Preprint – Not peer reviewed

Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits: Protocol for a Systematic Review and Meta-Analyses

Guillaume Chevance[1,2], Natalie M. Golaszewski[1], Elizabeth Tipton[3], Eric B. Hekler[1], Matthew Buman[4], Gregory Welk[5], Kevin Patrick[1], Job G. Godino[1,2]

[1]Center for Wireless & Population Health Systems, University of California, San Diego, San Diego, United States;
[2]Exercise and Physical Activity Resource Center, University of California, San Diego, San Diego, United States;
[3]Department of Human Development, Teachers College, Columbia University, New York, United States;
[4]School of Nutrition & Health Promotion, Arizona State University, Phoenix, AZ, United States;
[5]Department of Kinesiology, Iowa State University, Ames, IA 50011, United States

Corresponding authors: Job Godino
University of California San Diego, Qualcomm Institute, Atkinson Hall,
9500 Gilman Drive,
La Jolla, CA 92093, USA

Email: jgodino@eng.ucsd.edu

Compliance with Ethical Standards: All authors declare that they have no competing interests with the present research; This article does not contain any studies with human participants performed by any of the authors, informed consent was therefore not required.

Author agreement statement: We the authors agree to the sharing of this pre-print on SportRχiv.

Pre-print DOI: 10.31236/osf.io/7mcdw

Twitter handles: @GuillChevance

Citations:

APA
Chevance, G., Golaszewski, N., Tipton, E., Hekler, E. B., Buman, M., Welk, G., Patrick, K., & Godino, J. (2019, December 5). Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits: Protocol for a Systematic Review and Meta-Analyses. https://doi.org/10.31236/osf.io/tgk7a

MLA
Chevance, Guillaume, et al. “Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits: Protocol for a Systematic Review and Meta-Analyses.” SportRxiv, 5 Dec. 2019. Web.

Chicago
Chevance, Guillaume, Natalie Golaszewski, Elizabeth Tipton, Eric B. Hekler, Matthew Buman, Gregory Welk, Kevin Patrick, and Job Godino. 2019. “Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits: Protocol for a Systematic Review and Meta-Analyses.” SportRxiv. December 5. doi:10.31236/osf.io/tgk7a.

Introduction

It is widely recognized that physical activity is an important determinant of health (1–3). However, there is considerable variation in the observed associations between physical activity and a variety of health outcomes (4,5). Although some of this heterogeneity may result from true biological mechanisms, the challenge of accurately assessing this complex behavior is likely a major contributor. In order to determine temporal trends and further characterize the dose- response relationship with various health outcomes, objective and valid measures of physical activity remain necessary (6,7).

Tools for the objective assessment of the frequency, intensity, and duration of physical activity in adults and children have largely been developed for short-term use within research or public health surveillance environments (8,9). However, recent advances in microtechnology, data processing, wireless communication, and battery capacity have resulted in the proliferation of low-cost, non-invasive, wrist-worn devices with attractive designs that can easily be used by consumers to track their physical activity over long periods of time (10). In the past years (2017-2018), ~102.4 millions of such devices were shipped worldwide, and this number is predicted to grow to 237.5 million by the year 2021 (11). The majority of devices contain Bluetooth connectivity that allows users to transmit data at varying resolutions to the cloud via mobile, web, or computer applications. Corresponding application programing interfaces (APIs) and standardized OAuth procedures (i.e., an authorization procedure that allows data to be

passed between apps via the Internet) enable researchers to gain access to the data stored by manufacturers.

Although it is impossible to directly measure physical activity, the latest generation of consumer-level activity monitors combine multi-sensor devices using *triaxial accelerometry* to measure movement (i.e., gravitational acceleration in the anterior-posterior [x], cranial-caudal [y], and medial-lateral [z] planes) and *photoplethysmography* to measure heart rate (i.e., number of beats per minute). Importantly, a combined sensing approach may address many of the limitations of using either accelerometry or photoplethysmography alone (12,13). For example, heart rate monitors can accurately assess forms of physical activity that cannot be effectively captured by accelerometers (e.g., cycling on a stationary bicycle), whereas accelerometers can accurately assess low intensity physical activity that are not effectively captured by heart rate (e.g., slow walking, (14)). The combination of these data streams through branched equation modeling and/or machine learned algorithms may result in a more accurate assessment of physical activity (14,15) and allows for the assessment of newly developed metrics such as cardiorespiratory fitness and resting heart rate.

