Published in final edited form as:

J Phys Act Health. 2016 February ; 13(2): 140–144. doi:10.1123/jpah.2014-0592.

Validation of a New Counter for Direct Observation of Physical Activity in Parks

Bing Han¹, Deborah A. Cohen¹, Kathryn Pitkin Derose¹, Terry Marsh¹, Stephanie Williamson¹, and Laura Raaen¹

¹RAND Corporation, Santa Monica, CA 90401

Abstract

Purpose—Prior tools to observe large groups of people in parks have not allowed disaggregation of physical activity levels by age group and gender simultaneously, making it impossible to determine which subgroups engaged in moderate to vigorous physical activity (MVPA). This study aims to examine the reliability of a 12-button counter to simultaneously assess MVPA by age and gender subgroups in park settings.

Methods—A total of 1,160 pairs of observations were conducted in 481 target areas of 19 neighborhood parks in the great Los Angeles area between June 2013 and March 2014. Inter-rater reliability was assessed by Pearson's correlation, intra-class correlation (ICC), and agreement probability in the total metabolic equivalents (METs) and METs spent in MVPA. Cosine similarity was used to check the resemblance of distributions among age and gender categories. Pictures taken in a total of 112 target areas at the beginning of the observations were used as a second check on the reliability of direct observation.

Results—Inter-rater reliability was high for the total METs and METs in all age and gender categories (between 0.82 and 0.97), except for male seniors (correlations and ICC between 0.64 and 0.77, agreement probability 0.85 to 0.86). Reliability was higher for total METs than for METs spent in MVPA. Correlation and ICC between observers' measurement and picture-based counts are also high (between 0.79 and 0.94).

Conclusion—Trained observers can reliably use the 12-button counter to accurately assess PA distribution and disparities by age and gender.

Keywords

direct observation; counter; reliability; validation; neighborhood parks

INTRODUCTION

Direct observation is a popular method for assessing physical activity (PA), which requires a designated observer to document the PA behavior of one or more subjects for a short

Corresponding author: Bing Han, 1776 Main St., Santa Monica, CA 90401, Telephone: 310-393-0411 x6175, Fax: 310-393-4818, bhan@rand.org.

CONFLICT OF INTEREST

The authors claim no conflict of interest. The results of the present study do not constitute endorsement by ACSM.

duration (11, 13, 16). Sometimes multiple observations are used to summarize the PA behavior according to a pre-specified protocol (11, 13, 16). Coding of PA is often conducted in real time but can also be processed post-hoc from pictures and videotapes. Direct observation can be used to measure a single individual's physical activity over time, and to measure the physical activity of a group of people in various built environments, e.g., gymnasiums, neighborhood parks, walking paths, bike lanes, etc. While costly for following a single subject over time (11), direct observation is widely adopted as a feasible and cost-effective approach to studying PA in built environments (1, 7, 10, 12, 17–20).

Some built environments can accommodate a large group of people in different PA levels at the same time, and the composition of the crowd can vary over time. For example, under clement weather a public park with 10 acres of land in an urban neighborhood may contain 300 to 1,000 people at the same time, all engaging in different activities (8, 9). There can also be a flow of new arrivals and departures throughout a day. Measuring PA in such a dynamic and complex setting using traditional instruments such as accelerometers or self-report is difficult. In recent years, direct observations have become the main measurement method for studying park-based PA (4–6, 14). Moreover, direct observations have the additional advantage of being able to simultaneously record contextual and environmental information about where PA occurs, which can help to explain the observations (2, 3).

Many direct observation protocols, with the exception of those using video recordings, are limited in their ability to associate the PA level of individuals with numerous personal characteristics, preventing researchers from accurately assessing the effects of PA promotion interventions in built environments such as neighborhood parks. For example, the SOPARC protocol separately assesses one characteristic at a time by gender: PA (sedentary, moderate, and vigorous), age group (children, teenagers, adults, and seniors), or race/ethnicity (white, black, Hispanic, Asian and others), and repeats the counting twice for each gender (12). The resultant data cannot be disentangled to identify which subpopulations are engaging in PA in parks, except by gender. For example, it is reported that most people using parks are males and that they are more physically active than females in parks. However, it is unclear whether park facilities and programming particularly support PA among children, teens, adults or seniors, and whether males are the majority in all age groups.

