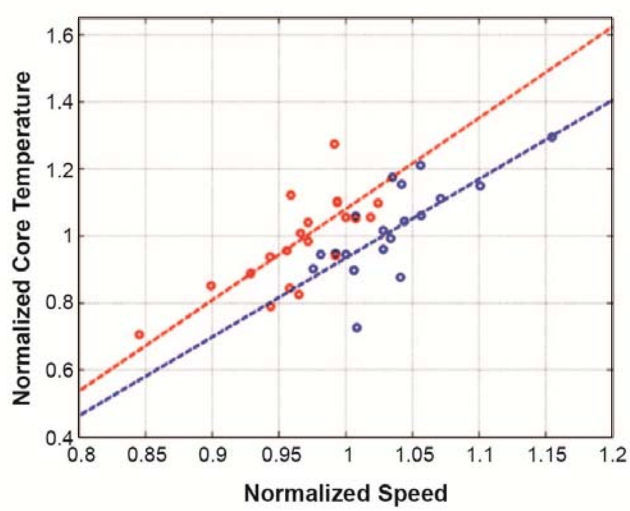


Figure 34. Normalized Soldier speed versus normalized core temperature. Configuration C (20 Soldiers) is shown in blue and Configuration D (20 Soldiers) is shown in red.

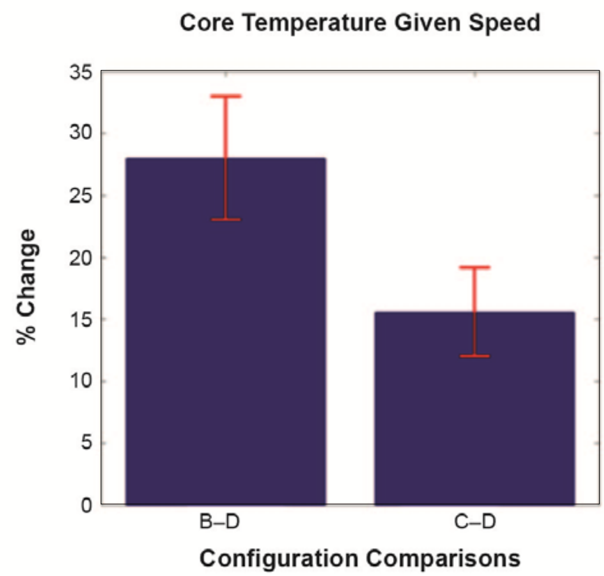


Figure 35. Change in core temperature on the 5 km march given the same average speed, resulting from a shift to more encapsulating configurations (from B to D and from C to D). The plot used the following number of Soldiers for computation: B–D: 20 Soldiers, and C–D: 20 Soldiers. Error bars represent 95% confidence intervals.

Comparisons between Configurations A and D were not made due to the fact that the configurations did not have significant overlapping normalized velocities. This can be seen in the plot of Figure 36. Without overlapping velocities, the obtained fits may not be valid in regions where there is a high degree of extrapolation.

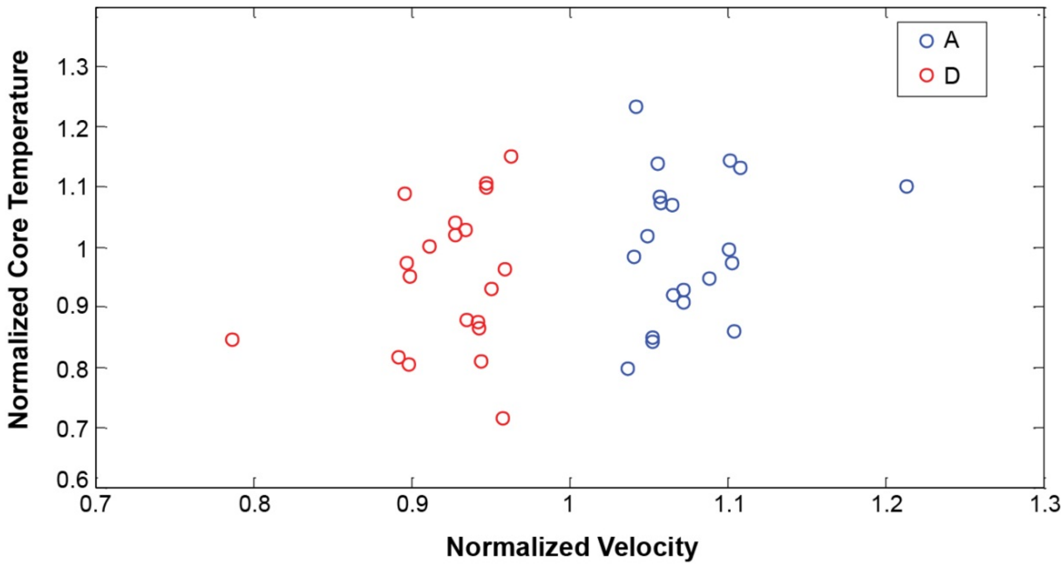


Figure 36. Normalized Soldier speed versus normalized core temperature. Configuration A is shown in blue and Configuration D is shown in red.

