Title: A smartphone activity monitor that accurately estimates energy expenditure

Authors: Haedo Cho[1] and Patrick Slade[1,2]*

Affiliations:
[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University; Boston, MA, USA.
[2]Kempner Institute for the Study of Artificial and Natural Intelligence, Harvard University; Boston, MA, USA.
*Corresponding author. Email: slade@seas.harvard.edu

Abstract: Physical inactivity is the fourth largest cause of global mortality. Health organizations have requested a tool to objectively measure physical activity because many relationships between activity and health are not clearly understood. Existing activity monitors are either unsuitable for large-scale use or have substantial error. We present OpenMetabolics, a biomechanically-informed activity monitor that employs a smartphone in a pants pocket which measures leg motion to estimate energy expenditure. OpenMetabolics uses a data-driven machine learning model to capture the relationship between underlying leg muscle activity and energy expended during common physical activities. OpenMetabolics estimated energy expenditure with a 18% cumulative error across all real-world activities, approximately two times lower than existing tools. We developed a pocket motion artifact correction model to accurately monitor energy expenditure when the smartphone is in a pocket of various types of clothing. A week-long, at-home monitoring study highlighted individual and population-level activity patterns across various timescales. We have made the data, code, and smartphone application open source. This accurate and accessible activity monitor could be deployed for large-scale studies with many patient populations to relate activity to health outcomes, inform health policy, and develop interventions.

Main Text: Physical activity is a powerful intervention for many health conditions, but activity monitoring tools limit studies relating activity to health outcomes. Regular physical activity improves many aspects of health including musculoskeletal function[1], mental health[2], weight management[3], disease prevention[4], cognitive function[5], and sleep[6]. Understanding the relationship between activity and health outcomes is important for developing interventions, evaluating efficacy of clinical trials, or advancing health research. However, many of these relationships remain unclear, such as weight loss[7], quality of life in dementia and Alzheimer's disease[8,9], and cognitive function in multiple sclerosis[10]. The World Health Organization[11] and Physical Activity Guidelines Advisory Committee[1] have requested improved activity monitoring tools to address the difficulty of connecting activity and health outcomes.

Health organizations have specified that activity monitoring tools should accurately capture the frequency, duration, intensity, and type of physical activity[1,11]. Activities of any duration contribute to health benefits[12–14]. Frequent estimation, ideally monitoring each step, is necessary to capture all activity like the many short walking bouts that comprise most daily movement[15]. Accurately measuring activity intensity is important as it can vary dramatically even within one type of activity, such as walking at different speeds or inclines[16]. The activity intensity can be quantified using discrete categories like moderate-to-vigorous intensity[11], units of movement like steps[17,18], or direct and continuous measures of energy expenditure[16,19,20]. Monitoring technology should be widely accessible to improve health equity[1] and be suitable for longitudinal studies as many health outcomes are influenced by long-term physical activity[21].

Existing activity monitoring tools have trade-offs in monitoring capabilities. Respirometry[19] and doubly labeled water[20] are the gold-standard, laboratory-based methods for accurately estimating energy expenditure, but are not suitable for large-scale deployment due to their cost and inconvenience for everyday use. Self-report surveys are commonly used to monitor activity but they rely on subjective responses based on coarse categories like light, moderate and vigorous intensity, leading to substantial errors[22]. Wearable devices may offer a practical and affordable method to assess physical activity using objective activity metrics[11,23]. Pedometers approximate activity intensity based on total step counts[24] or cadence[25], but they are restricted to a few walking-related activities. Accelerometer-based devices, like the ActiGraph, are worn on the leg, waist, or wrist and quantify activity intensity by counting the number of acceleration measurements that reach a threshold. ActiGraph studies report high errors in estimating energy expenditure[26,27], possibly due to thresholds not holding consistently across different wear locations, populations, and activities[28,29]. Smartwatches are practical for daily use but are predominantly used in high-income countries[23], provide limited access to algorithms or raw data[22], and estimate energy expenditure with large errors[30–32]. Researchers have developed many different activity monitors by integrating custom wearable sensors and utilizing data-driven models[16,33–35]. For example, our previous work used a wearable system with two custom sensors on the leg to estimate energy expenditure during laboratory experiments[16]. However, custom sensor sets are not capable of large-scale deployment and infrequently validated during real-world conditions.

Smartphones may be a promising activity monitor, but existing methods have not been thoroughly evaluated for real-world use with a diverse participant cohort. Smartphones are used by approximately 70% of the global population[36], provide informative sensor measurements related to physical activity[16], and can download custom software applications with openly accessible algorithms. However, several smartphone activity monitoring methods do not collect ground-truth energy expenditure measurements and cannot evaluate estimation accuracy[37–39]. Some studies only test young and healthy participants[39,40], or use the same participants for training and testing their model[39]. Most studies focus on a few activities rather than many common activities including daily living tasks[37–41].

We present OpenMetabolics, a data-driven method that estimates energy expenditure during real-world activity using leg motion data from a smartphone carried in a pants pocket. OpenMetabolics estimates energy expenditure more accurately than existing tools, provides an estimate per gait cycle, monitors many common physical activities, and can be easily deployed as a smartphone application. OpenMetabolics relies on a data-driven model that captures the relationship between leg motion and the energy expended during common physical activities. Active energy expended during common activities that involve ambulation like walking, stair climbing, running, and biking is primarily due to leg muscle activity[42]. Leg motion encapsulates information related to energy expenditure because leg accelerations are directly related to muscle forces, and thus muscle activity, based on Newton's second law of motion[16].

The subsequent paragraphs describe the series of experiments performed to develop and validate each component of OpenMetabolics and provide an extensive comparison to state-of-the-art activity monitoring tools. This includes training a data-driven machine learning model that uses leg motion to estimate energy expenditure once per gait cycle. We validated the accuracy of the data-driven model during real-world activity monitoring by having subjects not included in the training data wear a smartphone strapped to their thigh. We chose to isolate the smartphone on the thigh for metabolic testing to provide ground-truth thigh motion data. Respirometry experiments require substantial time, evaluating the smartphone in the pocket of a single type of pants for each participant would provide limited examples of clothing. We then proposed a pocket motion correction model that minimizes motion artifacts from a smartphone carried in the pocket of different types of clothing. We validated that a smartphone in a pocket has the same monitoring accuracy as a smartphone strapped to the thigh by evaluating the system with many different participants and articles of clothing. Finally, we deployed OpenMetabolics to continuously monitor real-world activity throughout one week.

OpenMetabolics processes smartphone data to estimate energy expenditure once per gait cycle. A smartphone in the pocket provides inertial measurements from the accelerometer and gyroscope (Fig. 1a). An orientation calibration algorithm aligns the smartphone worn in a pocket with the thigh to provide a consistent frame of reference for measurements (Fig. 1b and Supplementary Fig. 1). The angular velocity data is segmented by gait cycle (Fig. 1c) and downsampled to a fixed size (Fig. 1d). Motion artifacts caused by phone movement in the pocket are removed with a pocket motion correction model to isolate leg motion (Fig. 1e). A data-driven model estimates energy expenditure using angular velocity data and the subject's weight and height (Fig. 1f and Supplementary Fig. 2). OpenMetabolics estimates energy expenditure once per gait cycle during common activities (Fig. 1g).

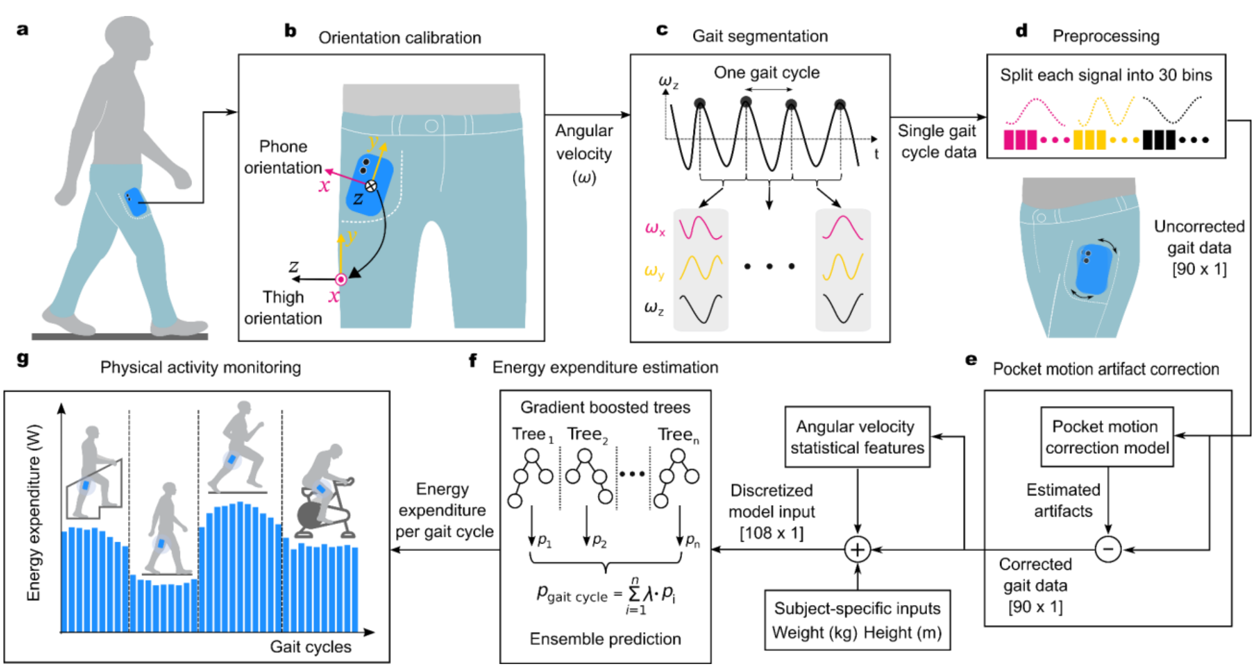

Fig. 1. OpenMetabolics energy expenditure estimation method. a, A participant walks with a smartphone in their pocket. **b,** An orientation calibration algorithm aligns the smartphone data with the thigh's frame of reference during each bout of activity, regardless of the smartphone's orientation in the pocket. **c,** Individual gait cycles are segmented by detecting peaks in the sagittal plane angular velocity (ω_z). **d,** Each axis of angular velocity data from a single gait cycle is

downsampled to a fixed size of 30 values. **e,** A linear regression model estimates motion artifacts caused by the phone shifting in the pocket during walking and removes these artifacts from the uncorrected gait data. **f,** A pre-trained data-driven model, an ensemble of gradient boosted trees, estimates the energy expenditure once per gait cycle. This model takes an input of the corrected gait data, statistical features of the gait data, and the subject's height and weight. Each tree estimate (p) is aggregated using a learning rate (λ) to produce the ensemble estimate. **g,** OpenMetabolics estimates energy expended during real-world activities.