The expanding use of consumer-level activity monitors in population and clinical health research has led to an array of independent studies aimed at evaluating the validity of the various metrics. No devices have received more attention than those manufactured by Fitbit (Fitbit Inc, San Francisco, CA, USA). From community-based health interventions that aim to motivate individuals to increase their physical activity level, to interventions that aim to improve patient-health professional interactions

(16,17), Fitbits are likely the most widely used. Illustrating this, approximately 200 clinical trials registered at ClinicalTrials.gov over the past 3 years (2017-2019) report using measures from Fitbits as an outcome (18). Hence, a major concern for consumers and researchers alike is whether or not Fitbits provide accurate estimates of physical activity behavior.

Several studies have evaluated the validity of different versions of Fitbits to estimate energy expenditure, intensity, heart rate, and/or steps, mostly in controlled laboratory settings [e.g., (19)] and sometimes in free-living conditions [e.g.,(20)]. Moreover, there have been three systematic reviews that have been conducted to examine the accuracy of measures derived from consumer-level activity monitors in general (21,22) and from Fitbits specifically (18). Taken together, these reviews concluded that Fitbit devices accurately measure steps, while estimates of energy expenditure are less than optimal (to our knowledge, no synthesis is available for the estimates of heart rate). Although previous systematic reviews have been informative, several limitations exist within these reviews. First, two of the three systematic reviews (18,22) have included studies with very small sample sizes or compared Fitbits to questionable criterion measures, such as other wearable devices (i.e., accelerometers), instead of ground-truth or “gold-standard” measures of energy expenditure (23), heart rate (24) or steps (25) [e.g., indirect calorimetry, electrocardiogram and steps count video recorded]. Second, all previous reviews have included older versions of the Fitbit that do not use photoplethysmography combined with accelerometry, which are likely to result in more bias than more recent Fitbits (21). Third, there has yet to be a quantitative synthesis of the validity of Fitbits through a meta-analysis.

The purpose of the proposed systematic-review and meta-analysis is to systematically examine, quantify and report the validity of energy expenditure, intensity, heart rate, and steps measured by combined-sensing Fitbits.

Methods

A systematic review of the literature will be conducted to retrieve “validation studies” reporting a measure of energy expenditure, heart rate, or steps, and a criterion, “gold-standard”, measure of these parameters (see the inclusion criteria below). After the systematic review, a series of meta-analyses for each outcome will be conducted. Summary effect sizes for the accuracy of Fitbits compared to the criterion measures will be calculated from the effect sizes of all the individual studies included in the systematic review.

This protocol is registered with PROSPERO (ID 161937) and is report here according to the PRISMA-P guidelines [(26); see supplemental material]. All the study materials and data will be provided in supplemental materials on SportRxiv page (see https://osf.io/preprints/sportrxiv/).

Search strategy

Consistent with best practices (27), a systematic review of the literature will be conducted in three iterations to retrieve both published and unpublished studies. The search will be conducted using databases PubMed and EMBASE from January 2015

(i.e., commercialization of the first Fitbit device including a heart rate monitor) and the date in which the search will be performed. The grey literature will also be inspected through Open Grey. A first iteration will be performed with the following combination of terms: ‘Fitbit’ AND ‘validation’ AND ‘precision’ OR ’comparison’ OR ‘equivalent’ OR ‘agreement’ OR ‘accuracy’ AND ‘heart rate’ OR ‘steps’ OR ‘energy expenditure’ OR ‘exercise’ (see supplemental materials). In a second iteration, studies will be also sourced from previously published systematic reviews (18,21,22). Finally, a third iteration will be performed in which reference lists within studies included in the first iterations will be examined. Published conference abstracts will also be included if sufficient detail is reported to assess study quality. Study selection will be performed by one coder (GC), and check by an independent second coder (NG). Any discrepancies will be identified and resolved. No language restrictions will be applied.

