

PAPER • OPEN ACCESS

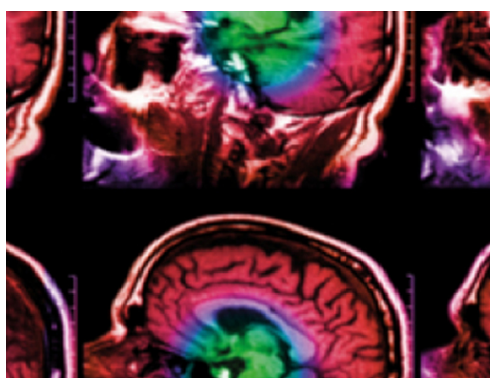
## Validation of ambulatory monitoring devices to measure energy expenditure and heart rate in a military setting

To cite this article: Rahel Gilgen-Ammann *et al* 2021 *Physiol. Meas.* **42** 085008

View the [article online](#) for updates and enhancements.

### You may also like

- [THE STRUCTURE OF THE CIRCUMGALACTIC MEDIUM OF GALAXIES: COOL ACCRETION INFLOW AROUND NGC 1097](#)  
David V. Bowen, Doron Chelouche, Edward B. Jenkins *et al.*
- [THE NEGATIVE EFFECTIVE MAGNETIC PRESSURE IN STRATIFIED FORCED TURBULENCE](#)  
Axel Brandenburg, Koen Kemel, Nathan Kleeorin *et al.*
- [Relationship between anomalous Ettingshausen effect and anomalous Nernst effect in an FePt thin film](#)  
T Seki, R Iguchi, K Takanashi *et al.*



**IPEM | IOP**

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,  
biomedical engineering and related subjects.

Start exploring the collection—download the  
first chapter of every title for free.



## PAPER

## OPEN ACCESS


RECEIVED  
1 March 2021REVISED  
29 July 2021ACCEPTED FOR PUBLICATION  
2 August 2021PUBLISHED  
27 August 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# Validation of ambulatory monitoring devices to measure energy expenditure and heart rate in a military setting

Rahel Gilgen-Ammann<sup>1</sup>, Lilian Roos<sup>1</sup>, Thomas Wyss<sup>1</sup>, Bertil J Veenstra<sup>2</sup>, Simon K Delves<sup>3</sup>, Nadja Beeler<sup>1</sup> , Mark J Buller<sup>4</sup> and Karl E Friedl<sup>4</sup>

<sup>1</sup> Swiss Federal Institute of Sport Magglingen SFISM, Hauptstrasse 247, Magglingen, Switzerland

<sup>2</sup> Institute of Training Medicine & Training Physiology, MOD/TGTF, Herculeslaan 1, Utrecht, The Netherlands

<sup>3</sup> Institute of Naval Medicine, Crescent Rd, Alverstoke, Hampshire, United Kingdom

<sup>4</sup> United States Army Research Institute of Environmental Medicine, 10 General Greene Avenue, Natick, Massachusetts, United States of America

E-mail: [rahel.gilgen@baspo.admin.ch](mailto:rahel.gilgen@baspo.admin.ch)

**Keywords:** physiological monitoring, military personnel, wearable electronic devices, fitness trackers

Supplementary material for this article is available [online](#)

## Abstract

**Objectives.** To investigate the validity of different devices and algorithms used in military organizations worldwide to assess physical activity energy expenditure (PAEE) and heart rate (HR) among soldiers. **Design.** Device validation study. **Methods.** Twenty-three male participants serving their mandatory military service accomplished, firstly, nine different military specific activities indoors, and secondly, a normal military routine outdoors. Participants wore simultaneously an ActiHeart, Everion, MetaMax 3B, Garmin Fenix 3, Hidalgo EQ02, and PADIS 2.0 system. The PAEE and HR data of each system were compared to the criterion measures MetaMax 3B and Hidalgo EQ02, respectively. **Results.** Overall, the recorded systematic errors in PAEE estimation ranged from  $0.1 (\pm 1.8) \text{ kcal} \cdot \text{min}^{-1}$  to  $-1.7 (\pm 1.8) \text{ kcal} \cdot \text{min}^{-1}$  for the systems PADIS 2.0 and Hidalgo EQ02 running the Royal Dutch Army algorithm, respectively, and in the HR assessment ranged from  $-0.1 (\pm 2.1) \text{ b} \cdot \text{min}^{-1}$  to  $0.8 (\pm 3.0) \text{ b} \cdot \text{min}^{-1}$  for the PADIS 2.0 and ActiHeart systems, respectively. The mean absolute percentage error (MAPE) in PAEE estimation ranged from 29.9% to 75.1%, with only the Everion system showing an overall MAPE <30%, but all investigated devices reported overall MAPE <1.4% in the HR assessment. **Conclusions.** The present study demonstrated poor to moderate validity in terms of PAEE estimation, but excellent validity in all investigated devices in terms of HR assessment. Overall, the Everion performed among the best in both parameters and with a device placement on the upper arm, the Everion system is particularly useful during military service, as it does not interfere with other relevant equipment.