We trained a data-driven machine learning model that incorporates the cyclic motion during physical activity to estimate energy expenditure using inputs from a smartphone's angular velocity measurements. These time series data are segmented into gait cycles to provide a simple structure for extracting information from leg motion. Estimating energy expenditure from a data-driven model with a gait cycle of input data was more accurate and required less computation than a time series model using a sliding window of input data[43]. This gait cycle structure in the data-driven model can relate the differences in leg kinematics to energy expended during varying physical activities and levels of intensity, enabling task-agnostic regression[16]. We trained an ensemble model of gradient boosted trees, combining many shallow decision trees to improve robustness against overfitting[44]. The model was trained on a previous laboratory dataset[16] of thigh kinematics and steady-state metabolic rate data from 36 participants (23 men 13 women; age, 31 ± 11 yr; body mass, 71.6 ± 12.0 kg; height, 1.73 ± 0.07 m; body mass index, 23.9 ± 3.3) performing activities like walking, running, stair climbing, and stationary biking at various intensities[16] (Supplementary Table 1).

OpenMetabolics estimated cumulative energy expenditure more accurately than existing physical activity monitoring tools during naturalistic real-world walking. Participants walked on a public sidewalk wearing a respirometry system, a commercial smartwatch, a wrist-worn heart rate monitor, a pedometer, a thigh-based accelerometer, and a smartphone strapped to their thigh (Fig. 2a and 2b). A new group of participants (n=28, 16 men 12 women; age, 35 ± 14 yr; body mass, 72.1 ± 13.1 kg; height, 1.72 ± 0.08 m; body mass index, 24.3 ± 3.5) was recruited to perform naturalistic walking bouts in a real-world setting for 16 minutes, taking approximately 33% of the average daily steps for U.S. adults[45] (Fig. 2c). These participants were separate from those in the training dataset (Supplementary Fig. 3). Participants varied their walking speed and bout duration in response to ecologically relevant[46] audio prompts[47], causing them to self-select walking speeds that closely matched the expected ground-truth distributions for the speed and duration of the walking bouts (Fig. 2d and 2e). OpenMetabolics had the lowest average error across all physical activities (Fig. 2f). OpenMetabolics had a cumulative energy expenditure error of 13% for real-world walking (multiple comparisons, n=28, *P<0.05, **P<0.005), significantly less than existing physical activity monitors (Fig. 2g). OpenMetabolics estimated cumulative energy expenditure with minimal bias and low variability compared to ground-truth respirometry, indicating more accurate and consistent estimation across a diverse adult population than existing activity monitors (Supplementary Fig. 4).

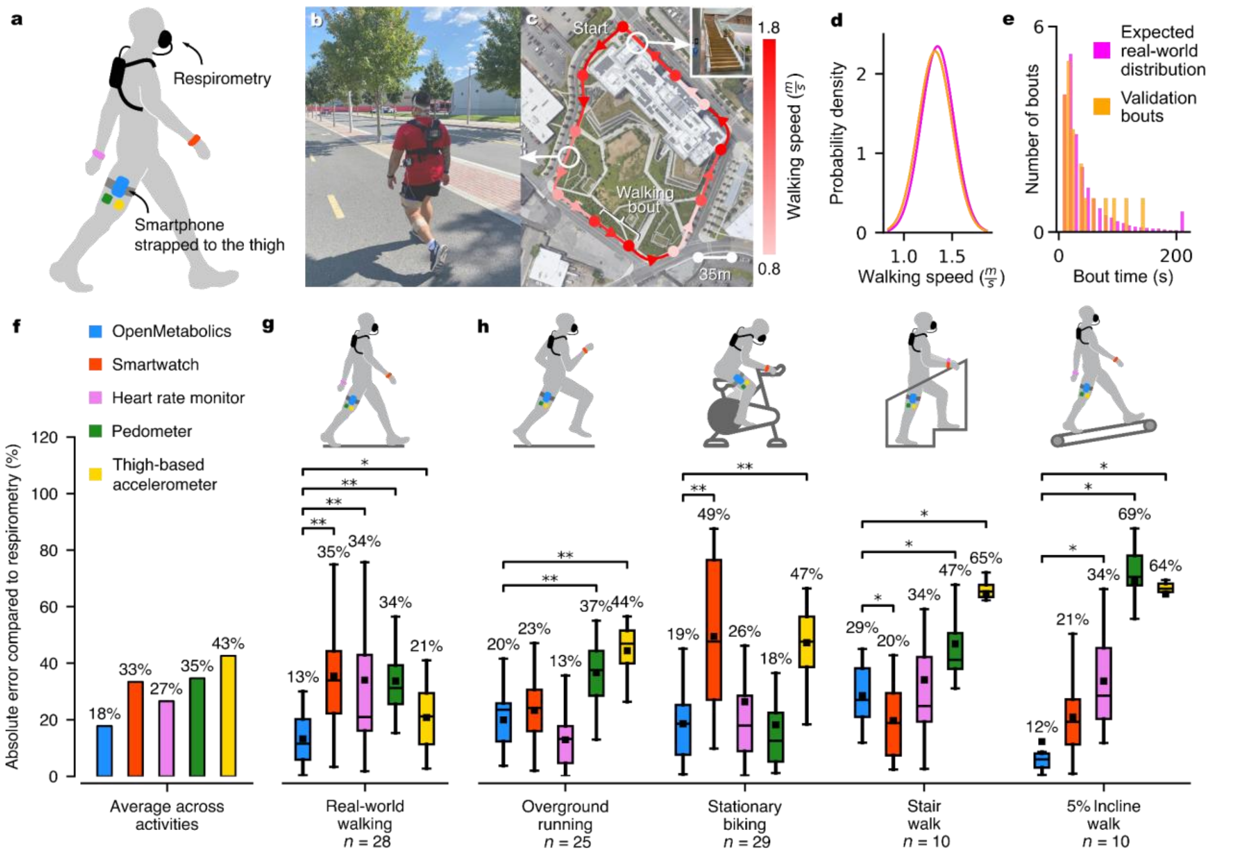

Fig. 2. Evaluating OpenMetabolics' real-world energy expenditure estimation using a smartphone strapped to the thigh. a, OpenMetabolics was compared to a commercial smartwatch and portable respirometry system. **b,** Participants performed physical activity experiments on public sidewalks. **c**, Map of the 650-meter course used for evaluation. Participants performed 20 bouts of walking through the course while following ecologically relevant[46] audio prompts[47] to elicit naturalistic walking. Each lap included two flights of stair climbing. **d,e,** Distribution of self-selected walking speeds **(d)** and walking bout durations **(e)** during the experiment, compared with previously recorded distributions of real-world walking data[15]. **f,** OpenMetabolics had the lowest weighted absolute error across all physical activities compared to existing physical activity monitors. **g,** OpenMetabolics had significantly lower cumulative energy expenditure error than a commercial smartwatch, heart rate model, pedometer, and thigh-based accelerometer during real-world walking, compared to respirometry (multiple comparison tests, n=28, *P<0.05, **P<0.005). **h,** OpenMetabolics had lower absolute error in cumulative energy expenditure across several physical activities (multiple comparison tests, n=10, *P<0.05, **P<0.005). The error for each activity was weighted based on the number of participants. The boxes show the interquartile range, with a line at the median and a dot and percentage value representing the mean. Whiskers extend to 1.5 times the interquartile range.