Criteria for study inclusion

Studies that simultaneously report outcome data from a Fitbit device (energy expenditure, heart rate, or steps) and a valid criterion measure will be considered. Only studies that evaluate Fitbit devices including a heart rate monitor (i.e., *Charge HR 2015, Surge 2015, Blaze 2016, Charge 2 2016, Alta HR 2017, Ionic 2017, Versa 2018, Charge 3 2018, Inspire HR 2019, Versa Lite Edition 2019*) will be included. Valid criterion measures of energy expenditure include doubly labeled water, or direct and indirect calorimetry; for heart rate, electrocardiograms (ECG), pulse oximeters and

specific chest worn systems (e.g., Polar) are accepted as criterion measures; for steps, direct observation is the only criterion (video recorded or not).

Importantly, in comparison with previous systematic reviews, these criteria will exclude studies that compare Fitbit devices to other wearable devices such as accelerometers. Only data from adults (age > 18 years old) will be considered. No other exclusion criteria will be set; however, sensitivity and sub-group analyses will be conducted.

Planned sensitivity and sub-group analyses

Studies that will be included in the systematic review are likely to be heterogenous in regards of the characteristics of the participants included, the type of device used to derive parameters, the specific activities and intensities recorded in each individual study, and study quality. Hence, these potential moderator variables will be examined in the meta-analysis in order to account for heterogeneity in the literature.

1. *Characteristics of the participants.* Adults with underlying health conditions may have atypical movement patterns or altered heart rate due to medication (28). These abnormal patterns may reduce accuracy estimation of both energy expenditure, heart rate, and steps and will thus be analyzed. Participants' age will also be considered as movement patterns may change between young and older adults. Depending of the numbers of studies available different age groups will be analyzed independently (e.g., 18-45

years old, 45-65 years old, over 65 years old). Age range will be defined according to the mean age reported in each individual study. If available, Body Mass Index (BMI) and gender will also be controlled for. Finally, if available, skin tone will be retrieved, as darker skin absorbs more green light, which presents a problem because most photoplethysmography use green LED's as light emitters, limiting their ability to accurately measure heart rate through dark skin (29).

2. *Type of Fitbit devices.* Fitbit currently (Dec. 2019) commercializes 10 devices that combine accelerometry and heart rate monitor; however, each device is likely to vary in the estimation of the parameters under scrutiny due to differences in technology and algorithms. Sensitivity analyses will be conducted to compare estimates derived from different Fitbit devices.

3. *Type of activity.* Accuracy of the Fitbit may vary depending of the activity recorded during the protocol. According to a previous meta-analysis (21), potential differences between (*i*) resting and sedentary activities, (*ii*) ambulation, household and walking tasks, (*iii*) cycling, (*iv*) running and (*v*) resistance exercises will be examined. Recovery periods, typically characterized with a resting period at the end of protocol, will be included in the main analysis, but excluded from these sub-group analyses as they appear difficult to categorize and pool together due to protocol differences. If

the study reported descriptive data for multiple activities, a ‘mixed’ category will be created.

4. *Intensity.* Previous reviews have highlighted that accuracy of the Fitbit may vary with the intensity recorded during the protocol (19). Hence, effect sizes recorded for sedentary and light activities will be combined together and examined independently from moderate and vigorous intensity activities. Limits between light and moderate intensity physical activity will be defined according to the compendium of physical activities [(18); e.g., walking above 3 mph/5 km/h and cycling above 7 mph/11 km/h, or 150 W, will be considered as moderate to vigorous physical activity]. Recovery time or cool-down activity after a hard workout, will be classified separately as a recovery period.

5. *Protocol duration.* Studies with longer protocols are likely to be negatively correlated with the accuracy of the devices since more errors could be accumulated over time. The total time of the protocol will thus be controlled for in the analyses.

6. *Study quality.* A custom risk of bias tool will be used to quantify potential biases in effect sizes related to study quality (see below). Studies will be then split according to their quality score.