## Introduction

Armed forces worldwide are monitoring the demands and activities performed by their personnel in different military occupations (Rosendal *et al* 2003, Pihlainen *et al* 2014, Wyss *et al* 2014, Friedl 2018, Buller *et al* 2021). This allows the commanders to make data based decisions about military performance tasks, missions and injury prevention. Concerning monitoring of physical demands in a military setting, commonly data about distance covered on foot, heart rate (HR), energy expenditure (EE), skin or core temperature are assessed and analyzed (Wyss and Mader 2011, Pihlainen *et al* 2014, Veenstra *et al* 2018, Buller *et al* 2021). Many different commercial or self-developed objective monitoring devices are available. They claim to assess the same parameters, however, these are based for example on HR data, on acceleration data or on a combination of both obtained on the upper arm, on the chest, on the hip using each time different technology (Brage *et al* 2007, Wyss *et al* 2012, Burrell *et al* 2016, Buller *et al* 2021). The major limitation of these devices is the limited knowledge about data validity and reliability for measuring physical demands in a military setting or limited comparability

(Crouter *et al* 2008, Dannecker *et al* 2013, Lee *et al* 2014, Sperlich and Holmberg 2017, Friedl 2018, Bent *et al* 2020, Carrier *et al* 2020). Yet, to adhere to and to benefit from data monitoring, a basic prerequisite is to validate technology against a criterion measure, ideally through a comparison of multiple systems simultaneously, in the environment and during activities of the user group (Sperlich and Holmberg, 2017). Validity is especially important when decisions based on these data may affect soldier health and safety. Knowledge about accuracy of a system can then be used to determine if it is good enough for the intended purpose.

Therefore, the aim of the present study was to simultaneously investigate the validity of different devices and algorithms used in military organizations worldwide to assess EE and HR among soldiers.

## Methods

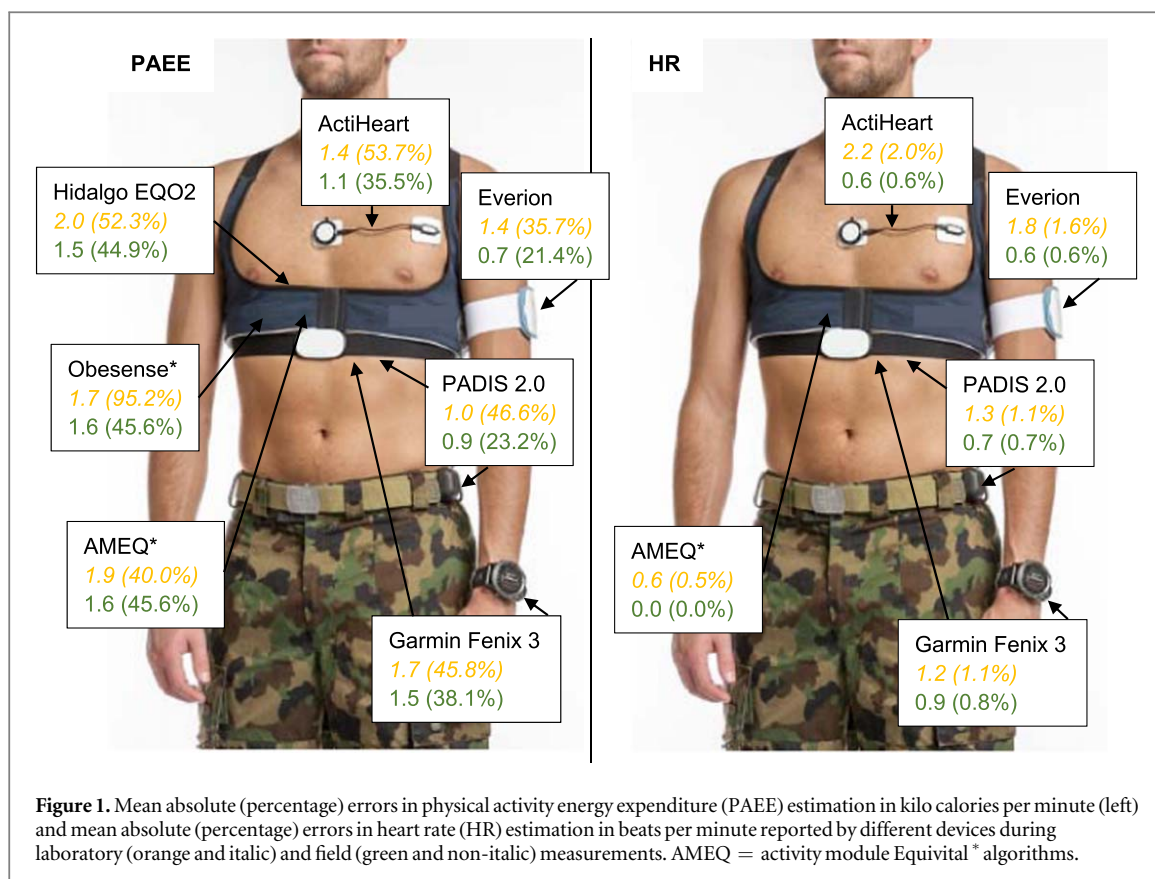
Twenty-three male participants (age:  $20.9 \pm 2.5$  years, weight:  $76.0 \pm 11.6$  kg, height:  $1.8 \pm 0.1$  m, body mass index:  $24.2 \pm 2.5$  kg m<sup>-2</sup>, estimated peak oxygen consumption (Wyss *et al* 2007):  $50.3 \pm 4.3$  ml kg<sup>-1</sup> min<sup>-1</sup>) serving their mandatory military service at a Swiss Army Infantry Training School participated in this study. Participation was voluntary and all participants provided their written informed consent. This study was performed in accordance with the principles of the Declaration of Helsinki and was approved by the local Ethics Committee of Northwest and Central Switzerland (Project-ID: 2016-01842). The present study was conducted within the international collaboration North Atlantic Treaty Organization (NATO) Human Factors & Medicine HFM-260, 'Enhancing Warfighter Effectiveness with Wearable Biosensors and Physiological Models'.