OpenMetabolics estimated cumulative energy expenditure more accurately than existing physical activity tools during several physical activities and activities of daily living. Participants performed overground running and stair walking at a self-selected pace, and stationary biking and incline walking at a steady state pace, all in 6-minute bouts (Fig. 2h). The smartwatch's high error may indicate heart rate and cadence estimated from wrist motion are less informative signals of energy expenditure[16]. The pedometer and thigh-based accelerometer had high errors during walking on stairs and inclines because cadence and acceleration intensity may not have a clear relationship with effort (multiple comparisons, n=10, *P<0.05). For example, slow walking on flat ground and walking up stairs may yield similar cadence and acceleration intensities but require different energy expenditures. Thigh-based accelerometers may benefit from using the gait cycle structure of motion[16] and gyroscope measurements[26] to better estimate energy expenditure. A data-driven pedometer model and data-driven heart rate model, trained on the same dataset with the same modeling approach as OpenMetabolics, performed similarly to the existing pedometer[25] and heart rate models[48] (Fig. 2g and Supplementary Fig. 5). This indicates that OpenMetabolics' low error is not due to the datasets or model structure, but instead may be due to the information encoded in cyclic leg motion that captures underlying muscle activity and energy expenditure. A linear mixed model indicated that the participants' age, gender, and body mass index had no significant effect on OpenMetabolics' error during the real-world experiments (Supplementary Table 2, n=38, P=0.365, P=0.099, P=0.276). Participants performed several common activities of daily living including organizing their home, cleaning, eating, and maintaining personal hygiene[49]. OpenMetabolics and the existing physical activity monitors had similar estimation errors during activities of daily living, indicating that monitoring leg motion is a substantial contributor to energy expenditure in everyday activities (Supplementary Fig. 6 and Supplementary Video 1, multiple comparisons, n=10, *P<0.05).

A pocket motion correction model minimizes motion artifacts caused by the smartphone moving in the pocket to enable accurate energy expenditure estimation when people wear different clothes. A smartphone's movement within the pocket introduces unwanted motion, distorting the ground-truth smartphone measurements of leg motion. This uncorrected smartphone data contains cyclic and consistent motion artifacts throughout each gait cycle[50]. We trained a linear regression model to remove motion artifacts related to smartphone movement in the pocket, so the corrected smartphone data only reflects leg motion (Fig. 3a). Participants performed a series of 20 second walking bouts at 5 self-selected speeds following ecological audio prompts while wearing each of 4 different types of clothing to evaluate the pocket motion correction model (n=10, 5 men 5 women; age, 26 ± 3 yr; body mass, 71.7 ± 9.1 kg; height, 1.72 ± 0.08 m; body mass index, 24.1 ± 1.3). The pocket motion correction model significantly reduced the angular velocity root mean square error by 28% compared to the ground-truth smartphone strapped to the thigh (Fig. 3b; two-sided Wilcoxon signed-rank test, n=10, **P<0.005), with the largest reduction observed when participants wore athletic shorts (Supplementary Fig. 7). The corrected smartphone data had a significantly lower relative difference in energy expenditure estimation across all types of clothing compared to the uncorrected smartphone data, indicating the pocket motion correction model effectively reduced the systematic bias in estimation caused by motion artifacts (Fig. 3c; two-sided Wilcoxon signed-rank test, n=10, **P<0.005; Supplementary Fig. 8). This approach may generalize to work well for many clothes with an extensive dataset to accurately estimate energy expenditure for people wearing any type of clothing in everyday life. Collecting leg motion data at home with a smartphone can personalize the pocket motion correction model which may improve estimates of motion artifacts for specific items of clothing (n=1, Supplementary Fig. 9).

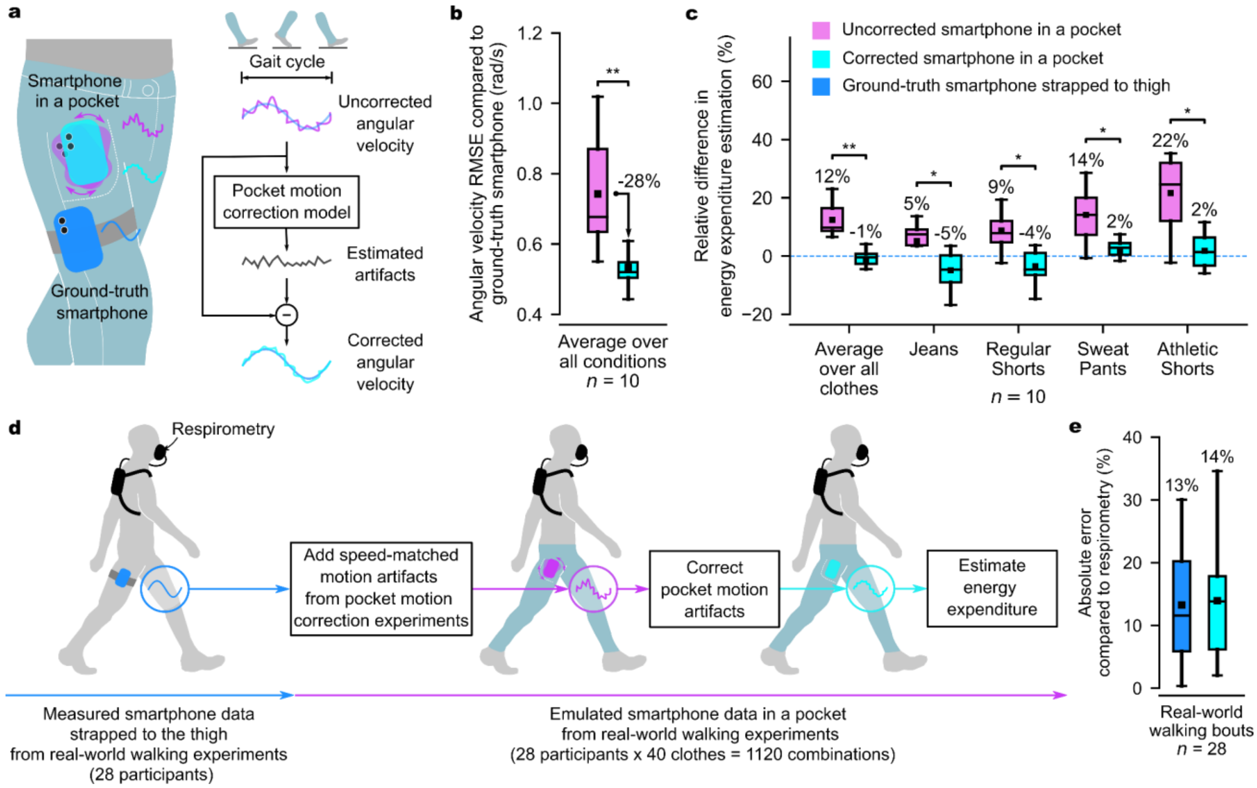

Fig. 3. Correcting and emulating motion artifacts for a smartphone carried in a pocket during real-world walking experiments. a, Participants performed a series of 20 second walking bouts at 5 self-selected speeds following ecological audio prompts while wearing each of 4 different types of clothing. Motion data was measured with a smartphone in the pocket and a smartphone strapped to the thigh. The pocket motion correction model was trained to estimate motion artifacts caused by the smartphone moving in the pocket during each gait cycle. The motion artifacts are the difference in angular velocity between the ground-truth smartphone strapped to the thigh, which reflects true leg motion, and the smartphone in the pocket, which provides uncorrected data. **b,** The pocket motion correction model significantly reduced the motion artifacts, measured with angular velocity root mean square error (RMSE), by 28% across all types of pants and walking speeds (two-sided Wilcoxon signed-rank test, n=10, **P<0.005). **c,** The relative difference in energy expenditure estimation was significantly reduced by correcting smartphone data (two-sided Wilcoxon signed-rank test, n=10, **P<0.005). This estimation improvement was true across all clothing conditions, especially looser fitting clothes that had larger motion artifacts (multiple comparison, n=10, *P<0.05). **d,** We added previously collected speed-matched motion artifacts (Fig. 3a-c) to the smartphone strapped to the thigh in the real-world walking experiments (Fig. 2) to emulate an uncorrected smartphone in a pocket for many participants and clothes. OpenMetabolics estimated energy expenditure during real-world walking experiments using the corrected smartphone data by removing motion artifacts with the pocket motion correction model. **e,** There was no significant difference between OpenMetabolics' energy expenditure estimates when using the corrected smartphone data or the ground-truth smartphone data during real-world walking experiments, indicating a smartphone in the pocket is an effective monitoring tool (two-sided Wilcoxon signed-rank test, n=28, P=0.245). The boxes show the interquartile range, with a line at the median and a dot and percentage value representing the mean. Whiskers extend to 1.5 times the interquartile range.

OpenMetabolics had similar real-world energy expenditure estimation errors when using measurements from a smartphone with pocket motion artifacts or a ground-truth smartphone strapped to the thigh, indicating feasibility for broad deployment. Evaluating OpenMetabolics' monitoring performance when a smartphone is carried in a pocket by many participants with many types of clothing is crucial for validation. Unfortunately, respirometry requires approximately 15 minutes of data collection for representing real-world activity, totalling hundreds of hours of testing to evaluate a few dozen participants, each wearing several types of clothing. However, collecting only smartphone data to isolate pocket motion artifacts for different participants and types of clothing requires approximately 15 seconds of data per condition, scaling favorably for experimental collection (Fig. 3c). We augmented the smartphone data strapped to the thigh from the real-world respirometry experiment (Fig. 2) with many different pocket motion artifacts recorded in separate experiments (Fig. 3a) to emulate a real-world respirometry experiment where participants wore many types of clothing (Fig. 3d). This emulation process resulted in 1120 possible combinations of 28 participants and 40 clothes, providing a rigorous evaluation of the pocket motion correction model. The OpenMetabolics energy expenditure estimates using the emulated smartphone in the pocket showed no statistical difference in cumulative error than the measured smartphone strapped to the thigh (Fig. 3e; two-sided Wilcoxon signed-rank test, n=28, P=0.245). OpenMetabolics can accurately estimate energy expenditure when carried in the pocket of many types of clothing because the pocket motion correction model removes estimation bias due to clothing motion artifacts.

