

ORIGINAL  
Research

Danielle E. Jake-Schoffman, PhD , Valerie J. Silfee, PhD, Meera Sreedhara, MPH, Milagros C. Rosal, PhD, Christine N. May, PhD, Andrea Lopez-Cepero, Stephenie C. Lemon, PhD, and Christina F. Haughton, MPH, PhD

# Reporting of Physical Activity Device Measurement and Analysis Protocols in Lifestyle Interventions

**Abstract:** *This systematic review examined the extent to which lifestyle physical activity interventions that used wearable devices (eg, pedometers, accelerometers) reported on the length of device wear time requested in their protocols, criteria for analytic inclusion of data, and participant compliance with device use protocols. Literature searches were conducted using PubMed, Cochrane Central Register, and PsychInfo. Studies were included if they were the main outcomes paper of a trial that reported on a randomized or quasi-randomized trial focused on increasing lifestyle physical activity and were published between January 1, 2006 and March 30, 2016. Titles and abstracts were screened by 2 independent reviewers; eligible full texts were retrieved and reviewed by 2 independent reviewers. A total of 104 studies used wearable devices (n = 57 pedometers, n = 47 accelerometers). Most studies (n = 65, 67.3%) asked participants to wear devices for 7 days. Almost half of the studies (n = 46, 44.2%) did not report minimum device wear time required for analytic inclusion of data, and variation existed among studies reporting these criteria. Most studies (n = 60, 57.7%) did not report*

*average device wear time, or participant compliance with device wear. Overall, there was heterogeneity in reporting of physical activity device data. Refinement and streamlining of guidelines for device use, analysis, and reporting of data could improve comparability across studies.*

31.1% of adults worldwide can be classified as physically inactive.<sup>2</sup> One strategy to raise population levels of physical activity is the development and testing of behavioral interventions to promote lifestyle physical activity, focusing on increasing leisure-time



Overall, there was immense heterogeneity in all aspects of the reporting of physical activity device use in the context of the lifestyle physical activity intervention studies reviewed...



**Keywords:** lifestyle intervention; exercise; physical activity; guidelines and recommendations; measurement

While the health benefits of physical activity are well documented, including reduced risk of several chronic diseases,<sup>1</sup>

physical activity with account for the cultural and environmental contexts in which individuals carry on their everyday lives.<sup>3,4</sup> Despite a lack of a “gold standard” of physical activity measurement,<sup>5</sup> there is nonetheless an increased expectation to use wearable devices (eg, pedometers, accelerometers) for the objective measurement of physical activity intervention outcomes.<sup>6</sup> Evidence

DOI:10.1177/1559827619862179. From Department of Health Education and Behavior, College of Health and Human Performance, University of Florida, Gainesville, Florida (DEJ-S); Clinical Training and Development, UPMC Health Plan, Pittsburgh, Pennsylvania (VJS); Division of Preventive and Behavioral Medicine, Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts (MS, MCR, AL-C, SCL); Department of Psychology, Springfield College, Springfield, Massachusetts (CNM); and Friedman School of Nutrition and Policy, Tufts University, Boston, Massachusetts (CFH). Address correspondence to: Danielle E. Jake-Schoffman, PhD, Department of Health Education and Behavior, College of Health and Human Performance, University of Florida, PO Box 118210, Gainesville, FL 32611; e-mail: djake-schoffman@ufl.edu.

For reprints and permissions queries, please visit SAGE's Web site at [www.sagepub.com/journals-permissions](http://www.sagepub.com/journals-permissions)

Copyright © 2019 The Author(s)

to support the validity and reliability of wearable devices for capturing levels of and changes in physical activity<sup>6</sup> have been established in controlled physical activity studies, but less is known about the use of these tools in more general lifestyle interventions (ie, conducted in the field as opposed to in highly controlled or supervised settings).