The investigated devices were the ActiHeart (CamNTEch, Cambridge, UK; Firmware 4.0.109), Everion (Biovotion, Zurich, Switzerland; Firmware 02.23.00), Garmin Fenix 3 (Garmin, Olathe, USA; Firmware 8.2), Hidalgo EQ02 (Equivital, Cambridge, UK; Algorithm Settings Version 1.0), PADIS 2.0 (Axiamo, Biel/Bienne, Switzerland; Firmware 30.03.2017, beta-version). Each device was compared to the criterion devices Hidalgo EQ02, for HR data, and MetaMax 3B (Cortex Biophysik GmbH, Leipzig, Germany; Firmware 2.8.6, July 2017), for EE data, respectively. The portable breath-by-breath gas analyzer MetaMax 3B was shown to measure metabolic demands reliably and has been used previously to assess the validity of wearables designed to monitor EE (Vogler *et al* 2010, Duking *et al* 2020). The Equivital LifeMonitor EQ02 is a multiparameter body-worn system including electrocardiography and demonstrated good HR validity (Akintola *et al* 2016). To estimate EE, the Garmin Fenix 3 and the PADIS 2.0 had to be combined with a HR monitor. As the available space around the chest region was a limiting factor for the number of investigated devices, the HR monitor TICKR X (Wahoo Fitness, Atlanta, USA) was chosen because it was simultaneously compatible with both devices (Bluetooth and ANT + sending mode). The sampling rate within all devices was set to 30 s. The ActiHeart was worn on the chest with two self-adhesive electrocardiogram electrodes, the Everion on the left upper arm, the Hidalgo EQ02 holster and the TICKR X strap around the chest and the Garmin Fenix 3 on the left wrist (figure 1). Two PADIS 2.0 sensors were worn, one on the right hip and one on the right side of the backpack. All tested devices as well as the criterion devices were calibrated according to the manufacturer's manual with the original equipment and participant's information was entered into the respective user profiles before each individual measurement. Furthermore, each measurement system was used with body location and wear as intended by the manufacturer.

One week prior to validation measurements participants completed a self-conducted endurance run at the Infantry Training School. After this initial endurance run measurements were taken on two different test days at least 24 h apart within two weeks of each other.

On measurement day one, military activities were obtained indoors in laboratory conditions. Firstly, information about the study was verbally repeated. Secondly, body weight (in only underwear) and body height were measured with a calibrated scale and portable stadiometer, respectively (Model 861 and 213, Seca GmbH, Hamburg, Germany), and upper arm (relaxed midbiceps) and chest (during normal breathing excursions) circumferences were obtained using a non-stretch nylon tape measure. Thirdly, the participants were fitted with all devices and put on their military trouser, t-shirt, and boots. Resting EE and HR were assessed for 15 min after an initial 5 min period with the participants lying awake on a nonconductive bed. This was followed by 9 activities with increasing intensities (table 1). The activity tasks were office work in a sitting position, cleaning weapon and boots, mopping the floor, walking, self-paced marching with loaded backpack ( $20.3 \pm 1.9$  kg) and weapon, lifting and lowering 30 kg loads in 10 s repetitions, lifting and carrying 30 kg loads 20 m with 10 s rest at every turn, shoveling sand for 30 s followed by 8 s rest periods, and running outdoors in running shoes at a steady pace that could be sustained for 60 min (speed  $7.9 \pm 0.9$  km h<sup>-1</sup>). Each of the nine activities lasted 5.5 min with 2 min breaks between activities. During the recovery phases, the upcoming activity was explained.

On measurement day two, field activities were investigated outdoors during a normal military day routine without interference by the study team. The participants were equipped with the same devices as during measurement day one. Activities differed according to the daily military program and could include a 25 km march, shooting, building combat, setting up a check point, material inspection, classroom lecture, or sports.



**Table 1.** Activity protocol of measurement day one during the structured tests of military tasks of increasing intensity (indoors).

Intensity level		
1	Lying awake, no task for 20 min (Resting EE and HR)	
2	Sitting and doing office work	
	Cleaning weapon and boots	
	Mopping the floor	
	Walking	
3	Self-paced marching with loaded backpack (20.3 ± 1.9 kg) and weapon	
	Lifting and lowering 30 kg loads in 10 s repetitions	
	Lifting and carrying 30 kg loads 20 m with 10 s rest at every turn	
	Shoveling sand for 30 s followed by 8 s rest periods	
4	Running at a steady pace that could be sustained for 60 min (speed 7.9 ± 0.9 km h <sup>-1</sup> )	

Note. Each of the activity lasted 5.5 min with 2 min breaks between activities. The task order within intensity level 2 and 3 was randomly assigned.

**Table 2.** Analysis of the overall validity of physical activity energy expenditure (PAEE) and heart rate (HR) measurements by each investigated system.

System	PAEE (kcal.min <sup>-1</sup> )			HR (b.min <sup>-1</sup> )		
	Systematic errors (limits of agreement)	MAE (MAPE)	20% accuracy, <i>n</i> (%)	Systematic errors (limits of agreement)	MAE (MAPE)	5% accuracy, <i>n</i> (%)
ActiHeart	0.8 (2.1)	1.3 (46.3)	17 (45.9)	0.8 (3.0)	1.5 (1.4)	38 (97.4)
AMEQ <sup>a</sup>	−1.7 (1.8)	1.8 (42.3)	3 (8.1)	0.1 (1.2)	0.3 (0.3)	39 (100)
Everion	−0.6 (1.6)	1.1 (29.9)	26 (70.3)	0.6 (2.4)	1.3 (1.1)	38 (97.4)
Garmin Fenix 3	−1.3 (2.2)	1.6 (42.8)	10 (27.8)	−0.2 (2.6)	1.1 (1.0)	39 (100)
Hidalgo EQ02	−1.3 (1.7)	1.8 (49.3)	12 (32.4)			
PADIS 2.0	0.1 (1.8)	0.9 (37.5)	24 (66.7)	−0.1 (2.1)	1.0 (0.9)	39 (100)
Obesense <sup>a</sup>	1.1 (2.5)	1.7 (75.1)	16 (43.2)			

AMEQ = activity module Equivital.