OpenMetabolics captures trends in energy expenditure, from individuals to populations, and across timescales, providing a capable tool for making many types of health inferences. A week-long monitoring study with 10 participants (5 men 5 women; age, 26 ± 4 yr; body mass, 66.5 ± 7.1 kg; height, 1.71 ± 0.07 m; body mass index, 22.7 ± 2.0) showcased OpenMetabolics' activity monitoring in the real world. The high granularity of energy expenditure estimates once per gait cycle reveals temporal patterns of activity within each day and throughout the week (Fig. 4a). These daily activity patterns could be used to investigate how a person's lifestyle, work, and environment impacts their energy expenditure. For example, the majority of energy expenditure for one representative participant occurred during commuting hours (Fig. 4b). OpenMetabolics found participants' physical activity varied throughout the week, capturing individual differences in activity level (Fig. 4c) and their unique distribution of physical activity intensity (Fig. 4d). OpenMetabolics can also monitor population scale trends in physical activity (Fig. 4e). Participants had the lowest physical activity level on Sunday compared to the weekday average (two-sided Wilcoxon signed-rank test, n=10, *P<0.05), which matches the results of a previous longitudinal study[51].

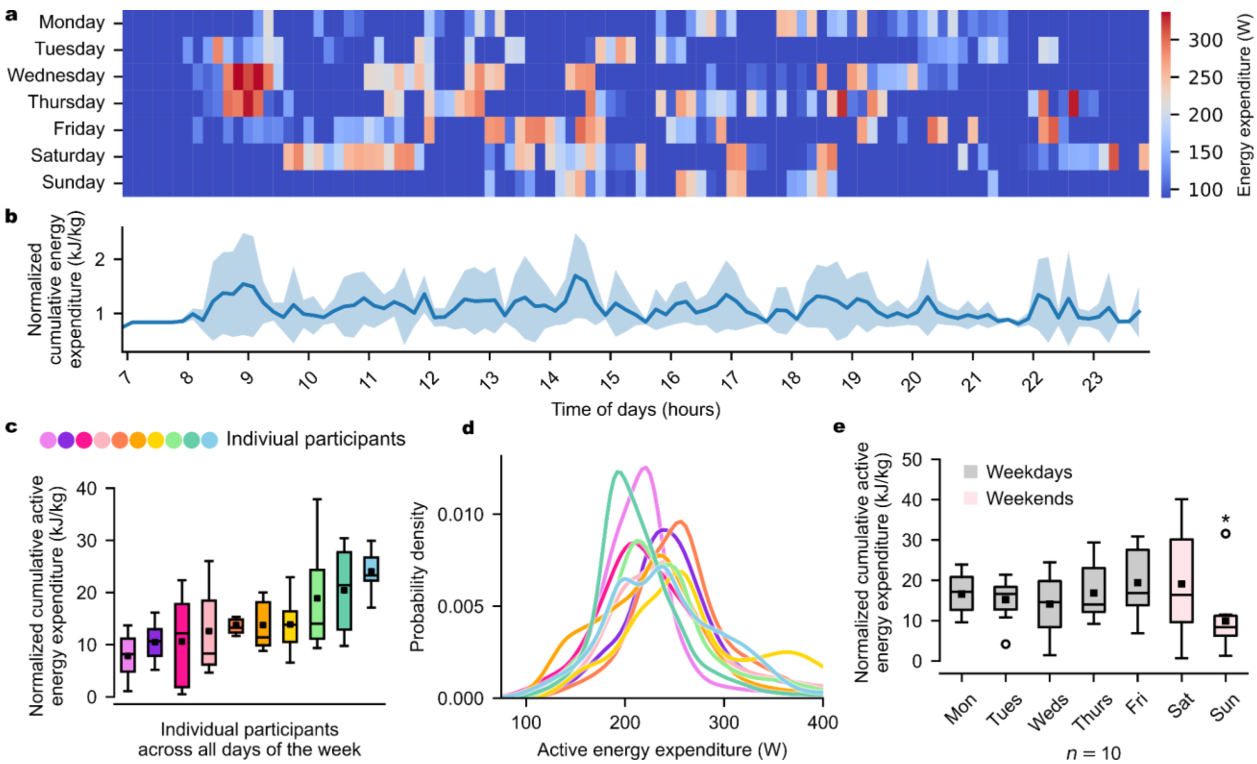

Fig. 4. Real-world activity monitoring with OpenMetabolics. a, OpenMetabolics provides granular energy expenditure estimates throughout a week for one representative participant, offering detailed insights into individual energy expenditure patterns. The heatmap visualizes the average energy expenditure for every 10-minute interval during the majority of active hours. **b,** OpenMetabolics captures the temporal pattern of total energy expenditure shown at 10-minute intervals, highlighting the distinctions between active and inactive portions of the day. The shaded band indicates one standard deviation. **c,** OpenMetabolics captures differences in active energy expenditure across individuals. **d,** Each participant has a unique distribution of physical activity intensity that characterizes their energy expenditure. **e,** People were less active on Sundays than weekdays (two-sided Wilcoxon signed-rank test, n=10, *P<0.05). The boxes show the interquartile range, with a line at the median and a dot and percentage value representing the mean. Circles represent outliers and whiskers extend to 1.5 times the interquartile range.

OpenMetabolics works well for common real-world activities. More data and sensing inputs are needed to accurately estimate energy expenditure for certain activities, such as those with similar thigh motion and different energy expenditure. A data-driven model using leg motion data accurately estimated energy expenditure for conditions like biking at a fixed cadence with varying resistance levels[16] and walking with different body weight loads[43] based on similar but differentiable leg motion measurements. However, adding sensor inputs like electromyography or wearable strain sensors[52] could provide additional information related to muscle activity to improve estimation for more challenging conditions where motion kinematics are similar but muscle recruitment strategies differ. Further validation on upper limb or full-body exercise could enhance the generalizability of OpenMetabolics. This energy expenditure estimation approach could extend to full-body exercises, such as rowing, by adding upper limb sensor data segmented by movement repetitions. Men carry smartphones in their pocket more frequently than women[22], likely due to women's clothing having fewer pockets. Models that incorporate wearable sensor information, such as data from a smartwatch, could improve the amount of activity captured by a smartphone carried in other common locations such as the hand, backpack, jacket, or purse. OpenMetabolics data-driven models could benefit from additional data to represent broader participant demographics to ensure it performs well when monitoring a diverse population. We provide an open-source framework for training these models[53]. The scientific community can contribute datasets to this framework to help build generalized models that include broader participant demographics, sensing modalities, and types of clothing.

OpenMetabolics is an accessible tool that could provide insights into physical activity that impact many fields like clinical research, nutrition, biomechanics, robotics, and public health. OpenMetabolics is an effective physical activity monitor that can be deployed at large-scale and meets the requirements specified by health organizations to overcome the scientific gaps in knowledge relating physical activity to health outcomes[1]. We provide open-source data, code, and a smartphone application to enable this tool to be widely used and extended to support scientific and community use across many different patient populations and health challenges[53]. The accessibility of OpenMetabolics makes it a promising tool to perform physical activity monitoring in areas that are underserved or have limited resources[18,54]. Many scientific fields could leverage this technology: clinical researchers could use findings to inform health policy, public health researchers and urban planners could use the large-scale dataset to design activity-friendly urban environments[55], nutritionists could personalize health interventions, biomechanists could study real-world mobility, and roboticists could develop assistive technology to maximize patient outcomes[56]. OpenMetabolics may provide an effective and accessible activity monitoring tool to address fundamental health questions about physical activity and develop new interventions to help people lead healthier lives.