4.8 FOOT CONTACT TIME

An exploratory analysis was performed on accelerometer data to extract features related to ambulatory stride period and foot-to-ground contact time during events like the road march and the starting and finishing 200-meter runs (S200, F200) within the mACRT. This type of analysis may provide useful insight into gait asymmetry. Several analytical approaches have been developed. One involves the autocorrelation of accelerometer magnitude signals to extract features related to the periodicity of the stride. Another is a first principles physics-based approach that seeks to directly measure foot-to-ground contact time using the accelerometers attached to the boot laces. This method identifies foot-to-ground contact as the period of time in which the acceleration is stable at the level of gravity with low jerk (the derivative of acceleration), indicating that the foot is firmly planted on the ground. The method appeared to work well for the road march, but not as well for noisy periods of locomotion, like the 200-meter runs within the mACRT. Further development is needed.

5. CONCLUSIONS

The two overarching objectives in the work performed by MIT LL for the SPBE were met. The first objective, gathering physiological datasets to use for equipment evaluations, was successfully achieved by outfitting the Soldiers with physiological sensors that recorded data throughout the SPBE study. These data had relatively few artifacts throughout the collections and, due to the high Soldier participation rate, resulted in a diverse range of Soldier size, weight, fitness, and effort data.

The second objective, determining the best physiological parameters to use as standard metrics for equipment evaluations, was met by defining preliminary sets of metrics and then assessing the differences between the uniform configurations based on these metrics. Of special interest to the metric evaluations was the amount of effort each Soldier contributed to an event. It was found that Soldier effort had a profound effect on most of the metrics evaluated, and if normalization for effort were omitted, the effects of effort alone could bias results.

Test participants' level of effort during the test was estimated using average heart rate compared to Soldier minimum and maximum average heart rate during the test. In this way, heart rate helped to separate level of effort effects from effects due to the particular uniform configuration. There were progressive decreases in speed and movement, as measured by body-worn accelerometers, as the weight of the configurations increased. Slower completion times and lower heart rates indicated test participants were not working as hard with Configurations B–D compared with Configuration A. This is perhaps because of additional weight and discomfort with the more protective configurations. Statistically significant ($p < 0.01$) differences in speed and movement were found between all configurations on the 5 km road march and morning mACRT (total mACRT, sprints alone, and select individual obstacles).

The different equipment configurations appeared to affect the motion of test participants in some individual obstacles of the mACRT. In particular, the heavier equipment configurations had progressively less variability in accelerometer motion measurements during the high crawl, indicating that the configurations inhibited the test participants' motion (i.e., restricting mobility). There was a reduction in acceleration magnitude, in addition to an increase in time required to complete each obstacle event. The reduction in magnitude was not simply due to the subjects' slower motion, as there were instances in which Soldiers moved at the same speed for different configurations, but still showed a reduction in acceleration magnitude.

It is not currently clear that motion inhibition holds for other mACRT obstacles, like the balance beam. Further refinement of obstacle time segmentation is necessary to determine if decreased motion is seen across all events in the mACRT. Preliminary analysis shows that the configuration makes less of a difference for the balance beam than it does for the high crawl. The high crawl requires significant amounts of motion for multiple body parts, whereas some of the other obstacles only necessitate substantial movement of one or two body parts. This is consistent with the idea that configuration differences may be more apparent for specific obstacles. For instance, in heavy shoulder movement events, one would show more shoulder movement variability. Therefore, it would be worthwhile, for future events, to quantitatively determine how sensitive each obstacle may be to different parts of an equipment configuration. By incorporating these targeted events, the evaluation may increase measurement sensitivity to configuration differences. This philosophy could also incorporate the range-of-motion measurements that were performed. For example, if there is a configuration that reduces the ability to elevate or rotate the arms, the Soldier might have to shoot at elevated targets to emphasize this effect.