<sup>a</sup> Algorithms.

The daily study duration included 30 min of preparation and 90 min of data collection.

Resting EE and HR were averages of measurements obtained between minutes 10 and 14 during the 15 min resting measurement (Compher *et al* 2006). Peak oxygen consumption was estimated based on the formula of Wyss *et al* (2007) from the results of the progressive endurance run the week prior the laboratory measurements. For the analysis of the laboratory tasks, the average 30 s EE and HR data of 120 s duration from seconds 180–300 were used to calculate kilo calories per min (kcal.min<sup>-1</sup>) and beats per minute (b.min<sup>-1</sup>) for each activity. For the field activities, the average 30 s EE and HR over the 90 min measurements were calculated for each device and compared to the respective criterion device. The oxygen consumption and carbon dioxide expiration values were used to estimate EE based on the formula by Peronnet and Massicotte (1991). In addition, based on the HR data from the Hidalgo EQ02, two further formulas for EE estimation in soldiers were investigated and validated. Firstly, the algorithm of Obesense (Gilgen-Ammann *et al* 2017), and secondly, the algorithm activity module Equivital (AMEQ) developed and in use by the Royal Dutch Army (Veenstra *et al* 2018). Some devices recorded total EE (Everion, Garmin Fenix 3, MetaMax 3B, and Obesense) and others recorded physical activity EE (PAEE). In the present study, PAEE was used. In the devices recording total EE, PAEE was computed by calculating the total EE minus the resting EE. After each measurement, the data of each device was downloaded with the respective software and the raw data was exported as an Excel file (Windows 2013, Microsoft Corporation).

In case of technical difficulties, and therefore, missing data of a reference device, the entire measurement, PAEE (5/42, 12%) or HR (3/42, 7%) analysis, was deleted. In case of missing data in an investigated device (<4%), the missing data was interpolated applying the expectation maximization method (Blankers *et al* 2010).

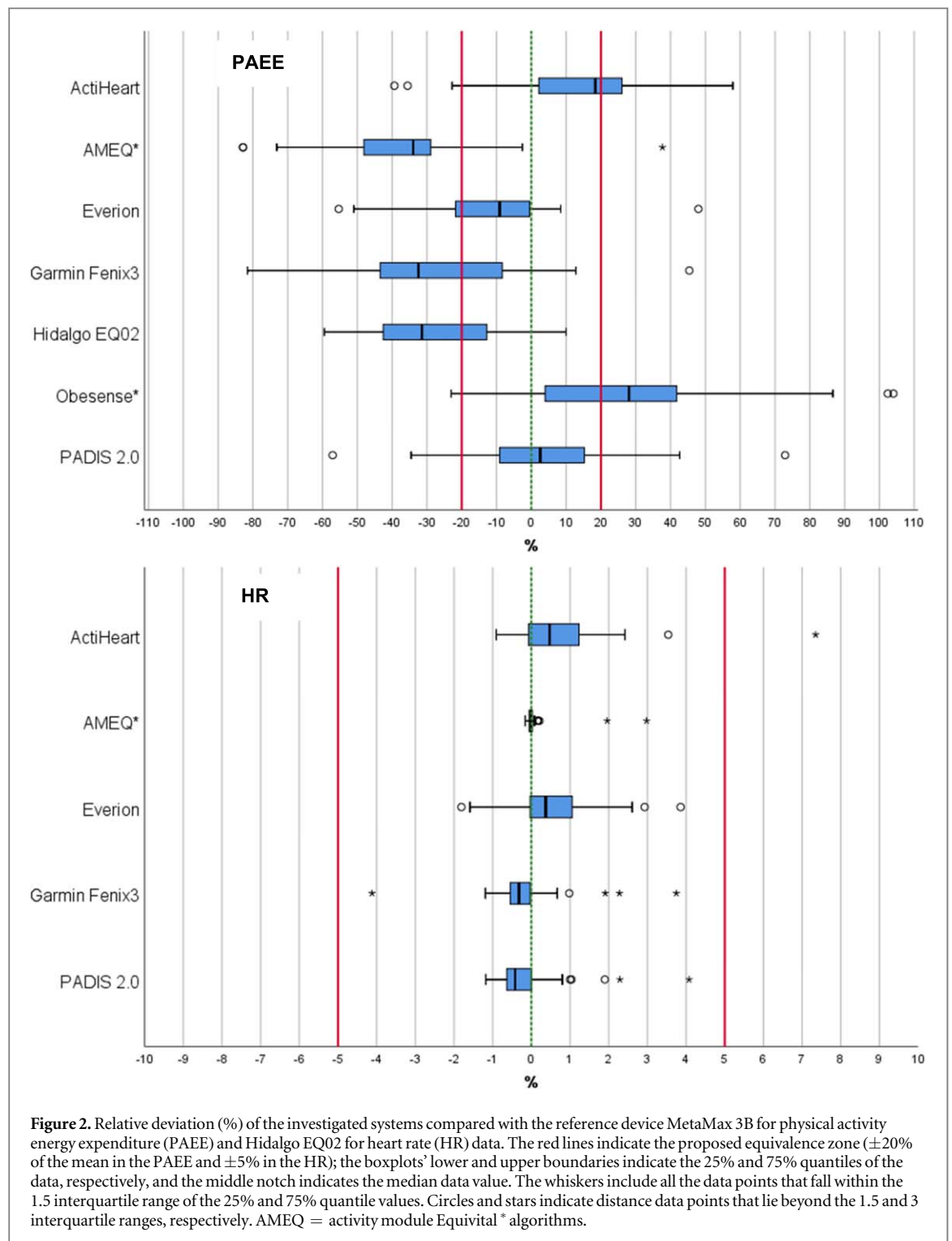
Statistical analyses were performed using IBM SPSS Statistics 24 (IBM Corporation, Armonk, NY, United States). The alpha level was set at  $p < 0.05$ . For PAEE and HR data, the mean absolute errors (MAE) and mean absolute percentage errors (MAPE), Bland–Altman analyses with corresponding systematic biases and 95% limits of agreement (Bland and Altman 1986), Pearson correlations, and % accuracy were conducted for analyses. For the % accuracy, a percentage of the PAEE data within  $\pm 20\%$  and of the HR data within  $\pm 5\%$  from the criterion values was deemed meaningful (Schweizer and Gilgen-Ammann 2018, Gilgen-Ammann *et al* 2019). Furthermore, univariate analyses with post-hoc tests least significant difference were used to detect activities with significant influence on the MAE in the PAEE and HR estimations, respectively.