Quality assessment (risk of bias)

A custom tool, developed based on previous study using the COSMIN criteria (31), will be used to assess study quality including: (*i*) sample size calculation [explicitly reported = 1 point, see (32)], (*ii*) peer reviewing (study peer reviewed = 1 point), (*iii*) quality of the criterion measures (utilization of a direct calorimetry or doubly labeled water to estimate energy expenditure, ECG to estimate heart rate, and video recorded estimation of steps = 1 point), (*iv*) placement of the Fitbit device [device up to three finger widths above the wrist bone = 1 point; see (33)], (*v*) the validation of multiple devices simultaneously (validation study including one device at a time = 1 point), (*vi*) and the description of missing data (percentage of missing data for each analysis described = 1 point). A score comprises between 0 and 6 will be computed for each outcome of interest. This score will then be used to categorized individual studies as 'low', 'moderate' and 'high' risk of bias. A specific extraction form is provided in supplemental materials for the quality score (see supplemental materials).

Sensitivity analyses will be performed for the primary meta-analyses (i.e., average energy expenditure, heart rate, steps) based on risk of bias by removing "high risk of bias" studies from the analyses. Subgroup analyses will also be conducted according to the potential moderators identified previously.

Data extraction and management

Information about the study characteristics (authors, year of publication, design, sample size, number of observations for each outcome), population characteristics (age, health conditions, BMI), descriptive statistics, the type of Fitbit, and features of the criterion measures will be extracted.

A major characteristic of the literature concerning the validation of physical activity wearables devices is the multitude of different statistical strategies and related effect sizes used to estimate the validity of these devices (38). Researchers have used, separately or in combination, analysis of variance (e.g., ANOVA), correlations (e.g., ICC), and measures of agreement (e.g., Bradley-Blackwood test, Bland-Altman analyses, Mean Absolute Percentage Error). This makes the meta-analysis of individual studies challenging. Based on a previous examination of the literature, we decided to use the mean bias and standard deviation from the *Bland Altman analyses*, in combination with the *Mean Absolute Percentage Error* (MAPE). These two methods will mirror the approach in individual studies, and will be directly interpretable instead of traditional effect sizes used in a previous meta-analysis [e.g., Hedges' g, (21)].

- For the Bland-Altman analyses, effect sizes extracted will be the mean bias (i.e., accuracy) and variance or standard deviation (i.e., precision) in kilocalories per minutes (kcals/min), beat per minutes (bpm), and difference of steps per minutes (steps/min) between the Fitbit and criterion measures of energy expenditure, heart rate, and steps respectively. It's important to note

that kcals and steps are not always reported in function of time (i.e., per minutes). Some authors have preferred the total amount of kcals or steps recorded during a specific task or an entire protocol. To make comparisons between studies, and interpretation of the results, possible, we will retrieve the time spend during each protocol tasks. We will then convert the absolute number of kcals and steps in kcals and steps per minute by dividing the mean bias and standard deviation reported by the duration of each specific task in minutes. For example, a mean bias of 20 ±10 kcals recorded over a three minutes tasks will be converted to 6 ±3 kcals/min.

- For the MAPE meta-analysis, MAPE for absolute kcals, steps and kcals/min, steps/min and bpm will be extracted and pooled together as they are expressed in percentage. The potentially different algorithms used to compute the MAPEs will also be extract and controlled for in the analyses (e.g., MAPE = Mean{|Y - X|/X} *versus* Mean{|Y - X|} / Mean(X)).

These outcomes will be extracted directly from eligible studies when available, or compute with other statistics reported (see in supplemental materials). In a third case, authors will be contacted and ask to provide the necessary information (i.e., means and standard deviations, and/or Bland Altman estimates, and/or MAPE estimates).

Data will be extracted and coded from one coder (GC), then check by a second coder (NG). Discrepancies will be identified and resolved by re-referencing the articles. Two data extraction forms are provided in supplemental materials, one for the extraction

of the study characteristics and one for the meta-analyses (see supplemental materials).

Data synthesis and analyses

A specific meta-analytic framework will be used for the -analyses of agreement between measures for both the Bland-Altman and MAPE indicators (34). The main outcome of the Bland-Altman meta-analyses will be the population limits of agreement (LoA) between Fitbit devices and criterion measure of energy expenditure, heart rate, and steps. LoA combines the bias of a test (i.e., the average difference between the tested measure and a criterion measure) and the standard deviation of these differences. The results from the individual studies will be converted into a standard format to conduct meta-analyses, with bias meaning 'criterion measures – Fitbit' measured in kcals/min, bpm and steps/min for energy expenditure, hear rate and steps respectively.