References

1. Physical Activity Guidelines Advisory Committee. *Physical Activity Guidelines Report*. (2018).
2. Harris, M. A. The relationship between physical inactivity and mental wellbeing: Findings from a gamification-based community-wide physical activity intervention. *Health Psychol Open* **5**, 2055102917753853 (2018).
3. Cox, C. E. Role of Physical Activity for Weight Loss and Weight Maintenance. *Diabetes Spectr.* **30**, 157–160 (2017).
4. Durstine, J. L., Gordon, B., Wang, Z. & Luo, X. Chronic disease and the link to physical activity. *Journal of Sport and Health Science* **2**, 3–11 (2013).
5. Esteban-Cornejo, I., Tejero-Gonzalez, C. M., Sallis, J. F. & Veiga, O. L. Physical activity and cognition in adolescents: A systematic review. *J. Sci. Med. Sport* **18**, 534–539 (2015).
6. Wang, F. & Boros, S. The effect of physical activity on sleep quality: a systematic review. *European Journal of Physiotherapy* **23**, 11–18 (2021).
7. Cook, C. M. & Schoeller, D. A. Physical activity and weight control: conflicting findings: conflicting findings. *Curr. Opin. Clin. Nutr. Metab. Care* **14**, 419–424 (2011).
8. Potter, R., Ellard, D., Rees, K. & Thorogood, M. A systematic review of the effects of physical activity on physical functioning, quality of life and depression in older people with dementia. *Int. J. Geriatr. Psychiatry* **26**, 1000–1011 (2011).
9. Tavares, B. B., Moraes, H., Deslandes, A. C. & Laks, J. Impact of physical exercise on quality of life of older adults with depression or Alzheimer's disease: a systematic review. *Trends Psychiatry Psychother.* **36**, 134–139 (2014).
10. Morrison, J. D. & Mayer, L. Physical activity and cognitive function in adults with multiple sclerosis: an integrative review. *Disabil. Rehabil.* **39**, 1909–1920 (2017).
11. Bull, F. C. *et al.* World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br. J. Sports Med.* **54**, 1451–1462 (2020).
12. Stamatakis, E. *et al.* Association of wearable device-measured vigorous intermittent lifestyle physical activity with mortality. *Nat. Med.* **28**, 2521–2529 (2022).
13. Ahmadi, M. N. *et al.* Brief bouts of device-measured intermittent lifestyle physical activity and its association with major adverse cardiovascular events and mortality in people who do not exercise: a prospective cohort study. *Lancet Public Health* **8**, e800–e810 (2023).
14. Stamatakis, E. *et al.* Vigorous intermittent lifestyle physical activity and cancer incidence among nonexercising adults: The UK Biobank accelerometry study. *JAMA Oncol.* **9**, 1255–1259 (2023).
15. Orendurff, M. S., Schoen, J. A., Bernatz, G. C., Segal, A. D. & Klute, G. K. How humans walk: bout duration, steps per bout, and rest duration. *J. Rehabil. Res. Dev.* **45**, 1077–1089 (2008).
16. Slade, P., Kochenderfer, M. J., Delp, S. L. & Collins, S. H. Sensing leg movement enhances wearable monitoring of energy expenditure. *Nat. Commun.* **12**, 4312 (2021).
17. Hamaya, R. *et al.* Time-vs step-based physical activity metrics for health. *JAMA Intern. Med.* (2024).
18. Althoff, T. *et al.* Large-scale physical activity data reveal worldwide activity inequality. *Nature* **547**, 336–339 (2017).
19. Holdy, K. E. Monitoring energy metabolism with indirect calorimetry: instruments, interpretation, and clinical application. *Nutr. Clin. Pract.* **19**, 447–454 (2004).
20. Westerterp, K. R. Doubly labelled water assessment of energy expenditure: principle,

practice, and promise. *Eur. J. Appl. Physiol.* **117**, 1277–1285 (2017).
21. Reiner, M., Niermann, C., Jekauc, D. & Woll, A. Long-term health benefits of physical activity--a systematic review of longitudinal studies. *BMC Public Health* **13**, 813 (2013).
22. Organization, W. H. *Physical Activity Measurement and Surveillance in Adults: Report of a Scoping and Planning Meeting, 27-28 November 2023*. iv, 34 p. (World Health Organization, Genève, Switzerland, 2024).
23. Troiano, R. P., Stamatakis, E. & Bull, F. C. How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. *Br. J. Sports Med.* **54**, 1468–1473 (2020).
24. Foster, R. C. *et al.* Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Prev. Med.* **41**, 778–783 (2005).
25. Tudor-Locke, C., Sisson, S. B., Collova, T., Lee, S. M. & Swan, P. D. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Can. J. Appl. Physiol.* **30**, 666–676 (2005).
26. Hibbing, P. R., Lamunion, S. R., Kaplan, A. S. & Crouter, S. E. Estimating energy expenditure with ActiGraph GT9X inertial measurement unit. *Med. Sci. Sports Exerc.* **50**, 1093–1102 (2018).
27. Aguilar-Farias, N., Peeters, G. M. E. E. G., Brychta, R. J., Chen, K. Y. & Brown, W. J. Comparing ActiGraph equations for estimating energy expenditure in older adults. *J. Sports Sci.* **37**, 188–195 (2019).
28. Migueles, J. H. *et al.* Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Med.* **47**, 1821–1845 (2017).
29. Bammann, K., Thomson, N. K., Albrecht, B. M., Buchan, D. S. & Easton, C. Generation and validation of ActiGraph GT3X+ accelerometer cut-points for assessing physical activity intensity in older adults. The OUTDOOR ACTIVE validation study. *PLoS One* **16**, e0252615 (2021).
30. Shcherbina, A. *et al.* Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *J Pers Med* **7**, (2017).
31. Lu, K. *et al.* Wearable Cardiorespiratory Monitoring System for Unobtrusive Free-Living Energy Expenditure Tracking. in *World Congress on Medical Physics and Biomedical Engineering 2018* 433–437 (Springer Nature Singapore, 2019).
32. Pope, Z. C., Zeng, N., Li, X., Liu, W. & Gao, Z. Accuracy of Commercially Available Smartwatches in Assessing Energy Expenditure During Rest and Exercise. *Journal for the Measurement of Physical Behaviour* **2**, 73–81 (2019).
33. Lee, C. J. & Lee, J. K. IMU-Based Energy Expenditure Estimation for Various Walking Conditions Using a Hybrid CNN-LSTM Model. *Sensors* **24**, (2024).
34. Yuan, J., Zhang, Y., Liu, S. & Zhu, R. Wearable Leg Movement Monitoring System for High-Precision Real-Time Metabolic Energy Estimation and Motion Recognition. *Research* **6**, 0214 (2023).
35. Lu, K. *et al.* Fusion of Heart Rate, Respiration and Motion Measurements from a Wearable Sensor System to Enhance Energy Expenditure Estimation. *Sensors* **18**, (2018).
36. Website. Global smartphone penetration 2016-2023. Statista https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/.
37. Aguiar, B., Silva, J., Rocha, T., Carneiro, S. & Sousa, I. Monitoring physical activity and

energy expenditure with smartphones. in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* 664–667 (IEEE, 2014).
38. Guidoux, R. *et al.* A smartphone-driven methodology for estimating physical activities and energy expenditure in free living conditions. *J. Biomed. Inform.* **52**, 271–278 (2014).
39. Pande, A. *et al.* Using Smartphone Sensors for Improving Energy Expenditure Estimation. *IEEE J Transl Eng Health Med* **3**, 2700212 (2015).
40. Johnson, M., Turek, J., Dornfeld, C., Drews, J. & Hansen, N. Validity of the Samsung Phone S Health application for assessing steps and energy expenditure during walking and running: Does phone placement matter? *Digit. Health* **2**, 2055207616652747 (2016).
41. Carneiro, S. *et al.* Accelerometer-based methods for energy expenditure using the smartphone. in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings* 151–156 (IEEE, 2015).
42. Piercy, K. L. *et al.* The Physical Activity Guidelines for Americans. *JAMA* **320**, 2020–2028 (2018).
43. Slade, P., Troutman, R., Kochenderfer, M. J., Collins, S. H. & Delp, S. L. Rapid energy expenditure estimation for ankle assisted and inclined loaded walking. *J. Neuroeng. Rehabil.* **16**, 67 (2019).
44. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **7**, 21 (2013).
45. Bassett, D. R., Jr, Wyatt, H. R., Thompson, H., Peters, J. C. & Hill, J. O. Pedometer-measured physical activity and health behaviors in U.S. adults. *Med. Sci. Sports Exerc.* **42**, 1819–1825 (2010).
46. Schmuckler, M. A. What Is Ecological Validity? A Dimensional Analysis. *Infancy* **2**, 419–436 (2001).
47. Brinkerhoff, S. A., Murrah, W. M., Hutchison, Z., Miller, M. & Roper, J. A. Words matter: instructions dictate 'self-selected' walking speed in young adults. *Gait Posture* **95**, 223–226 (2022).
48. Keytel, L. R. *et al.* Prediction of energy expenditure from heart rate monitoring during submaximal exercise. *J. Sports Sci.* **23**, 289–297 (2005).
49. Edemekong, P. F., Bomgaars, D., Sukumaran, S. & Levy, S. B. Activities of Daily Living. (2019).
50. Jayasinghe, U., Hwang, F. & Harwin, W. S. Comparing Loose Clothing-Mounted Sensors with Body-Mounted Sensors in the Analysis of Walking. *Sensors* **22**, (2022).
51. Racette, S. B. *et al.* Influence of weekend lifestyle patterns on body weight. *Obesity* **16**, 1826–1830 (2008).
52. Alvarez, J. T., Gerez, L. F. & Araromi, O. A. Towards soft wearable strain sensors for muscle activity monitoring. *IEEE Trans. Neural Syst. Rehabil. Eng.* (2022).
53. Open-source code and data for this paper. https://github.com/Harvard-Slade-Lab/A-smartphone-activity-monitor-that-accurately-estimates-energy-expenditure.
54. Huhn, S. *et al.* The Impact of Wearable Technologies in Health Research: Scoping Review. *JMIR Mhealth Uhealth* **10**, e34384 (2022).
55. Sallis, J. F. *et al.* Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *Lancet* **387**, 2207–2217 (2016).
56. Slade, P., Kochenderfer, M. J., Delp, S. L. & Collins, S. H. Personalizing exoskeleton assistance while walking in the real world. *Nature* **610**, 277–282 (2022).
57. Brockway, J. M. Derivation of formulae used to calculate energy expenditure in man. *Hum.*