## Results

Complete day 1 laboratory measurements were obtained for all 23 participants but only 19 participants completed day 2 field measurements due to withdrawal from the study or assignment transfer. From these, useable data was obtained for laboratory ( $n = 22$ ) and field ( $n = 15$ ) PAEE validation, and laboratory ( $n = 22$ ) and field ( $n = 17$ ) HR validation.

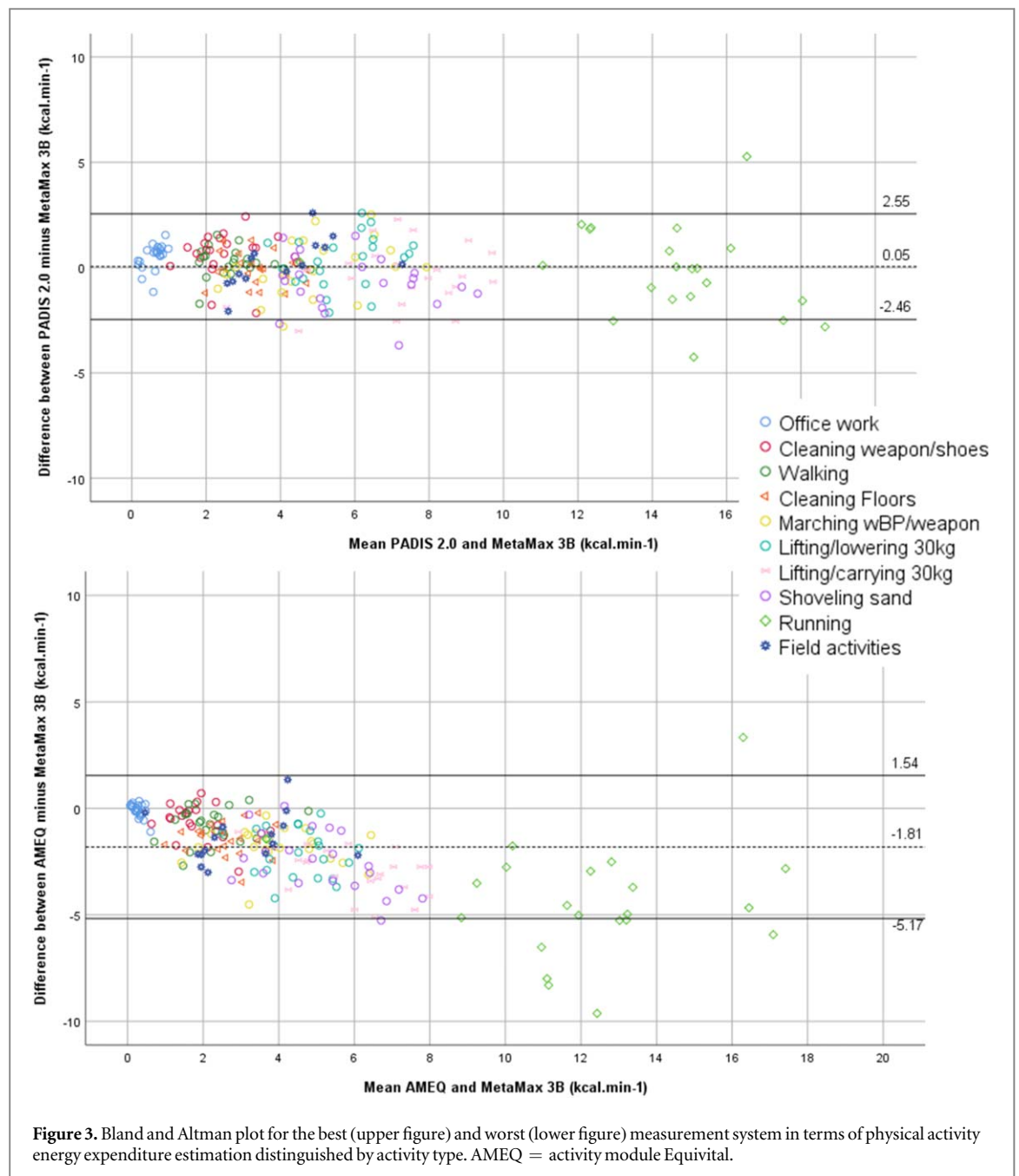
For all laboratory and field measurements combined, the average reference value for PAEE was  $4.6 \pm 1.3$  kcal.min<sup>-1</sup> (range of mean 0.3–13.9 kcal.min<sup>-1</sup>) and for HR  $108.9 \pm 12.7$  b.min<sup>-1</sup> (range of mean 72.6–153.2 b.min<sup>-1</sup>). The recorded systematic errors (limits of agreements) in PAEE estimation ranged from 0.1 ( $\pm 1.8$ ) kcal.min<sup>-1</sup> to −1.7 ( $\pm 1.8$ ) kcal.min<sup>-1</sup> for the PADIS 2.0 and AMEQ, respectively, and in the HR assessment ranged from −0.1 ( $\pm 2.1$ ) b.min<sup>-1</sup> to 0.8 ( $\pm 3.0$ ) b.min<sup>-1</sup> for the PADIS 2.0 and ActiHeart, respectively (table 2). The MAPE in PAEE estimation ranged from 29.9% to 75.1% for the Everion and the