In the present study a 'population LoA' will be computed to account for two sources of variation, the average within-study variation, and the between-study variation. The 'population LoA' is wider (i.e., more 'conservative'/'robust') than those typically reported in other meta-analyses of Bland-Altman studies (34). In the present study the pooled LoAs are calculated using $\delta \pm 2\sqrt{(\sigma^2+\tau^2)}$, where δ is the average bias across studies, σ^2 is the average within-study variation in differences and τ^2 is the variation in bias across studies. Both δ and σ^2 will be estimated using a weighted least-

squares model (similar to a random-effects approach) and their standard errors (SEs) using robust variance estimation (RVE). RVE will be used instead of model-based SEs because we expect that most of the studies included in our review used repeated measures designs without accounting for the correlation between measurements (i.e., multi-level approach). The method-of-moments estimator from (35) will be used for the τ^2 parameter. According to (34), we will also include measures of uncertainty when interpreting the LoA estimates by calculating the outer 95% CIs for pooled LoA; and adjusted repeated measurements, which were not properly adjusted in individual studies. Multiple effect sizes from a same individual study will also be handled with the use of RVE method [see (36,37)]. Finally, this method will be similarly applied to the meta-analysis of MAPEs.

In the absence of clear clinically acceptable agreement between the Fitbit and gold-standard measures of energy expenditure, heart rate, and steps, we didn't specify any a-priori hypotheses. Final interpretation will be done by combining results from the Bland-Altman and MAPE meta-analyses.

All analyses will be conducted in the R statistical program (11). The R code [adapted from (34)] and all data used in the meta-analyses will be made available.

Discussion

Previous systematic reviews have reported mixed results regarding the accuracy and precision of Fitbit devices to estimate energy-expenditure, heart-rate and steps (18,21,22). Comprehensive synthesis of previous validation studies is thus needed,

especially given the increasingly use of Fitbits for public health, medical and health promotion research.

The technology behind the Fitbit is ever-evolving, with 8 new models on the market including a heart rate monitor since 2015. The present systematic review and meta-analysis will thus focus on recent versions of the device that include a heart rate monitor in addition to the tri-axial accelerometer. More importantly, in comparison with previous literature reviews, the present study will exclude validation studies where energy expenditure, heart rate, and steps were estimated with other wearable devices (i.e., accelerometers) known to provide biased estimation of these parameters (21). Finally, to our knowledge, this study will be the first to provide a quantitative synthesis of the accuracy of the Fitbit, through a meta-analysis, for heart rate and steps. The frameworks that will be used for this meta-analysis will provided directly interpretable measure of the accuracy and precision of the most recent Fitbit models.

This up-to-date synthesis will help determine if future research can use the Fitbit as an accurate measure of energy expenditure, heart rate, and steps. Subgroup analyses from the present meta-analysis may also indicate that Fitbit devices are only accurate for the estimation of certain parameters under certain conditions.

References

1. Hallal PC, Bauman AE, Heath GW, Kohl HW, Lee I-M, Pratt M. Physical activity: more of the same is not enough. Lancet Lond Engl. 2012 Jul 21;380(9838):190–1.

2. Luyster FS, Strollo PJ, Zee PC, Walsh JK. Sleep: A Health Imperative. Sleep. 2012 Jun 1;35(6):727–34.

3. Ross R, Blair SN, Arena R, Church TS, Després J-P, Franklin BA, et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. Circulation. 2016 13;134(24):e653–99.

4. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet Lond Engl. 2012 Jul 21;380(9838):247–57.

5. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: A systematic review and meta-analysis of prospective studies. Sleep J Sleep Sleep Disord Res. 2010;33(5):585–92.

6. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? Int J Obes Relat Metab Disord J Int Assoc Study Obes. 1998 Aug;22 Suppl 2:S30-38.

7. Lynch JW, Smith GD. A life course approach to chronic disease epidemiology. Annu Rev Public Health. 2005;26:1–35.

8. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014 Jul;48(13):1019–23.

9. Van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res. 2011 Mar;20(1 Pt 2):183–200.

10. Dobkin BH, Dorsch A. The Promise of mHealth: Daily Activity Monitoring and Outcome Assessments by Wearable Sensors. Neurorehabil Neural Repair. 2011;25(9):788–98.