Nutr. Clin. Nutr. **41**, 463–471 (1987).
58. Seethapathi, N. & Srinivasan, M. The metabolic cost of changing walking speeds is significant, implies lower optimal speeds for shorter distances, and increases daily energy estimates. *Biol. Lett.* **11**, 20150486 (2015).
59. Brown, G. L., Seethapathi, N. & Srinivasan, M. A unified energy-optimality criterion predicts human navigation paths and speeds. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
60. Selinger, J. C. & Donelan, J. M. Estimating instantaneous energetic cost during non-steady-state gait. *J. Appl. Physiol.* **117**, 1406–1415 (2014).
61. Blokland, I. J. *et al.* Estimation of Metabolic Energy Expenditure during Short Walking Bouts. *Int. J. Sports Med.* **42**, 1098–1104 (2021).
62. Staacks, S., Hütz, S., Heinke, H. & Stampfer, C. Advanced tools for smartphone-based experiments: phyphox. *Phys. Educ.* **53**, 045009 (2018).
63. White, T. *et al.* Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *Int. J. Obes. (Lond)* **43**, 2333–2342 (2019).
64. Mifflin, M. D. *et al.* A new predictive equation for resting energy expenditure in healthy individuals. *Am. J. Clin. Nutr.* **51**, 241–247 (1990).
65. Camargo, J., Flanagan, W., Csomay-Shanklin, N., Kanwar, B. & Young, A. A Machine Learning Strategy for Locomotion Classification and Parameter Estimation Using Fusion of Wearable Sensors. *IEEE Trans. Biomed. Eng.* **68**, 1569–1578 (2021).
66. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163 (2016).
67. Baroudi, L. *et al.* Estimating Walking Speed in the Wild. *Front Sports Act Living* **2**, 583848 (2020).
68. Schimpl, M. *et al.* Association between walking speed and age in healthy, free-living individuals using mobile accelerometry--a cross-sectional study. *PLoS One* **6**, e23299 (2011).
69. Brisswalter, J., Hausswirth, C., Smith, D., Vercruyssen, F. & Vallier, J. M. Energetically optimal cadence vs. freely-chosen cadence during cycling: effect of exercise duration. *Int. J. Sports Med.* **21**, 60–64 (2000).

Acknowledgements: We thank A. Chin and S. Benabou for assistance with data collection; S. Delp, D. Ba, R. Howe, and F. Calmon for feedback on the study; S. Collins and S. Slade for editorial suggestions.

Funding: Harvard Dean's Competitive Fund for Promising Scholarship and the Raj Bhattacharyya and Samantha Heller Assistive Technology Initiative Fund.

Author contributions: H.C. contributed to conceptualization, methodology, investigation, visualization and writing. P.S. contributed to conceptualization, methodology, visualization, and writing.

Competing interests: The authors declare no competing interests.

Data availability: All materials necessary to replicate the results from this work are available in a public repository: https://github.com/Harvard-Slade-Lab/A-smartphone-activity-monitor-that-accurately-estimates-energy-expenditure. This includes the experimental data, comprising the training dataset of 36 participants used for training the data-driven energy expenditure estimation model and validation dataset for real-world walking experiments using OpenMetabolics and other activity monitors with a diverse population.

Code availability: All custom code necessary to replicate the results from this work is available in a public repository: https://github.com/Harvard-Slade-Lab/A-smartphone-activity-monitor-that-accurately-estimates-energy-expenditure. This includes code to train the data-driven energy estimation model in OpenMetabolics, validate OpenMetabolics and other activity monitors during real-world walking experiments, and replicate the entire OpenMetabolics process.

Methods

Experimental design

The research objective was to develop an accurate tool for large-scale monitoring of energy expenditure during physical activity. To achieve this, we proposed a smartphone-based approach that can be carried in everyday pants and continuously monitor energy expenditure across various timescales. We hypothesized that a data-driven method using leg motion captured by a smartphone would have lower error than existing physical activity monitors in estimating energy expenditure during real-world physical activity. We conducted a power analysis based on a previous study[16] that compared the estimation performance between a Wearable system using leg motion from multiple sensors on the leg and a smartwatch, and found that a sample size of 26 participants would be sufficient to see statistical significance. This analysis used a power of 0.8 with a significance level of 0.05, the difference in mean absolute error in cumulative energy expenditure between the Wearable system (12%) and the smartwatch (40%), and the standard deviation of the Wearable system (25%) from the same study. All experiments and study protocols were approved by the Institutional Review Board of Harvard University (IRB22-1561) and Stanford University (IRB-17282). We collected data from 30 participants for the real-world validation experiments, 10 participants for activities of daily living and stair and incline walk, 10 participants for the pocket motion artifacts correction experiment, and 10 participants for the week-long, at-home monitoring experiment. For the real-world validation experiments, we tested 4 more participants than the minimum number determined by the power analysis in case of any missing data during later analysis. All participants were volunteers and provided written informed consent before study procedures. Participants provided consent for the publication of their identifiable images. The experiments included human participant testing in laboratory, outdoor settings, and at home in free-living conditions. The subsequent sections cover the details of each experiment.

Measuring the ground-truth cumulative energy expenditure of physical activities

The ground-truth metabolic cost of physical activities was calculated using measurements from respirometry equipment. This equipment measured the volume of carbon dioxide and oxygen exchanged with each breath. The standard Brockway equation was applied to convert gas exchange data into metabolic cost in Watts for each breath[57]. Participants carried portable respirometry equipment (K5, COSMED) in the form of a small backpack to collect respirometry measurements during real-world validation experiments (Fig. 2a). Participants abstained from all food and drink except for water for at least 3 hours prior to the study visit to avoid confounds from the thermal effect of food. At the beginning of each session, a 6-minute quiet standing condition was recorded as the baseline metabolic cost. Cumulative metabolic cost was calculated by summing the metabolic cost during the condition[58], including any metabolic cost above the baseline metabolic cost of quiet standing for 3 minutes following the condition[16,59]. Excess oxygen consumption and carbon dioxide production during the return to steady state in quiet standing reflect delays between immediate energy use by muscles and the measured expired gas that arise due to mitochondrial, transport and respiratory dynamics[60]. Accounting for the residual metabolic cost after physical activity enables more accurate measurement of cumulative energy expended during non-steady activities[61]. The cumulative energy expenditure was computed as the cumulative metabolic cost divided by the total time period of activity.

Evaluation metrics

We evaluated the performance of the physical activity monitors based on two error metrics: absolute error, and relative difference.

The absolute error quantifies the discrepancy between the estimated and ground-truth energy expenditure, calculating the percentage error between the estimate and the ground truth and taking the absolute value of these errors.

The relative difference measures the percentage difference between estimated values and the ground-truth. We calculated the difference between the estimated values and the ground-truth, divided by the ground-truth, and multiplied by 100 to obtain the relative change percentage.

The cumulative energy expenditure of the physical activity monitors was calculated by summing all the estimates, multiplied by the time difference between consecutive estimates, divided by the total time period of activity. The error of cumulative energy expenditure was reported with absolute error.

Existing physical activity monitoring tools for comparison

We compared OpenMetabolics to existing methods including smartwatch, pedometer, heart rate model and thigh-based accelerometer during real-world walking experiments.

The smartwatch estimated energy expenditure every minute using an undisclosed model provided by Fitbit. The smartwatch used was a Fitbit Charge 4 with model number FB417 and software version 48.20001.100.76. Prior to data collection, subject-specific information was entered into the Fitbit app and the smartwatch was worn snugly on the participant's left wrist. The energy expenditure estimates from the smartwatch were exported from the Fitbit app. The estimates were in units of kilocalories per minute and converted into Watts by applying a scaling factor. The converted energy expenditure estimates were then interpolated at 5-second intervals.

The pedometer estimates energy expenditure by using a regression equation based on cadence at every gait cycle. Data was collected using a smartphone strapped to participants' thigh (Fig. 2a). The gait cycle was segmented using the detected peak of sagittal plane angular velocity to compute cadence. The smartphone was iPhone 12 Pro with a model number of MGK13LL/A and software version 17.0.3. We used the Phyphox smartphone application to collect angular velocity of leg motion[62]. A regression equation converts each cadence into Metabolic Equivalent values[25] and energy expenditure is then calculated by multiplying participant's body weight.

The heart rate model estimates energy expenditure every 3 seconds using a regression equation that incorporates heart rate along with subject-specific data such as gender, age, height, and weight as parameters[48]. Heart rate data was collected using a smartwatch with activity-specific modes as it provided more frequent measurements of heart rate than without activity selection. The smartwatch was an Apple watch series 1 (42 mm) with model number A1154 and software version 4.3.2. The device was worn snugly on the right wrist. The heart rate data was exported from the Apple Health app.

The thigh-based accelerometer estimates Metabolic Equivalent values every 5 seconds using a quadratic regression equation that incorporates thigh acceleration intensity[63]. The thigh acceleration intensity was calculated by segmenting the three-axis acceleration data into 5-second windows, calculating the euclidean norm of the windows, and then filtering using a fourth-order high-pass filter with 0.2 Hz cutoff frequency. The average of the euclidean norms computed a single value representing the thigh acceleration intensity for each window of acceleration data. The quadratic regression equation then converted each thigh acceleration intensity into Metabolic Equivalent values, and energy expenditure then calculated by multiplying participant's body weight. The acceleration data was collected using a smartphone strapped to participants' thigh

(Fig. 2a). The smartphone was iPhone 12 Pro with a model number of MGK13LL/A and software version 17.0.3. We used the Phyphox smartphone application to collect linear acceleration of leg motion[62].

Data-driven energy expenditure estimation

A data-driven regression model estimated energy expenditure using angular velocity measurements of leg motion during real-world physical activities. The dataset used to train this model was collected by having participants perform various physical activities while wearing a respirometry system and a portable inertial measurement unit on the thigh during laboratory-based experiments. The following paragraphs detail how we pre-process the data and train the data-driven regression model to estimate energy expenditure.The pre-processing steps include aligning to the desired frame of reference, segmenting leg movements by gait cycle, eliminating motion artifacts, and combining features to create an input for the data-driven energy expenditure model.