Obesense, respectively, and in the HR assessment from 0.3% to 1.4% for the AMEQ and the Obesense, respectively. The correlations ranged from  $r = 0.647$  to  $0.841$  (all  $p$ -values  $< 0.001$ ) for the PAEE estimations and from  $r = 0.993$  to  $0.999$  (all  $p$ -values  $< 0.001$ ) for the HR measurements. On average, the percentage of participants where the PAEE estimation was within  $\pm 20\%$ , when compared to the MetaMax 3B, ranged from 8.1% (3/37) in the AMEQ to 70.3% (26/37) in the Everion system (figure 2, PAEE) and the percentage of participants where the HR estimations was within  $\pm 5\%$ , when compared to the Hidalgo EQ02, was 100% (39/39) in all systems, except the 97.4% (38/39) in the ActiHeart and the Everion (figure 2, HR). For the laboratory and field measurements separately, the MAE and MAPE in PAEE and HR estimations are presented in figure 1.

The univariate analyses revealed significant differences in the PAEE estimation depending on the activity type ( $F_{9,1488} = 95.53$ ,  $p < 0.01$ ). However, no single activity or activity level (i.e. intensity) could be shown to



increase MAE in the PAEE estimation. Hence, all the investigated devices have their (dis)advantages depending on activity types, however, they differ among each other. Bland and Altman plots that illustrate the best and the worst system, respectively, estimating PAEE with a distinction by activity types are presented in figure 3. For the HR estimation, the univariate analysis resulted in  $F_{10,236} = 18.63$  ( $p = 0.084$ ). Yet, the running activity resulted in significantly higher MAE in HR estimation than all other activity types. Extended tables of device results are presented in an online appendix for PAEE and HR estimations (supplemental tables (available online at [stacks.iop.org/PMEA/42/085008/mmedia](https://stacks.iop.org/PMEA/42/085008/mmedia))).

## Discussion

The present study investigated the validity of different devices and algorithms used in military organizations worldwide to assess PAEE and HR among Swiss Army soldiers. Data outputs were assessed in both structured tests of military tasks of increasing intensity (indoors) and during a normal day of routine military activities (outdoors). Our results showed that the Everion and the PADIS 2.0 were the most precise devices in estimating PAEE overall, yet, with a systematic errors of  $-0.6$  and  $0.1 \text{ kcal.min}^{-1}$ , MAPE of 29.9% and 37.5%, correlations of  $r = 0.841$  and  $0.814$ , and 70.3 and 66.7% of all PAEE estimations within  $\pm 20\%$  of the reference values,

respectively. In terms of HR measurements, all investigated devices demonstrated very good validity with a systematic error of  $-0.1 \text{ b} \cdot \text{min}^{-1}$  in the PADIS 2.0, a MAPE 0.3% in the AMEQ, excellent correlations from  $r = 0.993$  to  $0.999$ , and 3/5 devices having 100% of their HR data within  $\pm 5\%$  of the reference values.

In terms of PAEE estimation, similar error rates and differences between investigated devices and intensity-levels were observed elsewhere. Roos *et al* (2017) validated three commercially available sport watches and corresponding chest belts estimating EE and reported MAPE of 10%–42% when assessed during low- and moderate-intensity running. Also, MAPE of 20.6%, ranging from 9.1% to 31.4%, in the PAEE estimation were observed in the Polar Vantage M compared to the MetaMax 3B when obtained during various activities from sitting in a chair to accomplishing a floorball course (Gilgen-Ammann *et al* 2019). On average, 59.5% of the mean PAEE values were accurate to within 20%, which is comparable to some of the systems investigated in the present study (Everion 70.3%, PADIS 2.0 66.7%, ActiHeart 45.9%). Notably, the Polar Vantage M and some of the systems in our study involve only a single sensor, whereas other systems consist of two sensors (e.g. PADIS 2.0 or Garmin Fenix 3), which may hamper the usability without improving validity. The proprietary algorithms used in the devices to estimate PAEE are not publicly disclosed and the results can only be assessed with empirical testing such as this experiment (Sperlich and Holmberg 2017, Duking *et al* 2020). Historically, PAEE has been derived from activity or from HR but published studies have suggested that a combination of these two measurements, particularly with individual calibration, might produce even better PAEE estimates (Brage *et al* 2007).