Configuration effects were readily apparent in the raw test participant skin temperature data. The test participant core temperature data required a more sophisticated analysis method to understand configuration effects due to the body's ability to regulate core temperature. It was found that test participants moved more slowly for heavier configurations, thus impacting the recorded heart rates and computed effort for those configurations. Slowing down during events is one way the Soldiers helped to regulate their core temperature. By normalizing the data for speed, it was shown that if test participants moved at the same speed as they did for the lighter configurations, their core temperatures would be 15% to 28% higher for the heavier configurations.

6. LESSONS LEARNED

Several lessons learned from the SPBE test could improve the separability and reliability of the physiological data if used in future evaluations. Conducting the test in an area in which the climate is controlled or just less cold would help to eliminate confounding factors such as temperature, wind, and precipitation. Executing events in higher temperature and humidity ambient conditions may allow for greater separability in the core temperatures of a given Soldier in all the different uniform configurations. Performing physical fitness baseline tests to determine maximum heart rate and oxygen consumption (VO_2) in a laboratory setting prior to the evaluation would make the effort calculations more accurate and robust, which would facilitate differentiation of the effects of the various configurations. Also, better quantification of exertion levels for each Soldier during each event would allow for analysis of more parameters.

Other observations and lessons learned from the SPBE give insight into protocols and planning that should continue to be implemented in future tests. An essential piece of the study was the inclusion of "preparation" days before formal testing with the test participants. This time allowed both the test administration staff as well as the participating Soldiers an opportunity to gain an understanding of the fit, function, and protocols related to the physiologic equipment. Allowing all people involved in the test time to become comfortable with the equipment facilitated more efficient use of time during the formal test days. The hardware failure rates for the Hidalgo EQ-02 were low, but it was still beneficial to have a 20% backup inventory. The core temperature capsules had a much higher defective/failure rate than expected, and future test supplies should include a 30% over-supply. The core temperature capsules and the SEMs, if on and transmitting, can be detected by handheld devices. Using these devices to verify that the sensors are functioning at the beginning and end of each day allows for a high-level identification of potential faulty equipment.

The evaluation was a unique opportunity to collect physiological data on a large group of Soldiers, not only in several equipment configurations, but over a long period of time and in a fielded environment as well. Collecting multiple physiological parameters for a group of Soldiers on numerous days leads to a high volume of data. In order to analyze and identify the useful information, the data must first be separated, segmented, and synchronized, a time-consuming task. Finding more automated and reliable ways to parse the data for future evaluations will allow for more rapid and informative analysis of the data.

Based on this evaluation, the most useful physiologic parameters for differentiating between uniform configurations are heart rate, body/foot acceleration, and skin temperature. Incorporating additional sensors in future tests, such as a gyrometer to measure absolute orientation of body parts, may lead to discovery of additional useful parameters for evaluation of how equipment affects Soldier performance. The collected physiological data from this test will help the Army perform a benefit analysis of the evaluated uniform configurations. These data may also be used to develop uniform configuration models, train investigators, and develop standard operating procedures and metrics for future evaluation tests.

This page intentionally left blank.

7. BIBLIOGRAPHY

1. Lab., Maneuver Battle. *Soldier Protection Benchmark Evaluation.* Ft. Benning, GA: s.n., 2012.

2. Hidalgo. The EQ02 LifeMonitor. [Online] January 7, 2013. [Cited: January 7, 2013] http://www.equivital.co.uk/products-new/tnr-products-new/sense-and-transmit.

3. Philips. Equivital. [Online] [Cited: January 7, 2013] http://www.healthcare.philips.com/us_en/homehealth/sleep/equivital/.

4. ActiGraph. GT3X+ Monitor. [Online] [Cited: January 7, 2013] http://www.actigraphcorp.com/products/gt3x-monitor/.

5. Garmin. [Online] [Cited: January 7, 2013] https://buy.garmin.com/shop/shop.do?pID=30026&ra=true.

6. Kestrel Meters – Handheld Weather and Wind Meters. [Online] [Cited: January 7, 2013] http://www.kestrelmeters.com/products/kestrel-4400-heat-stress-meter.

7. Mark J. Buller et al. *Estimation of Human Internal Temperature from Wearable Physiological Sensors.* Atlanta: In *Proceedings of the 22nd Conference on Innovative Applications of Artificial Intelligence* (IAAI), 2010.

8. Welk, Gregory J. *Physical Activity Assessments for Health-Related Research.* Champaign, Illinois: Human Kinetics; 1 edition, 2002.

