



Summer Meeting, 10–12 July 2018, Getting energy balance right

How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies

R. O'Driscoll¹, J. Turicchi¹, S. Scott¹, K. Beaulieu¹, J. Matu², K. Deighton³,
G. Finlayson¹ and R.J. Stubbs¹

¹*Appetite Control and Energy Balance Group, School of Psychology, University of Leeds, Leeds, LS2 9JT*, ²*Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, LS2 9JT* and ³*Institute for Sport, Physical Activity & Leisure, Leeds Beckett University, Leeds, LS6 3QS.*

Qualifying energy intake (EI) and energy expenditure (EE) remains a central challenge to nutritional science. Precise, accurate estimate of day-to-day EI and EE remain elusive. Wrist and arm-worn activity monitors have become popular methods to estimate EE⁽¹⁾ for research and consumer purposes. However, their accuracy compared with criterion measures remains uncertain. Recent devices include triaxial accelerometers, thermometers, evaporative heat loss sensors and photoplethysmography heart rate sensors⁽²⁾, which may improve EE estimates over accelerometry alone. This meta-analysis of criterion validation studies was conducted to determine the validity of current devices and technologies.

SportDISCUS, PubMed, SCOPUS, Medline, PsycINFO, EMBASE and CINAHL were searched for studies published before January 2018. We included studies validating EE estimates from wrist or arm-worn activity monitors against criterion measures (indirect calorimetry, room calorimeters and doubly labelled water) in healthy adult populations. A random effects meta-analysis was performed to establish Hedges' *g* (ES) and 95 % confidence intervals (95 % CI). Moderator analyses were conducted to determine the benefit of the inclusion of additional sensors, and to compare the accuracy of research-grade devices to consumer devices.

60 studies (104 effect sizes) comparing 41 devices were included in the meta-analysis. The pooled mean estimate of EE by all devices showed a significant underestimation relative to criterion measures (ES: -0.23, 95 % CI: -0.44 to -0.03; *p* = 0.03). The Garmin vivofit (ES: -1.09, 95 % CI: -1.61 to -0.56; *p* < 0.001), SenseWear Armband (ES: -0.31, 95 % CI: -0.62 to -0.01; *p* = 0.04) and the Jawbone UP24 (ES: -1.16, 95 % CI: -1.79 to -0.53; *p* < 0.001) were the only devices that significantly underestimated relative to criterion measures across all activities. Large heterogeneity was observed for many devices ($I^2 \geq 50\%$). Combining heart rate or heat sensing technology with accelerometry decreased the error in most activity types, aside from ambulatory activity where a difference was observed between sensors (*p* = 0.007), in which accelerometry was the only level of sensor not different from criterion measures (ES: -0.23, 95 % CI: -0.513 to 0.057; *p* = 0.12). Research-grade devices were statistically less accurate than commercial devices during ambulatory activity (*p* = 0.036) and sedentary tasks (*p* = 0.006) but were more accurate for total EE (*p* = 0.02).

Estimates of EE from wearable devices are heterogeneous and no device performs sufficiently across all activity types. Addition of physiological sensors generally improves estimates of EE and in some activities commercial level devices outperform research grade-devices. These data highlight the need to improve EE estimates from wearable devices for research and consumer purposes. PROSPERO CRD42018085016

1. Evenson KR, Goto MM & Furberg RD (2015) *Int J Behav Nutr Phys Act* **12**, 159.
2. Stahl SE, An HS, Dinkel DM, *et al.* (2016) *BMJ Open Sport Exerc Med* **2**, 1.