There is a need for researchers to report on factors related to wearable physical activity devices in their research, in particular protocols for data collection, criteria for data inclusion, and analysis procedures. Knowledge of these data can aid in interpretation of the results of study outcomes by the scientific community. First, researchers should report how participants were instructed to wear the device (eg, study protocols often ask participants to wear a device during all waking hours except for during water-based activities). Second, researchers should report their criteria for inclusion of the device data in their analyses based on parameters related to the data collection (eg, participants may be asked to wear a device for 7 consecutive days during all waking hours,<sup>7</sup> but their data may be included only if they wear the device for a minimum of 4 days each with 10+ hours of wear time). Third, researchers can report the extent to which participants were compliant with the instructions and measurement protocols (eg, percentage of participants who wore their device for the minimum required time required for analytic inclusion). Comprehensive reporting on the range of factors relating to device data collection and analysis procedures would help other researchers interpret the results of a study, assess the quality of the data, and draw meaningful comparisons across studies.

The goal of this systematic review is to evaluate the extent to which lifestyle physical activity interventions that use wearable devices for the measurement of physical activity outcomes report on the measurement protocols, criteria for analytic inclusion of data, and participant compliance with device use protocols. This review is a secondary analysis of data abstracted for a systematic review of

the physical activity measurement tools used in lifestyle interventions.<sup>8</sup>

## Materials and Methods

### Data Sources

This study used data from a systematic review of lifestyle physical activity interventions; complete details of the review have been published elsewhere.<sup>8</sup> Briefly, a computerized search was conducted using 3 databases, PubMed, Cochrane Central Register, and PsychInfo, searching for peer-reviewed original research published between January 1, 2006 and March 30, 2016. Keywords used in the searches included (“physical activity” OR “physical activities” OR “exercise” OR “leisure time physical activity” OR “leisure time physical activities”) AND (“intervention” OR “interventions” OR “randomized controlled trial” OR “comparative study” OR “clinical trial”).

### Inclusion and Exclusion Criteria

To be included in the review, studies had to be randomized controlled trials or quasi-experimental interventions focused on increasing lifestyle physical activity among adults (>18 years of age) and published in English. Only primary data reports (no secondary analyses) were included.

### Data Extraction

Titles and abstracts of identified studies were screened by a group of 5 reviewers (CFH, ALC, CNM, MS, VJS). Each title and subsequent abstract were screened by 2 independent reviewers and discrepancies were discussed until consensus of inclusion or exclusion was reached. All reviewers were rotated through pairs, such that they were assigned to screen a portion of titles and abstracts with each other reviewer. Interrater agreement (IRA) for titles was 99.6% agreement and for abstracts was 89.7% agreement. Full texts of the remaining eligible articles were retrieved and screened by 6 reviewers (CFH, DEJS, ALC, CNM, MS, VJS) using a standardized data abstraction form in REDcap. Each full text was screened by 2 independent reviewers and discrepancies

in data abstraction were discussed until consensus on the coding was reached. All reviewers were rotated through pairs, such that they were assigned to screen a portion of full texts with each other reviewer. The present review only includes information about the articles reporting on objective measures of physical activity. The data include details on the measurement tools used, data collection protocols, criteria for analytic inclusion, and participant compliance with measurement protocols.

*Type of Physical Activity Measurement Tool.* Reviewers coded the measures used in the study, selecting all applicable measures used in the study from a list of tools (eg, pedometers, accelerometers, heart rate monitors, multi-sensor devices, self-report; IRA=92.2%). Multisensor devices (eg, BodyMedia armbands), were excluded from the present analysis due to small sample size (n = 3) and lack of comparability to other devices (eg, collection of many additional metrics such as body temperature and heart rate with varying standards for data collection). Studies that used both pedometers and accelerometers for physical activity outcomes (n = 4) were counted once in each category.

*Length of Device Wear Time Requested in Protocol.* Reviewers coded how long participants were asked to wear the device during the measurement period via open response (eg, 7 days, 10 hours a day for 1 week; IRA = 83.5%).