Simultaneously observing two or more factors among a group of people in a short time has been considered challenging for observers in the field. In the SOPARC protocol, since an observer works with one factor at a time that has either three or four levels, a mechanical counter with four buttons has been recommended in an observed area. Observing more factors simultaneously would require a counter with more buttons (e.g., combining PA levels and age group would yield 12 buttons). A mechanical counter having so many buttons would be difficult to manipulate and it would be easy to make a mistake but hard to recognize and correct it.

In this paper we introduce an enhanced direct observation tool, a 12-button counter implemented on a tablet computer. With the new 12-button counter, an observer can simultaneously work with two factors: PA levels (sedentary, moderate, and vigorous) and age groups (children, teenagers, adults, and seniors). In conjunction with two rounds of

observations by each gender, this counter enables researchers to measure the detailed distribution of PA by age group and gender. Each time a button is pressed, a click sound is played and the corresponding button is highlighted to give the observer feedback about the input. There is also a function to correct mistakes in the preceding entry. The counter can be run on many smart devices including iOS, Android, and laptop PC. Internet connection is not required to use the counter in the field. The new counter can replace the mechanical counter commonly used in SOPARC and other similar study protocols for measuring PA in built environments.

Working with one observation factor at a time, entailing the use of at most four buttons, observers using the current SOPARC protocol have a relatively low work load, resulting in very high inter-rater reliability (12). While the new 12-button counter will offer more detailed data for analysis, it also could increase the work load of observers and potentially reduce the reliability. In the study described in this paper we conducted two validation experiments to check the reliability of the new 12-button counter. The first experiment checked whether two observers could agree on their assessments of the same area at the same time, i.e., inter-rater reliability. The second experiment checked whether the observers' measurements were similar to the counts retrieved from a picture taken at the beginning of the observation. If the picture processing was deemed as a different type of measurement, the second experiment could be seen as a check for inter-instrument reliability.

METHODS

The tablet-based 12-button counter and the observation protocol

The 12-button counter has a three by four layout, where rows are for PA levels and columns are age groups. Rows and columns are ordered in the same order as the factor levels. Different colors differentiate age groups, and three thumbnail icons differentiate PA levels. Both iOS and Android devices are supported and both have been extensively tested. In iOS the counter is implemented as a static html 5 webpage, and in Android it is an Android App. In both platforms, after downloading the counter to a tablet, network connection is not required to use the counter. The counter can also be loaded to a laptop PC and used in the offline mode of an Internet browser. The upper frame in Figure 1 shows a screen snapshot of the counter taken from an Android device. The counter can be downloaded from <http://mmicdata.rand.org/parkcounter>.

The protocol for using the new counter is similar to previous SOPARC protocols. To scan a target area, an observer selects a vantage point with no visual obstructions. If this is not possible, the area can be split into subareas and observed separately. The size of an area needs to be manageable so that all individuals in an area can be counted accurately. Each area needs two rounds of observations, one for each gender. In each round, an observer scans from his or her left side to the right side. For each new individual, an observer uses his or her best judgment on the age group and the person's momentary physical activity level, and press the corresponding button on the counter. The scan continues until all females are observed, and then repeats for males. An area is finished when both genders are observed. A recording error can be corrected by the "Edit Previous" link at the bottom. To avoid duplicated counts in an area, scanning and recording need to be as fast as accuracy allows.

After the counting is finished, the observer can press the link titled “Observation Count” at the bottom of the screen to show the counting results. (See the lower frame of Figure 1 for an example.) After recording the detailed counts into the database, a “Reset Counters” link next to it is used to reset and prepare for the next observation.

Observation data

Between June 2013 and March 2014, eleven community health workers (“*promotoras*”) visited 481 target areas in 19 neighborhood parks in the great Los Angeles area. All target areas were visited three times a day and on three or four days, where all days were in the same week. In each instance, two observers synchronized the start of their scans and independently observed the target area, and another two persons helped recording the final counts into the database. Team composition varied across all park visits due to logistical reasons. In 1,160 pairs of observations at least one observer recorded one or more person. These 1,160 pairs of observations were used to check the inter-rater reliability.