9. Qstarz International Co., Ltd. [Online] [Cited: January 7, 2013] http://www.qstarz.com/Products/GPS%20Products/BT-Q1000XT-F.htm.

This page intentionally left blank.

APPENDIX A
TABLES OF RESULTS

5 km March

	Speed Given Effort		Movement Given Effort	
	% Change	**p-Value**	**% Change**	**p-Value**
A–B	–9.0	0.00	–13.3	0.00
A–C	–10.5	0.00	–15.2	0.00
A–D	–13.9	0.00	–17.7	0.00
B–C	–1.7	0.00	–2.3	0.02
B–D	–5.4	0.00	–5.0	0.00
C–D	–3.4	0.00	–2.4	0.01

AM ACRT (Total)

	Speed Given Effort		Movement Given Effort	
	% Change	**p-Value**	**% Change**	**p-Value**
A–B	–15.5	0.00	–22.3	0.00
A–C	–19.8	0.00	–28.5	0.00
A–D	–26.4	0.00	–30.6	0.00
B–C	–3.8	0.00	–4.2	0.00
B–D	–11.6	0.00	–10.9	0.00
C–D	–6.1	0.00	–6.7	0.00

AM ACRT (Sprints)

	Speed Given Effort		Movement Given Effort	
	% Change	p-Value	% Change	p-Value
A–B	–18.9	0.00	–25.3	0.00
A–C	–22.7	0.00	–31.3	0.00
A–D	–31.7	0.00	–34.5	0.00
B–C	–5.2	0.00	–3.6	0.02
B–D	–14.1	0.00	–10.2	0.00
C–D	–7.3	0.00	–8.1	0.00

AM ACRT (Obstacles)

	Speed Given Effort		Movement Given Effort	
	% Change	p-Value	% Change	p-Value
A–B	–31.1	0.00	–18.7	0.00
A–C	–18.0	0.00	–25.7	0.00
A–D	–22.6	0.00	–28.6	0.00
B–C	–2.9	0.00	–6.2	0.00
B–D	–9.8	0.00	–13.7	0.00
C–D	–4.9	0.00	–4.6	0.00

5 km March

Core Temperature Given Speed

	% Change	p-Value
B–D	28.8	0.00
C–D	15.6	0.01

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: 19 August 2013

2. REPORT TYPE: Technical Report

3. DATES COVERED *(From - To)*:

4. TITLE AND SUBTITLE

Soldier Protection Benchmark Evaluation (SPBE) Physiological Data Collection and Analysis, Fort Greely, Alaska, 17 September–5 October 2012

5a. CONTRACT NUMBER: FA8721-05-C-0002

5b. GRANT NUMBER:

5c. PROGRAM ELEMENT NUMBER:

6. AUTHOR(S)

T.B. Hughes, J.R. Williamson, A.R. Hess, W.T. Young, A. Dumas, K.D. Fischl, and B.A. Telfer

5d. PROJECT NUMBER: 2355

5e. TASK NUMBER: 41

5f. WORK UNIT NUMBER:

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

MIT Lincoln Laboratory
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-1174

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Natick Soldier Research, Development, and Engineering Center
Natick, MA

10. SPONSOR/MONITOR'S ACRONYM(S):

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The United States Army is continually evaluating and improving the form, fit, and function of protective equipment for the individual Soldier. To improve upon the evaluations, the Product Manager Soldier Protective Equipment (PM SPE) asked Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) to incorporate a physiological status monitoring capability into their tests. During the execution of the Army's Soldier Protection Benchmark Evaluation (SPBE), MIT LL outfitted 32 Soldiers with physiological monitors, accelerometers, and GPS's to collect objective physical performance data. The Soldiers performed a predefined set of activities wearing four different protective gear configurations. Data from the sensors supplied by MIT LL were analyzed to assess the effects of the various configurations on the physical performance of the Soldiers. Accounting for percent effort exerted, statistically significant ($p\text{-value} < 0.01$) differences in the speed and movement of Soldiers wearing the different configurations could be seen for a 5-kilometer road march and an obstacle course. During the 5-kilometer march, skin temperature was significantly higher for the more protective and encapsulating configurations. Normalizing for speed on the 5-kilometer march, significantly higher core temperatures were observed for heavier configurations. The use of on-body sensors to collect physiological data during the SPBE provided the Army with an objective dataset to use in their evaluation of the various configurations of protective gear.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Same as report	65	