*Criteria for Analytic Inclusion of Data.* Reviewers coded the amount of physical activity data required for each participant to be included in the analyses via open response (eg, 4 days with 10 hours of data; IRA = 87.4%).

*Average Device Wear Time and Participant Compliance.* Reviewers coded if the study reported the amount of time participants wore the device during the measurement periods. Data on wear time and compliance were categorized in the following manner: study reported average device wear time (eg, an average of 5 days with 14 hours of wear time) and/or

the study reported information about how many participants had data that met the criteria for analytic inclusion (eg, 75.5% of participants had at least 4 days with 12 hours of wear time; IRA = 81.6%).

### Data Synthesis

Descriptive analyses of the abstracted studies included frequencies and proportions of studies using different measurement tools, measurement protocols, and criteria for analytic inclusion. Statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC).

## Results

The systematic literature search yielded a total of 13 718 unique articles (Figure 1). Of those, 10 903 were excluded based on the title review, leaving 2815 articles for abstract review. Of these, 1498 were excluded, leaving 1317 for full text review. Overall, 103 studies used objective physical activity measures, of which 3 used multisensor devices and were excluded, leaving 100 studies for inclusion in the present analysis. The full list of included articles is available in the Supplemental Material (available online).

### Results of Data Abstraction

*Type of Physical Activity Device.* Of the 100 studies using an objective measure of physical activity, 53 used pedometers, 43 used accelerometers, and 4 used both pedometers and accelerometers. For the purposes of our analyses, studies that used pedometers and accelerometers were counted as a separate study under each device type, yielding an analytic sample of 104 study entries (Figure 1).

*Length of Device Wear Time Requested in Protocol.* Most of the studies (n = 70, 67.3%) reported that participants were asked to wear a physical activity device for exactly 7 days during the outcome measurement periods (Table 1). An additional 7 studies (6.7%) did not report how long participants were asked to wear a device.

*Criteria for Analytic Inclusion of Data.* Overall, about half of the studies (n = 46, 44.2%) did not report the minimum number of days or hours required to include a participant's data in their analysis, although more studies using accelerometers than pedometers reported their criteria for analytic inclusion (n = 35 vs n = 23; Table 2). Of the 58 studies that reported on the minimum data needed for inclusion in analysis, there was considerable variation in the criteria used. Over half (n = 24, 51.1%) of the studies using accelerometers studies and over a third (n=21, 37.5%) of the studies using pedometers reported a requirement for at least 3 valid days of participant data for inclusion in the analysis. The most common set of criteria for analytic inclusion of pedometer data was 3 days (n = 6, 10.5%) and accelerometer data was 3 days with at least 10 hours of wear time (n = 7, 14.9%).

*Average Device Wear Time and Participant Compliance.* Most studies did not report on average device wear time or percent of participant compliance with measurement protocols (n = 60, 57.7%), and more studies using pedometers than accelerometers lacked this information (n = 40 vs n = 20; Table 2). By device type, 29.8% (n = 17) of studies using pedometers and 57.4% (n = 27) of studies using accelerometers reported on participant compliance with measurement protocols (Table 3).