In another 112 observations where at least one person was present, still pictures were taken before observing the target area. Usually two pictures were taken per target area (one to the left and one to the right of the observer, with some space overlapping). Additional pictures were taken from different angles if a target area was large (e.g., an indoor gym) or contained visual obstructions (e.g., a playground). These pictures were intended to provide a second check of the observer’s reliability. We post-processed the pictures to retrieve the total number of people and total physical activity level in metabolic equivalents. However, it was not possible to fully identify the age group and gender for all individuals in the scene, (e.g., some people with their backs to the camera, some figures partially blocked by others).

Statistical analysis

Physical activity for all observed subjects in a gender and age category was measured on a scale of total metabolic equivalents (METs), using the conversion rule of 6 METs for vigorous activity, 3 METs for moderate activity, and 1.5 METs for any activity below moderate. The overall agreement within a pair of measurements, i.e., between two observers (inter-rater reliability) or between an observer and a picture (inter-instrument reliability), was assessed by Pearson’s correlation and intra-class correlation (ICC), where the ICC was equal to the ratio of between-pair variance divided by the sum of between-pair variance and within-pair variance. The two variance components were estimated by the linear mixed-effect model.

We used two additional measures to check detailed inter-rater reliability: the probability of agreement and the mean cosine similarity. We considered two METs measurements to be in agreement if they were identical or differed by no more than 1.5 METs (e.g., missing or over counting no more than one sedentary person, or misclassifying no more than one person between sedentary and moderate activity levels). This is similar to the relaxation of agreement used in testing reliability of the original SOPARC tool (12) to allow for a small level of measurement error when the observers’ work load is high. The cosine similarity is a mathematical measure of similarity between two vectors, in particular, when the entries of the two vectors take non-negative values only (15). Each observation can be represented by a

vector of METs, where each entry represents the METs for a category of age group by gender. The cosine similarity is the cosine of the angle between the two vectors in the multidimensional Euclidean space. It represents the level of similarity in the entire distribution of METs by age and gender between two records. For non-negative measures, the cosine similarity is between 0 and 1, where 0 is independent or perpendicular and 1 is identical or overlapping.

RESULTS

Table 1 provides descriptive statistics of the average METs in each target area. The standard deviations in Table 1 are measures of variation among target areas. The mean METs in a target area is 17.6 (95% CI: 16.0~19.2), roughly equivalent to 12 sedentary people or six moderately active people or three vigorously active people. Most of the total METs were attributed to males, children, and adults. Males accounted for about 70% of METs spent in MVPA. There was a very small amount of METs generated by seniors.

Inter-rater reliability of total METs was high. The Pearson's correlation between the 1,160 pairs of observations is 0.97, the ICC is 0.98, and the agreement probability is 0.82. The mean cosine similarity is 0.95. Table 2 shows the inter-rater reliability measures by age and gender categories, which are generally high (>0.80) or very high (>0.90). However, correlations and ICC are only moderate for senior males (between 0.64 and 0.77). Inter-rater reliability is higher for total METs than for METs spent in MVPA in each age by gender category.

Inter-instrument reliability check was limited to total number of persons and total METs of an area, because the pictures did not have sufficient details to tell the age and gender of every person. The correlation between the picture-based measurements and field measurements by the 12-button counters was 0.94 for the total number of persons and 0.80 for total METs. The ICC was 0.92 for the total number of persons and 0.79 for total METs.

DISCUSSION

A major limitation of existing direct observation tools for measuring park-based PA, e.g., SOPARC, is their inability to associate the PA level of individuals with multiple personal characteristics. With these previously reported tools we cannot fully investigate disparities in PA or evaluate the differential effects of an intervention among different subpopulations, except between males and females. This limitation severely hampers the study of physical activity in park settings as well as other built environments. Our 12-button counter provides a reliable solution. Despite the increased work load of operating 12 buttons at a time, trained observers were able to measure the total physical activity levels and details by age and gender reliability with a high level of reliability. Implementation of the counter is by tablet computers, which are convenient to use in the field and can potentially be combined or used in conjunction with other survey or data measurement tools implemented with smart devices.