The data-driven regression model was trained to estimate energy expenditure using thigh angular velocity and steady-state metabolic data from a previous study[16]. The training dataset consisted of 36 participants performing four common activities including walking, running, stair climbing and stationary biking at various intensities (Supplementary Table 1), resulting in 265 unique conditions and 13300 gait cycles. We defined thigh orientation as a fixed coordinate system aligned with the superior-inferior and medial-lateral axes of the thigh segment (Fig. 1b). The inertial measurement unit on the thigh was aligned with the thigh orientation and rigidly attached using a Velcro strap, cohesive bandage and medical tapes. Each gait cycle data consisted of three time series signals from each axis of angular velocity data of the thigh motion, paired with the corresponding steady-state energy expenditure value. We trained an ensemble data-driven model, gradient boosted trees[44], to estimate the steady-state energy expenditure using the angular velocity data from a single gait cycle. The model hyperparameters were tuned using 5-fold cross validation with the negative mean squared error as the loss function. These hyperparameters included the number of trees in the ensemble, the maximum depth of each tree, the learning rate, and the fraction of samples and features to be used for each tree. We tuned the hyperparameters with the RandomizedSearchCV function from the scikit-learn library. The RandomizedSearchCV function randomly sampled 10 sets of hyperparameters from a uniform distribution. We selected the set of hyperparameters with the lowest loss value during the cross-validation.

Understanding which input features contributed to the model's predictions is helpful for interpreting the learned relationship between leg motion and energy expenditure. We calculated the relative importance of model features in percentage (Supplementary Fig. 2). The feature importance was calculated based on how frequently each model feature was used in splits across all trees in the model, indicating the feature was informative for distinguishing the energy expended during physical activity. The most important features were related to the sagittal plane angular velocity (ω_z), accounting for nearly 70% of the importance of all features. This aligns with the biomechanics and muscle activity of leg motion, as sagittal plane angular velocity primarily represents the hip, knee, and ankle joint flexion and extension during walking, running, stair climbing and stationary biking. The subject-specific information was not dominant compared to the angular velocity of leg motion, accounting for 3% of the total importance. This helps to understand key aspects of leg motion data for predicting energy expenditure and gain some intuition for future model refinement and data collection. The model training and feature importance visualization were performed using python (version 3.10.8) with packages xgboost (version 1.7.6) and scikit-learn (version 1.1.2).

We developed a bout detection algorithm to estimate energy expenditure only when motion was detected (Supplementary Fig. 1a). Running the energy expenditure estimation algorithm only when motion is detected can reduce the excess battery usage and data storage on the smartphone. We used a sliding window to continuously monitor the smartphone's angular velocity data. The data was stored only if the average second norm within the window exceeded an angular velocity threshold. We experimentally determined the sliding window length of 4 seconds and the angular velocity threshold of 0.5 rad/s. The stored data was then used to pre-process the angular velocity data into an input for the energy expenditure estimation model. When motion was not detected, the data-driven model identified quiet standing and estimated scaled basal energy expenditure. Basal energy expenditure was calculated using a previous equation[64] based on height, weight, age, and gender, and then multiplied by a scaling factor from a prior study[16].

The data-driven energy expenditure model requires an input aligned with thigh orientation, necessitating a calibration step for smartphone sensor data in the pocket. We developed an orientation alignment algorithm that finds two rotation matrices to match the smartphone sensor data in the pocket to the thigh orientation used for model training (Fig. 1b). These rotation matrices determined the thigh orientation: one for superior-inferior axis alignment and the other for mediolateral axis alignment (Supplementary Fig. 1b). The orientation alignment algorithm first computed a rotation matrix that aligned the smartphone's positive y-axis with the superior-inferior axis of the thigh segment, involving a rotation around the smartphone's z-axis. The algorithm computed a z-axis rotation angle within ±180 degrees that made the average of the smartphone's y-axis acceleration during the detected bout close to the acceleration of gravity. If no angle was detected, the algorithm assigned an identity matrix, assuming the current location of the smartphone in the pocket aligns with the superior-inferior axis of the thigh segment. This rotation matrix was used to correct the smartphone's angular velocity data, before correcting the mediolateral axis. Next, we computed a rotation matrix to align the smartphone's z-axis with the mediolateral axis of the thigh by rotating around the smartphone's y-axis. The algorithm computed a y-axis rotation angle within ±180 degrees that maximized the approximate sagittal plane angular velocity during leg motion. Together, these two rotation matrices calibrated the smartphone frame of reference in the pocket to the thigh orientation (Supplementary Fig. 1c).

The calibrated smartphone angular velocity data were filtered and segmented by gait cycle using the sagittal plane angular velocity (Fig. 1c). We used a fourth-order 6 Hz low-pass filter to remove high-frequency motion artifacts from the calibrated smartphone angular velocity data. The filtered sagittal plane angular velocity was used to segment the movement into gait cycles by detecting two consecutive peaks. The angular velocity peaks needed to cross a threshold of 70 degrees per second and be spaced at least 0.6 seconds apart to avoid errant motions that occur faster than the shortest gait duration for our activities, including running or biking.

The three axes of segmented velocity were discretized into 30 bins each and then concatenated into a single input vector of 90 values (Fig. 1d). All hyperparameter values for filtering, segmenting, and discretization during this process were selected from previous experiments[16].

The segmented angular velocity was used as the input for the pocket motion correction model to estimate and remove pocket motion artifacts (Fig. 1e). The pocket motion correction model used the segmented angular velocity and corresponding gait duration as a model input with a size of 91 by 1. The motion artifacts were subtracted from the uncorrected smartphone angular velocity to get the corrected smartphone angular velocity. Details regarding the pocket motion artifact correction model development, training, and experiment will be discussed in 'Pocket motion artifact correction experiment'.

The data-driven energy expenditure model used input data of the corrected smartphone angular velocity and additional statistical and subject-specific features to estimate energy expenditure once per step. Statistical features could capture the characteristics of data in a more compact form and potentially enhance model performance[65]. We computed five statistical features from each axis of the corrected smartphone angular velocity data. These statistical features included the mean, standard deviation, median, skewness and second norm. Subject-specific data, such as body weight and height, were included in the energy expenditure estimation model to relate anthropometric measurements to muscle activity. Finally, all these features were combined and flattened into a single vector array of 108 values and used to estimate energy expenditure once per gait cycle (Fig. 1g).

Evaluating OpenMetabolics during real-world physical activities

We performed real-world physical activity experiments to compare OpenMetabolics to existing activity monitoring tools. Participants (*n*=30, 17 men 13 women; age, 34 ± 14 yr; body mass, 71.4 ± 13.0 kg; height,1.72 ± 0.08 m; body mass index, 24.1 ± 3.5) completed a one day experimental protocol. We recruited participants whose average age and body mass index were within half a standard deviation of those in a previous study[30] that validated the performance of smartwatches across diverse cohorts (Supplementary Fig. 3). Real-world physical activity data were collected from 11 participants using a portable inertial measurement unit (Adafruit Precision NXP Breakout Boards) strapped to the thigh and from 19 participants using a smartphone strapped to the thigh, all under the same testing conditions. We aggregated these datasets for the analysis of OpenMetabolics' performance, as they demonstrated excellent reliability between the two measurements based on the intraclass correlation coefficient of 0.97[66] (Supplementary Fig. 10). Participants were asked to complete four different real-world physical activities, including naturalistic walking at various speed and bouts for 16 minutes, running at self-selected speed for 6 minutes, and stationary biking at fixed cadence for 6 minutes. Participants took a 3 minute rest of quiet standing between each activity. All activities were performed outdoors along a path consisting of concrete and brick public sidewalks (Fig. 2b). Participants could skip any conditions they found too strenuous to complete to make the protocol feasible for participants with diverse physiological characteristics. Two subjects were excluded due to the connection issues with the portable respirometry system and no other exclusions were made.

We emulated naturalistic walking by providing a series of audio cues with participants to start and stop their walking bouts. We tried to match two naturalistic walking factors: bout duration and walking speed. The bout durations were randomly sampled from a predetermined distribution (Fig. 2e) from a previous real-world study characterizing bout duration[15]. Participants rested in a quiet standing position for a randomized duration of 5 to 10 seconds between each bout. We mimicked the naturalistic range of walking speed by providing audio prompts that encouraged participants to self-select their walking speed, such as "Walk as if you were walking across the street" or "Walk as if you were walking through a field". These audio prompts were associated with different self-selected walking speeds[47], and we used them as ecologically relevant guidance. We randomly sampled from a preselected distribution of speed (Fig. 2d) that mimicked the walking speed distribution of free-living environments from a previous study[67].

Participants performed real-world physical activities while wearing a portable respirometry system, a smartwatch, a smartphone strapped to their thigh, and a smartwatch for walking speed collection. For the real-world walking experiment, participants walked continuously along a 650-m public sidewalk, climbing up two flights of stairs each time they passed the starting point (Fig. 2c). Participants performed 20 bouts of walking while following audio prompts provided via wired earphones, which were connected to a Raspberry Pi 3b+ running a predetermined python script. The walking speed of participants was collected from the same smartwatch used for the heart rate model. The observed walking speed was adjusted to account for the age difference[68] between the individuals from the previous study[67] and our target population (Fig. 2d). The total walking time was around 16 minutes with an average number of steps of 1650 across all participants.