## Discussion

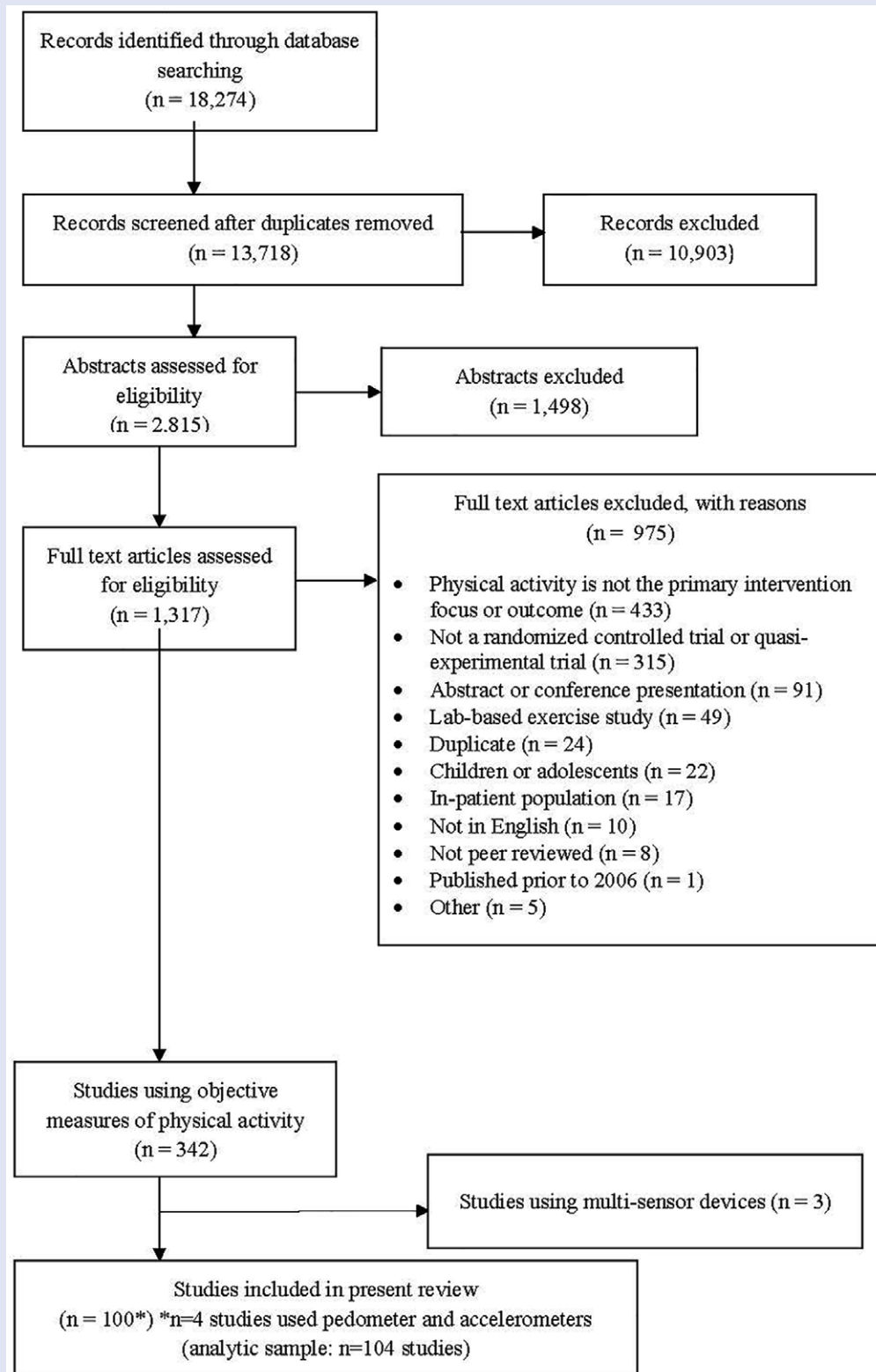
Overall, there was immense heterogeneity in all aspects of the reporting of physical activity device use in the context of the lifestyle physical activity intervention studies reviewed, including measurement protocols, criteria for analytic inclusion of data, and device wear time and participant compliance with measurement protocols. While some variation in reporting and criteria used is to be expected, our findings show that there is a great deal of diversity in cut-points used and lack of standardized reporting.