Generally, no activity type was revealed to particularly affect PAEE estimation accuracy in any measurement system. This might be explained by the fact that device placement was on different body parts, and i.e. strong arm movement did not play as a significant role as reported elsewhere (Gilgen-Ammann *et al* 2019). However, looking at the single systems, error measurements seem to be device- and activity-type dependent (supplemental tables). In line with this, Dooley *et al* (2017) stated that the examination of overall activities may lead to misinterpretation as differences would cancel each other and show minimal differences compared to a reference value. The existing devices and algorithms used in military organizations worldwide to estimate PAEE have only poor to moderate validity. Notably, only the Everion system showed overall MAPE  $< 30\%$  in the PAEE estimation. Users and commanders must be aware of these errors and take them into considerations while giving instructions or recommendations to personnel when relying on these systems. During military training, personnel often complete periods of intense exercise leading to high PAEE. Valid monitoring of PAEE is essential to applications involving safe limits of training and the prevention of training injuries (Epstein *et al* 1988, Edwards *et al* 2020). It was demonstrated that the mean sustained work intensity of soldiers was close to 50% of their maximal aerobic capacity (Pihlainen *et al* 2014). This is approximately the upper limit of sustainable effort, equivalent to a physical activity level of 2.25–2.50 or EE of  $\sim 4000 \text{ kcal} \cdot \text{d}^{-1}$  observed in many other studies (Wyss *et al* 2012). The devices tested in this study can provide useful quantification of military training intensity and daily workload at a group level but do not appear to be sufficiently accurate and precise for individual guidance.

In terms of HR assessment, the results were excellent in all investigated devices (MAPE  $< 1.4\%$ ). This finding was in line with recent research demonstrating high concordance in the HR measurement during various activities between optical HR monitors and the criterion measure Polar H10 chest strap (Schweizer and Gilgen-Ammann 2018). In the present study running resulted in significantly higher MAE in HR estimation than all the other activity types, which was in contrast to the findings by Schweizer and Gilgen-Ammann (2018), who found no differences, but had the lowest MAE in the running activity. In the present study,  $\pm 5\%$  accuracy was between 97% and 100%, when considering the mean values for each activity. Such high accuracy is required, as for meaningful prediction of heat strain from algorithms that estimate core body temperature from HR (Buller *et al* 2013, Buller *et al* 2021). Also, mental stress and discomfort increase HR without a simultaneous increase in the oxygen consumption, e.g. due to changes in the activity intensity (Lambiase *et al* 2012). Based on the present data, all investigated systems can be recommended for valid HR measurements during different activities.

When choosing a monitoring system, apart from validity, wearing comfort and feasibility should be taken into consideration as well. From a holistic point of view all these aspects have an effect on user compliance. Previously, sensors worn around the chest were rated to have a significantly greater negative impact on soldiers' bodies than sensors worn around the wrist and other body parts (Beeler *et al* 2018, Wyss *et al* 2020). A poor wearing comfort was the most frequently reported negative impact (21.0%), followed by interference with equipment (9.9%), and movement restrictions (7.4%). Considering this information and the present findings, the Everion on the upper arm was found to be the most valid and feasible system to assess PAEE and HR during structured tests of military tasks of increasing intensity (indoors) and during a normal day of routine military activities (outdoors).



## Conclusions

The present study demonstrated excellent validity in all investigated devices in terms of HR assessment (all MAPE < 1.4%) and poor to moderate validity in terms of PAEE estimations (all MAPE between 29.9% and 75.1%) when compared to a respective criterion measure. Overall, the Everion system performed among the best in both the PAEE estimation and HR assessment during structured tests of military tasks and a normal day of routine military activities. Moreover, with a device placement on the upper arm, the system is particularly useful during military services, as it has a good wearing comfort and does not interfere with other equipment.

## Acknowledgments

With the support of the staff at the Infantry Training School as well as Joel Bättig, Florian Herren, Jannik Brunschwiler, Nicolai Muff, and Daniel Agostino the data collection could be performed as planned. Thank you for your help. This study is a collaboration between Switzerland, the Netherlands, United States of America and the United Kingdom through NATO Panel HFM-260, 'Enhancing Warfighter Effectiveness with Wearable Biosensors and Physiological Models'. The authors have confirmed that any identifiable participants in this study have given their consent for publication.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Conflict of interest

The authors have no competing interests to declare.

## Disclaimers

The views expressed in this paper are those of the authors and do not reflect the official policy of their employers or respective governments. Citations of commercial products and trade names in this report do not constitute an official endorsement or approval of the products by the authors and affiliated organisations.

## ORCID iDs

Nadja Beeler  <https://orcid.org/0000-0001-7276-0752>

## References

- Akintola A A, van de Pol V, Bimmel D, Maan A C and van Heemst D 2016 Comparative analysis of the equivil EQ02 lifemonitor with holter ambulatory ECG device for continuous measurement of ECG, heart rate, and heart rate variability: a validation study for precision and accuracy *Frontiers Physiol.* **7** 391
- Beeler N, Roos L, Delves S K, Veenstra B J, Friedl K, Buller M J and Wyss T 2018 The wearing comfort and acceptability of ambulatory physical activity monitoring devices in soldiers *IJSE Trans. Occup. Ergon. Hum. Factors* **6** 1–10
- Bent B, Goldstein B A, Kibbe W A and Dunn J P 2020 Investigating sources of inaccuracy in wearable optical heart rate sensors *NPJ Digit. Med.* **3** 18
- Bland J M and Altman D G 1986 Statistical methods for assessing agreement between two methods of clinical measurement *Lancet* **1** 307–10
- Blankers M, Koeter M W and Schippers G M 2010 Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers *J. Med. Internet Res.* **12** e54
- Brage S, Ekelund U, Brage N, Hennings M A, Froberg K, Franks P W and Wareham N J 2007 Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity *J. Appl. Physiol.* **103** 682–92
- Buller M J, Delves S K, Fogarty A L and Veenstra B J 2021 On the real-time prevention and monitoring of exertional heat illness in military personnel *J. Sci. Med. Sport* (<https://doi.org/10.1016/j.jsams.2021.04.008>)
- Buller M J *et al* 2013 Estimation of human core temperature from sequential heart rate observations *Physiol. Meas.* **34** 781–98
- Burrell C, Love R J and Stergiopoulos S 2016 Integrated Physiological Monitoring (Vol. DRDC-RDDC-2016-R027): Defence Research and Development Canada – Toronto Research Centre [https://cradpdf.drdc-rddc.gc.ca/PDFS/unc256/p804911\\_A1b.pdf](https://cradpdf.drdc-rddc.gc.ca/PDFS/unc256/p804911_A1b.pdf)
- Carrier B, Creer A, Williams L R, Holmes T M, Jolley B D, Dahl S, Weber E and Standiford T 2020 Validation of garmin fenix 3 HR fitness tracker biomechanics and metabolics (VO2max) *J. Meas. Phys. Behav.* **3** 331–7
- Compher C, Frankenfield D, Keim N and Roth-Yousey L 2006 Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review *J. Am. Diet. Assoc.* **106** 881–903