Participants performed self-selected pace overground running for 6 minutes and 6 minutes of stationary biking at a steady state pace. We used a stationary bike, Wattbike Atom (Wattbike Ltd., Nottingham, UK), which has adjustable resistance and provides real-time measured cadence displayed on the screen. Participants selected a comfortable resistance and biked at a pedal rate of 83 revolutions per minute, which is a common freely chosen cadence[69]. We performed additional real-world walking with naturalistic environmental changes such as stair walk and incline walk. Participants (5 men and 5 women; age, 27 ± 3 years; body mass, 68.3 ± 9.3 kg; height, 1.73 ± 0.07 m; body mass index, 22.8 ± 2.2) completed a one-day experimental protocol. Participants climbed up and down two flights of stairs at self-selected pace for 6 minutes. Participants performed an incline walk on a treadmill set to a speed of 1.0m/s and an incline of 5 degrees. We used an incline treadmill (SOLE TT8 Treadmill, Salt Lake City, USA), which has adjustable velocity and incline settings.

Evaluating OpenMetabolics during activities of daily living

We performed naturalistic activities of daily living experiments to compare OpenMetabolics to existing activity monitoring tools. Participants (n=10, 5 men and 5 women; age, 27 ± 3 years; body mass, 68.3 ± 9.3 kg; height, 1.73 ± 0.07 m; body mass index, 22.8 ± 2.2) completed a one-day experimental protocol. Four different basic and instrumental activities essential for meeting basic physical needs and living independently were selected from previous research[49], including organizing the home, cleaning, eating, and personal hygiene. The video examples of activities of daily living highlights a participant performing these different activities to help with the understanding of the tasks (Supplementary Video 1). The organizing home activity involved managing and arranging living spaces, such as organizing clothes. The cleaning home activity included maintaining the living space clean and tidy such as sweeping the floor. The eating activity required participants to feed themselves while sitting. The personal hygiene activity involved the ability to bathe, groom oneself, and maintain dental hygiene. Participants performed these four activities of daily living in a randomized order while wearing a portable respirometry system, a smartwatch, a smartphone strapped to their thigh, and a wrist-worn heart rate monitor. Each activity lasted 6 minutes and was performed at a self-selected pace.

Pocket motion artifact correction experiment

We conducted naturalistic walking experiments in an indoor laboratory to investigate the effect of pocket motion artifacts on a range of clothes at varying speeds of walking bouts. Healthy young adults (*n*=10, 5 men 5 women; age, 26 ± 3 yr; body mass, 71.7 ± 9.1 kg; height,1.72 ± 0.08 m; body mass index, 24.1 ± 1.3) completed a one-day experimental protocol. We evaluated various types of clothes: jeans, regular shorts, sweatpants, and athletic shorts. Participants brought their own clothes to best emulate real-world use. Participants wore a smartphone strapped to their thigh as the ground-truth smartphone and carried another smartphone in their pocket (Fig. 4a). Both smartphones were iPhone 12 Pro with model number MGK13LL/A and software version 17.0.3. We used the Phyphox app[62] to collect angular velocity data from both smartphones, the same app used during real-world walking experiments (Fig. 2g). Each participant performed five walking bouts, each lasting 20 seconds with a 5-second pause between bouts. We provided participants with five specific audio prompts to create naturalistic variations with the widest range of speeds. The list of ecologically relevant[46] audio prompts[47] from slow to fast is as follows: "Walk as if you were walking home after a really bad day.", "Walk as if you were walking through a field.", "Walk across the walkway at your typical speed.", "Walk as if you were walking across the street.", and "Walk as if you were walking to catch a bus.". The order of the audio prompts was randomized to avoid any experimental bias.

A data-driven model was trained to eliminate motion artifacts from the pocket to isolate the smartphone's angular velocity of the underlying leg motion. We defined the angular velocity from the smartphone in the pocket as uncorrected smartphone data and from the smartphone strapped to the thigh as ground-truth smartphone data (Fig. 3a). The motion artifact was defined as the angular velocity difference over three axes between the uncorrected smartphone data and the ground-truth smartphone data. A linear regression model was trained as a pocket motion correction model using one gait cycle of the uncorrected smartphone data. The model input consisted of discretized angular velocity over three axes from one gait cycle with a size of 90 by 1 and the corresponding gait duration in second. The final input size was 91 by 1. The model output the estimated motion artifact with a size of 90 by 1, which was subtracted from the uncorrected smartphone data to obtain corrected smartphone data of the same size. We evaluated the root mean square error against the ground-truth smartphone data and the relative change in energy expenditure estimation with and without the pocket motion correction model (Fig. 3b-c). We used 10-fold cross-validation, holding out one subject's data for evaluation and averaging the results across all ten folds.

Personalizing pocket motion artifact correction model

We collected walking data from one pilot participant (n=1) to evaluate a process for personalizing the pocket motion artifact correction model that only requires smartphones. The participant completed two real-world walking sessions, each consisting of five different bouts for 10-seconds each, following the same audio prompts used in the pocket motion artifact correction experiment. In the first session, one smartphone was held against the thigh to collect ground-truth motion data while another smartphone was in a pocket as the participant performed real-world walking. This data was used to re-train the pocket motion correction model (Fig. 3a-c). In the second session, one smartphone was strapped to the thigh and the other smartphone was in a pocket as the participant performed real-world walking to collect evaluation data. We evaluated the motion artifact RMSE for three approaches: using uncorrected smartphone data, corrected smartphone data with the pocket motion correction model, and corrected smartphone data with the personalized pocket motion correction model (Supplementary Fig. 9). Due to the small sample size (n=1), no statistical tests were conducted and numerical results were not included in the main text.

Emulating smartphone data with pocket motion artifacts during real-world walking experiments

We emulated pocket motion artifacts during real-world evaluation experiments to understand the potential estimation performance of OpenMetabolics. Conducting respirometry experiments with a diverse cohort wearing many different types of clothing and performing naturalistic walking bouts would require hundreds of hours of data collection. We sampled motion artifacts from the pocket motion artifact experiment and matched them to the real-world walking experiments based on gait speed (Fig. 2g). We selected a group of motion artifacts within a 10% speed difference for each gait cycle data from the real-world walking experiments (Fig. 3d). We then added one of these speed-matched motion artifacts to the gait cycle data to emulate real-world walking data with motion artifacts. A motion artifact was randomly sampled from walking speeds within 10% of the real-world walking data to provide a range of similar pocket motion artifacts for rigorous evaluation. If no artifacts were found within the 10% speed difference range, we added the closest matching motion artifact to the real-world walking data. Adding appropriate motion artifacts based on walking speed prevents systematic bias and closely matches the motion artifacts produced during real-world motion. Once we emulated real-world walking data with motion artifacts, we processed and estimated energy expenditure using the pocket motion correction model, following the steps shown in Fig. 1. The pocket motion correction model was trained on a dataset separate from the test dataset of motion artifacts used for emulating real-world walking data, ensuring rigorous evaluation. This procedure was repeated 10 times using a 10-fold cross validation dataset, holding out one subject's data with 4 types of clothing for sampling motion artifacts and using the rest for training the pocket motion correction model. The entire emulation process was conducted for 28 participants from real-world walking experiments, multiplying by 40 clothes creates 1120 possible combinations. Finally, we averaged the estimated absolute error across all folds to determine the expected accuracy of OpenMetabolics when the smartphone was carried in a pocket during real-world walking scenarios (Fig. 3e).

<u>Monitoring real-world physical activity for a week</u>

We monitored participants' physical activity for a week to showcase the potential use case of OpenMetabolics. We developed a customized application, OpenMetabolics, to continuously collect smartphone inertial measurement unit data at 50 Hz, including acceleration and angular velocity. This application runs in the background even when the smartphone is locked, enabling continuous data collection without interruption. The development environment was Android Studio Hedgehog (version 2023.1.1) with OpenJDK 64-Bit Server VM (version 17.0.7). The application was installed on Galaxy A24 smartphones (model number of SM-A245M/DSN and Android version 13) and deployed to participants. Participants (n=10, 5 men 5 women; age, 26 ± 4 yr; body mass, 66.5 ± 7.1 kg; height,1.71 ± 0.07 m; body mass index, 22.7 ± 2.0) completed 7 days of an activity monitoring protocol. Participants were instructed to carry the smartphone in their pocket continuously for 7 days, except during sleep, swimming, and showering. Participants did not interact with the smartphone while the data collection was ongoing, ensuring continuous data collection of their leg motion during the protocol. The majority of observed walking bouts were less than 30 seconds, closely matching results from the previous real-world activity monitoring study[15] (Supplementary Fig. 11).

<u>Statistical analysis</u>

All statistical tests were implemented using python (3.10.8) and additional python libraries including scipy (1.10.1) and statsmodel (0.14.1). All data were assessed for normality using the Shaprio-Wilk test. Non-parametric tests were used if the normality test rejected the null hypothesis. We used Kruskal-Wallis test for comparisons involving more than two groups and two-sided Wilcoxon signed-rank test for pairwise comparisons. For multiple comparisons tests, p-values were adjusted to account for the number of comparisons of interest using Bonferroni correction to avoid the false positive error rate inflation. A post-hoc multiple comparison test was performed after assessing the main effect using the Kruskal-Wallis test. A Linear mixed model was used to investigate the effect of subject-specific information on the cumulative energy expenditure error of OpenMetabolics during the real-world validation experiments. There were no significant effects of the subject's age, gender and body mass index on the cumulative energy expenditure error, with p-values of 0.365, 0.099, and 0.276, respectively (Supplementary Table 2). All statistical tests were set to a significance level of 0.05, with significance reported below this threshold.