Our results demonstrate the lack of consistency across studies in terms of the criteria used to determine analytic inclusion of device data. While there are no set protocols for data inclusion, there is a precedent to follow the recommendations that have been established through longitudinal and intervention research, providing suggestions about of the minimum number of hours and days needed to provide a reliable estimate of physical activity in adults.<sup>6,9-11</sup> For pedometer data, a minimum of between 2 and 4 days of measurement data are recommended for a reliable estimate,<sup>10,12-16</sup> while for accelerometer data, the number of days of data required depends on the outcomes of interest. A minimum of 2 days of accelerometer data is recommended for a reliable measure of steps per day, 3 days of data are recommended for an estimate of total physical activity and time spent in moderate- to vigorous-intensity physical activity, and a minimum of 6 days are required to examine continuous 10-minute bouts of moderate- to vigorous-intensity physical activity.<sup>12,13,17</sup> Analyses of accelerometer data from the NHANES (National Health and Nutrition Examination Survey) used a minimum of 4 days with 10 hours of wear time to estimate the percentage of adults and children meeting national physical activity recommendations.<sup>9</sup> In our sample, only 21 (36.8%) of the studies using pedometers reported that they used the criterion of a minimum of 3 days of data for inclusion, while only 8 (17.0%) of the studies using accelerometers used the criteria of 4 days with at least 10 hours of wear time for inclusion. Furthermore, 46 (44.2%) of studies did not report the criteria they used to determine analytic inclusion, limiting the interpretability of their results.

One possible explanation for the lack of reporting consistency in the use of established cut-points for analytic inclusion could be the diverse fields that the included studies come from. Lifestyle physical activity interventions are designed and implemented in a range of

**Figure 1.**

Flow diagram of article inclusion and exclusion.



**Table 1.**

Number of Days of Physical Activity Device Wear Time Requested in Measurement Protocol of Included Studies (n = 104) by Device Type.

Number of Days	Studies Using Pedometers (N = 57), n (%)	Studies Using Accelerometers (N = 47), n (%)	Total Studies (N = 104), n (%)
1	1 (1.8)	0 (0)	1 (1.0)
3	2 (3.5)	0 (0)	2 (1.9)
4	0 (0)	2 (4.3)	2 (1.9)
5	4 (7.0)	2 (4.3)	6 (5.8)
6	2 (3.5)	0 (0)	2 (1.9)
7	35 (61.4)	35 (74.5)	65 (67.3)
8	0 (0)	1 (2.1)	1 (1.0)
14	2 (2.9)	1 (2.1)	3 (2.9)
Entire intervention	11 (19.3) <sup>a</sup>	5 (10.6) <sup>b</sup>	16 (15.4)
Not reported	6 (10.5)	1 (2.13)	7 (6.7)

<sup>a</sup>Range 28 to 180 days.

<sup>b</sup>Range 42 to 336 days.

contexts (eg, community centers, health care facilities, web-based interventions) by researchers and practitioners with diverse specializations that may fall outside of physical activity (eg, nursing, nutrition). With little guidance or expertise in physical activity measurement, researchers may find it challenging to select and incorporate a measurement tool that best matches the needs of their study.<sup>18</sup> Furthermore, with little standardization in measurement adherence and compliance protocols, researchers may lack an in-depth understanding of physical activity devices or the skills to report their outcomes based on published guidelines. However, despite the absence of rigid guidelines for physical activity measurement, future research should strive to report on the range of factors reported here (eg, measurement protocols used for data collection and analysis, participant compliance with protocols) to increase consistency.

The lack of consistency in reporting of physical activity protocols, adherence to

device use instructions, and data analysis procedures has important implications for research and public health practice. Critically, it limits the ability of experts to interpret study results, compare results across studies, and ultimately learn from what works well and for whom. Ultimately, standardized reporting of study procedures and results is a critical factor for researchers to be able to justify the validity, reliability, and impact of their outcomes.<sup>19</sup> The results of the present study demonstrate a need for improved consistency of reporting on the use of physical activity devices for measurement, including protocols used to instruct participants on device use, criteria for analytic inclusion, and average participant wear time or compliance to measurement protocols.