11. Ubrani J, Llamas R, Shirer M. Wristwear Dominates the Wearables Market While Clothing and Earwear Have Market-Beating Growth by 2021, According to IDC. International Data Corporation. https://www.idc.com/getdoc.jsp?containerId=prUS42371617. Published 2017. Accessed January 29, 2018.

12. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of Free-Living Energy Expenditure by Heart Rate and Movement Sensing: A Doubly-Labelled Water Study. PLOS ONE. 2015 Sep 8;10(9):e0137206.

13. Brage S, Ekelund U, Brage N, Hennings MA, Froberg K, Franks PW, et al. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. J Appl Physiol. 2007 Aug 1;103(2):682–92.

14. Brage S, Brage N, Franks PW, Ekelund U, Wong M-Y, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol. 2004;96(1):343–51.

15. Thompson D, Batterham AM, Bock S, Robson C, Stokes K. Assessment of low-to-moderate intensity physical activity thermogenesis in young adults using synchronized heart rate and accelerometry with branched-equation modeling. J Nutr. 2006 Apr;136(4):1037–42.

16. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis. JMIR MHealth UHealth. 2016 Apr 27;4(2):e40.

17. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable Technology and Physical Activity in Chronic Disease: Opportunities and Challenges. Am J Prev Med. 2018 Jan;54(1):144–50.

18. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR MHealth UHealth. 2018;6(8):e10527.

19. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. JMIR MHealth UHealth [Internet]. 2017 Mar 16 [cited 2018 Nov 26];5(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374271/

20. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions. JMIR MHealth UHealth [Internet]. 2017 Oct 20 [cited 2018 Nov 26];5(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5670315/

21. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med. 2018 Sep 7;

22. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act. 2015 Dec 18;12:159.

23. Hills AP, Mokhtar N, Byrne NM. Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures. Front Nutr [Internet]. 2014 Jun 16 [cited 2018 Nov 29];1. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4428382/

24. Horton JF, Stergiou P, Fung TS, Katz L. Comparison of Polar M600 Optical Heart Rate and Ecg Heart Rate during Exercise. Med Sci Sports Exerc. 2017 Dec 1;49(12):2600–7.

25. O'Connell S, ÓLaighin G, Quinlan LR. When a Step Is Not a Step! Specificity Analysis of Five Physical Activity Monitors. PLOS ONE. 2017 Jan 13;12(1):e0169616.

26. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Jan 1;4(1):1.

27. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev. 2015 Jun 26;4(1):82.

28. Van Remoortel H, Giavedoni S, Raste Y, Burtin C, Louvaris Z, Gimeno-Santos E, et al. Validity of activity monitors in health and chronic disease: a systematic review. Int J Behav Nutr Phys Act. 2012 Jul 9;9:84.

29. Valencell T. Valencell | Optical Heart Rate Measurement Top 5 Challenges [Internet]. Valencell. 2015 [cited 2019 Apr 12]. Available from: https://valencell.com/blog/2015/08/optical-heart-rate-measurement-top-5-challenges/

30. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, et al. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993 Jan;25(1):71–80.

31. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539–49.

32. Lu M-J, Zhong W-H, Liu Y-X, Miao H-Z, Li Y-C, Ji M-H. Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method. Int J Biostat. 2016 01;12(2).

33. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007 Mar 1;28(3):R1–39.

34. Tipton E, Shuster J. A framework for the meta-analysis of Bland-Altman studies based on a limits of agreement approach. Stat Med. 2017 Oct 15;36(23):3621–35.

35. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986 Sep;7(3):177–88.

36. Scammacca N, Roberts G, Stuebing KK. Meta-Analysis With Complex Research Designs: Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. Rev Educ Res. 2014 Sep 1;84(3):328–64.

37. Tanner-Smith EE, Tipton E. Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and spss. Res Synth Methods. 2014 Mar 1;5(1):13–30.

38. Welk G1, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing Analytic Methods and Reporting in Activity Monitor Validation Studies. Med Sci Sports Exerc. 2019 Aug;51(8):1767-1780. doi: 10.1249/MSS.0000000000001966.