- Crouter S E, Churilla J R and Bassett D R Jr 2008 Accuracy of the Actiheart for the assessment of energy expenditure in adults *Eur. J. Clin. Nutr.* **62** 704–11
- Dannecker K L, Sazonova N A, Melanson E L, Sazonov E S and Browning R C 2013 A comparison of energy expenditure estimation of several physical activity monitors *Med. Sci. Sports Exercise* **45** 2105–12
- Dooley E E, Golaszewski N M and Bartholomew J B 2017 Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices *JMIR Mhealth Uhealth* **5** e34
- Duking P, Giessing L, Frenkel M O, Koehler K, Holmberg H C and Sperlich B 2020 Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study *JMIR Mhealth Uhealth* **8** e16716
- Edwards V, Myers S, Siddall A, Needham-Beck S, Powell S, Jackson S, Greeves J, Wardle S and Blacker S 2020 Energy availability of officer cadets during british army training *ICSPP 5th Int. Congress on Soldier Physical Performance (Quebec City, Canada)* p 242
- Epstein Y, Rosenblum J, Burstein R and Sawka M N 1988 External load can alter the energy cost of prolonged exercise *Eur. J. Appl. Physiol. Occup. Physiol.* **57** 243–7
- Friedl K E 2018 Military applications of soldier physiological monitoring *J. Sci. Med. Sport* **21** 1147–53
- Gilgen-Ammann R, Koller M, Huber C, Ahola R, Korhonen T and Wyss T 2017 Energy expenditure estimation from respiration variables *Sci. Rep.* **7** 15995
- Gilgen-Ammann R, Schweizer T and Wyss T 2019 Accuracy of the multisensory wristwatch polar vantage's estimation of energy expenditure in various activities: instrument validation study *JMIR Mhealth Uhealth* **7** e14534
- Lambiase M J, Dorn J, Chernega N J, McCarthy T F and Roemmich J N 2012 Excess heart rate and systolic blood pressure during psychological stress in relation to metabolic demand in adolescents *Biol. Psychol.* **91** 42–7
- Lee J M, Kim Y and Welk G J 2014 Validity of consumer-based physical activity monitors *Med. Sci. Sports Exercise* **46** 1840–8
- Peronnet F and Massicotte D 1991 Table of nonprotein respiratory quotient: an update *Can. J. Sport Sci.* **16** 23–9
- Pihlainen K, Santtila M, Hakkinen K, Lindholm H and Kyrolainen H 2014 Cardiorespiratory responses induced by various military field tasks *Mil. Med.* **179** 218–24
- Roos L, Taube W, Beeler N and Wyss T 2017 Validity of sports watches when estimating energy expenditure during running *BMC Sports Sci. Med. Rehabil.* **9** 22
- Rosendal L, Langberg H, Skov-Jensen A and Kjaer M 2003 Incidence of injury and physical performance adaptations during military training *Clin. J. Sport Med.* **13** 157–63
- Schweizer T and Gilgen-Ammann R 2018 Accuracy of the optical heart rate device Polar OH1 during rest and exercise *CISM Sport Sci. Abstr.* <https://www.milsport.one/cism-academy/sport-science>
- Sperlich B and Holmberg H C 2017 Wearable, yes, but able...?: it is time for evidence-based marketing claims! *Br. J. Sports Med.* **51** 1240
- Veenstra B J, Wyss T, Roos L, Delves S K, Buller M and Beeler N 2018 An evaluation of measurement systems estimating gait speed during a loaded military march over graded terrain *Gait Posture* **61** 204–9
- Vogler A J, Rice A J and Gore C J 2010 Validity and reliability of the cortex MetaMax3B portable metabolic system *J. Sports Sci.* **28** 733–42
- Wyss T and Mader U 2011 Energy expenditure estimation during daily military routine with body-fixed sensors *Mil. Med.* **176** 494–9
- Wyss T, Marti B, Rossi S, Kohler U and Mäder U 2007 Assembling and verification of a fitness test battery for the recruitment of the Swiss army and nation-wide use *Schweiz. Z. Sportmedizin Sporttraumatologie* **55** 126–31
- Wyss T, Roos L, Friedl K, Buller M, Delves S K and Veenstra B J 2020 The feasibility of ambulatory physical activity monitoring devices in studies on soldiers *CISM Sport Sci. Abstr.* <https://www.milsport.one/cism-academy/sport-science>
- Wyss T, Roos L, Hofstetter M C, Frey F and Mader U 2014 Impact of training patterns on injury incidences in 12 Swiss army basic military training schools *Mil. Med.* **179** 49–55
- Wyss T, Scheffler J and Mader U 2012 Ambulatory physical activity in Swiss army recruits *Int. J. Sports Med.* **33** 716–22