This refinement is necessary to study the detailed MVPA distribution by age and gender, and to evaluate programs intended to promote MVPA targeting different subpopulations. For example, health promotion programs such as free exercise classes and diabetic prevention

programs in parks during business hours might target seniors and adult females, and improvements to fitness zones and various sports fields and facilities might target other particular age and gender groups. Inter-generational playgrounds may increase the caregivers' MVPA compared to ordinary playgrounds, but may or may not increase children's MVPA. The new 12-button counter can be used to evaluate whether particular target groups are benefiting from such interventions.

Our reliability study was based on a representative sample of the large Los Angeles neighborhood park system and reflects actual physical activity levels in parks. We observed a mixture of age groups, genders, and activity levels with a distribution very similar to the entire neighborhood park system in Los Angeles [see (9)] for the use of the LA neighborhood park system). Hence, our results have a high level of generalizability.

The relatively low correlation and ICC for senior males are largely due to the fact that there were very few senior male park users observed. A small difference between a pair of observers results in a major discrepancy. In almost all target areas there were no more than two senior males. Counting one senior male in moderate activity versus sedentary activity can result in a 30–100% difference between a pair of observers. Occasional discrepancies such as this example drove the correlation and ICC down since both statistical measures are known to be sensitive to outliers. However, the agreement probability measure allows for an inter-rater difference of 1.5 METs and thus is robust to such small discrepancies. Hence, the agreement probability in the senior male category is still high (>0.80).

In contrast, in Table 2 all inter-rater reliability measures for the senior female category are high. In one of the study parks we encountered a special exercise classes that involved an extraordinarily large group of senior females but almost no senior males. These observations had 10 or more senior females present in the same target area, and a small discrepancy did not yield a large relative difference. Small discrepancies in a regular target area with no more than two senior women present did not have the same level of influence due to these observations with higher counts. In theory, given sufficient numbers of senior park users in all levels of physical activity, we anticipate that measurements of seniors would be similar between the two genders.

Total METs are measured more reliably than the METs spent in MVPA for two reasons. First, it is more difficult to differentiate vigorous activity from moderate activity than to identify sedentary activity. METs spent in MVPA did not include those in sedentary status, who presumably would incur less measurement errors. Second, only about one third of the park users were in MVPA, and most areas did not have a large number of people in MVPA relative to sedentary activity (e.g., a supervised youth basketball game could have 12 players and referees in MVPA, but more than 50 sedentary people watching it). The relatively low METs spent in MVPA (see Table 1) makes the impact of measurement error (e.g., pressing a wrong button without correction) relatively larger. Inter-instrument reliability was somewhat lower (0.8) than the inter-rater reliability for total METs, because it was not always possible to definitively judge a person's PA status in still pictures. For the total number of people, the inter-instrument reliability is still very high.

Training for the new 12-button counter follows the same procedure as the current SOPARC protocol except for specifics on operating the new counter and the simultaneous observation of multiple factors. In training new users of SOPARC with the 12-button counter, our level of effort was about the same as with the mechanical counter. Since a tablet is often used for other purposes in the field (e.g., electronic survey forms, pictures, data recording, logs), it has the additional advantage of not having to carry to a separate and heavy mechanical counter. We also note that current SOPARC users with previous experience in using mechanical counters need some practice to adapt to the tablet-based counter due to its different button layout, lack of force feedback on touch, and different sound feedback.

In conclusion, we recommend the adoption of the new 12-button protocol for any future studies interested in assessing age-group specific physical activity levels in parks or other places where people are active.

ACKNOWLEDGMENTS

This study was supported by a grant from NIH/NHLBI (R01HL114283). We thank Mr. Alerk Amin at the information services of RAND Corporation for programming the 12-button counter.