The findings of the present review should be interpreted in the context of several limitations. First, the review was limited to studies available through electronic databases and did not include conference proceedings or unpublished

studies. Second, the review was limited to articles published in English, limiting the inclusion of international articles and potentially limiting the generalizability of the findings. Third, although extensive efforts to locate the full texts of articles were made, there was a small number (n = 2) of articles that were excluded from further review because a full text could not be obtained. Fourth, studies where physical activity was not the primary outcome were excluded from this review, such as studies that focused on weight loss or other chronic disease outcomes. Thus, the results of this review may not be generalizable to studies where objectively measured physical activity is a secondary outcome. Finally, because of constraints on data abstraction, this review does not include 2 additional important dimensions of physical activity measurement—device placement (eg, hip or wrist) and inclusion of weekday vs weekend days of device wear. These variables could also affect adherence to measurement protocols and should be reported alongside physical activity

**Table 2.**

Criteria Used to Determine Analytic Inclusion of Physical Activity Device Outcome Data in Included Studies (n = 104), by Device Type.

Criteria Used	Studies Using Pedometers (N = 57), n (%)	Studies Using Accelerometers (N = 47), n (%)	Total Studies (N = 104), n (%)
100 steps	1 (1.8)	0 (0)	1 (1.0)
>1000 steps	1 (1.8)	1 (2.1)	2 (1.9)
6 hours	0 (0.0)	1 (2.1)	1 (1.0)
8 hours	0 (0)	1 (2.1)	1 (1.0)
10 hours	0 (0)	4 (8.5)	4 (3.8)
1 day, 4 hours	0 (0)	1 (2.1)	1 (1.0)
3 days	6 (10.5)	0 (0.0)	6 (5.8)
3 days, 100 steps	1 (1.8)	0 (0)	1 (1.0)
3 days, 1200 steps	0 (0)	1 (2.1)	1 (1.0)
3 days with highest step count	1 (1.8)	0 (0)	1 (1.0)
3 days, 6 hours	0 (0.0)	1 (2.1)	1 (1.0)
3 days, 8 hours	1 (1.8)	1 (2.1)	2 (1.9)
3 days, 8 $\frac{1}{3}$ hours	0 (0)	1 (2.1)	1 (1.0)
3 days, 8 hours, 500 steps	1 (1.8)	0 (0)	1 (1.0)
3 days, 10 hours	1 (1.8)	7 (14.9)	8 (7.7)
4 days	3 (5.3)	1 (2.1)	4 (3.8)
4 days, 8 hours	0 (0)	1 (2.1)	1 (1.0)
4 days, 10 hours	0 (0)	5 (10.6)	5 (4.8)
5 days	2 (3.5)	2 (4.3)	4 (3.8)
5 days, 8 hours	1 (1.8)	1 (2.1)	2 (1.9)
5 days, 10 hours	0 (0)	2 (4.3)	2 (1.9)
6 days	1 (1.8)	0 (0.0)	1 (1.0)
7 days, 8 hours	1 (1.8)	0 (0.0)	1 (1.0)
7 days, 10 hours	0 (0)	1 (2.1)	1 (1.0)
4 weeks, 6 hours	0 (0)	1 (2.1)	1 (1.0)
All 56 days	1 (1.8)	0 (0)	1 (1.0)
Not reported	34 (59.6)	12 (25.5)	46 (44.2)

**Table 3.**

Criteria Used to Determine Average Wear Time and Participant Compliance With Physical Activity Device Measurement Protocols of Included Studies (n = 104), by Device Type.

Criteria	Studies Using Pedometers (N = 57), n (%)	Studies Using Accelerometers (N = 47), n (%)	Total Studies (N = 104), n (%)
Average wear time of device	5 (8.8)	11 (23.4)	16 (15.4)
Number of participants meeting criteria for analytic inclusion	9 (15.8)	11 (23.4)	20 (19.2)
Average wear time and number participants meeting criteria for analytic inclusion	3 (5.3)	5 (10.6)	8 (7.7)
Not reported	40 (70.2)	20 (42.6)	60 (57.7)

results to improve interpretation of data. Despite these limitations, the present review represents a novel overview of the reporting of physical activity measurement data and analysis criteria in lifestyle physical activity interventions.

### Conclusions

Measuring physical activity in the context of lifestyle physical activity interventions is critical to understanding the extent to which studies are stimulating meaningful changes in behavior. Results of this study indicate a need for improved reporting to harness the full benefits of using objective measures including the ability to compare outcomes across studies and summarize physical activity results to advance the science of lifestyle interventions.

### Acknowledgments

We acknowledge the contributions of our University of Massachusetts Medical School colleagues and staff (Nancy Harger and Christine Frisard) that made this research possible.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Author time during the preparation of this manuscript was provided by NIH grants R25CA172009 (DEJ-S, CNM), F31HL138970 (CFH), R25 GM113686-02 (ALC), UL1TR001453 (ALC, MS), F31HL142139 (MS), U46DP005031 (VS, CFH, MS, ALC, SCL, MCR), and 1 P60 MD0006912 (MCR, SCL).

### Ethical Approval

Not applicable, because this article does not contain any studies with human or animal subjects.

### Informed Consent

Not applicable, because this article does not contain any studies with human or animal subjects.

### Trial Registration

Not applicable, because this article does not contain any clinical trials.

### ORCID ID

Danielle E. Jake-Schoffman  <https://orcid.org/0000-0001-6381-7323>

### Supplemental Material

Supplemental material for this article is available online.

### References

1. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT; Lancet Physical

- Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380:219-229.
2. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U; Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380:247-257.
3. Dunn AL, Andersen RE, Jakicic JM. Lifestyle physical activity interventions. History, short- and long-term effects, and recommendations. *Am J Prev Med*. 1998;15:398-412.
4. Kahn EB, Ramsey LT, Brownson RC, et al. The effectiveness of interventions to increase physical activity. A systematic review. *Am J Prev Med*. 2002;22(4 suppl):73-107.
5. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. *Front Nutr*. 2014;1:5.
6. Dowd KP, Szeklicki R, Minetto MA, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. *Int J Behav Nutr Phy Act*. 2018;15:15.
7. Cain KL, Geremia CM. Accelerometer data collection and scoring manual. [http://sallis.ucsd.edu/Documents/Measures\\_documents/Accelerometer\\_Data\\_Collection\\_and\\_Scoring\\_Manual\\_Updated\\_June2012.pdf](http://sallis.ucsd.edu/Documents/Measures_documents/Accelerometer_Data_Collection_and_Scoring_Manual_Updated_June2012.pdf). Accessed April 15, 2017.
8. Silfee VJ, Houghton CF, Jake-Schoffman DE, et al. Objective measurement of physical activity outcomes in lifestyle

- interventions among adults: a systematic review. *Prev Med Rep.* 2018;11:74-80.
9. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40:181-188.
  10. Tudor-Locke C, Burkett L, Reis JP, Ainsworth BE, Macera CA, Wilson DK. How many days of pedometer monitoring predict weekly physical activity in adults? *Prev Med.* 2005;40:293-298.
  11. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc.* 2012;44(1 suppl 1):S1-S4.
  12. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? *Int J Behav Nutr Phys Act.* 2011;8:62.
  13. Rowe DA, Kemble CD, Robinson TS, Mahar MT. Daily walking in older adults: day-to-day variability and criterion-referenced validity of total daily step counts. *J Phys Act Health.* 2007;4:434-446.
  14. Felton GM, Tudor-Locke C, Burkett L. Reliability of pedometer-determined free-living physical activity data in college women. *Res Q Exerc Sport.* 2006;77:304-308.
  15. Strycker LA, Duncan SC, Chaumeton NR, Duncan TE, Toobert DJ. Reliability of pedometer data in samples of youth and older women. *Int J Behav Nutr Phys Act.* 2007;4:4.
  16. Kang M, Bassett DR, Barreira TV, et al. How many days are enough? A study of 365 days of pedometer monitoring. *Res Q Exerc Sport.* 2009;80:445-453.
  17. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR Jr. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med Sci Sports Exerc.* 2002;34:1376-1381.
  18. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: clinical and research applications: a scientific statement from the American Heart Association. *Circulation.* 2013;128:2259-2279.
  19. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.