References

1. Bocarro JN, Floyd M, Moore R, et al. Adaptation of the System for Observing Physical Activity and Recreation in Communities (SOPARC) to assess age groupings of children. *J Phys Act Health*. 2009; 6(6):699–707. Epub 2010/01/28. PubMed PMID: 20101912. [PubMed: 20101912]
2. Cohen D, Han B, Isacoff J, et al. Impact of park renovations on park use and park-based physical activity. *Journal of Physical Activity & Health*. 2013 In press. Epub 2014 Jun 20.
3. Cohen D, Marsh T, Williamson S, et al. The potential for pocket parks and to increase physical activity. *American Journal of Health Promotion*. 2013 In press. Epub 2014 Jan-Feb. 28(3).
4. Cohen DA, Han B, Derosé KP, Williamson S, Marsh T, McKenzie TL. Physical Activity in Parks: A Randomized Controlled Trial Using Community Engagement. *Am J Prev Med*. 2013; 45(5):590–597. [PubMed: 24139772]
5. Cohen DA, Han B, Derosé KP, et al. Neighborhood poverty, park use, and park-based physical activity in a Southern California city. *Social Science & Medicine*. 2012; 75(12):2317–2325. [PubMed: 23010338]
6. Cohen DA, Marsh T, Williamson S, Golinelli D, McKenzie TL. Impact and cost-effectiveness of family Fitness Zones: a natural experiment in urban public parks. *Health & place*. 2012; 18(1):39–45. PubMed PMID: 22243905; PubMed Central PMCID: PMC3308725. [PubMed: 22243905]
7. Floyd MF, Bocarro JN, Smith WR, et al. Park-Based Physical Activity Among Children and Adolescents. *Am J Prev Med*. 2011; 41(3):258–265. PubMed PMID: ISI: 000294002700005. [PubMed: 21855739]
8. Han B, Cohen D, McKenzie TL. Quantifying the contribution of neighborhood parks to physical activity. *Preventive Medicine*. 2013; 57(5):483–487. [PubMed: 23827723]
9. Han B, Cohen DA, Derosé KP, Marsh T, Williamson S, Raaen L. How Much Do Neighborhood Parks Contribute to Local Residents' Physical Activity in the City of Los Angeles: A Meta-Analysis. *Preventive Medicine*. 2014 In press. Epub 2014 Sep 4.
10. Hino AAF, Reis RS, Ribeiro IC, Parra DC, Brownson RC, Fermino RC. Using Observational Methods to Evaluate Public Open Spaces and Physical Activity in Brazil. *Journal of Physical Activity & Health*. 2010; 7:S146–S154. PubMed PMID: ISI: 000280451800005. [PubMed: 20702903]
11. Kohl HWI, Fulton JE, Caspersen CJ. Assessment of Physical Activity among Children and Adolescents: A Review and Synthesis. *Preventive Medicine*. 2000; 31(2):S54–S76. doi: <http://dx.doi.org/10.1006/pmed.1999.0542>.

12. McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Parks and Recreation in Communities (SOPARC): Reliability and feasibility measures. *Jl of Physical Activity and Health*. 2006; 3(Suppl 1):S208–S222.
13. McKenzie TL, Sallis JF, Nader PR. SOFIT: System for Observing Fitness Instruction Time. *Journal of Teaching in Physical Education*. 1991; 11:195–205.
14. Reed JA, Price AE, Grost L, Mantinan K. Demographic Characteristics and Physical Activity Behaviors in Sixteen Michigan Parks. *Journal of community health*. 2012; 37(2):507–512. [PubMed: 21922166]
15. Singhal A. Modern information retrieval: A brief overview. *IEEE Data Eng Bull*. 2001; 24(4):35–43.
16. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sports medicine*. 2001; 31(6):439–454. PubMed PMID:11394563. [PubMed: 11394563]
17. Spengler JO, Floyd MF, Maddock JE, Gobster PH, Suau LJ, Norman GJ. Correlates of park-based physical activity among children in diverse communities: results from an observational study in two cities. *Am J Health Promot*. 2011; 25(5):e1–e9. Epub 2011/05/04. PubMed PMID: 21534825. [PubMed: 21534825]
18. Suau LJ, Floyd MF, Spengler JO, Maddock JE, Gobster PH. Energy expenditure associated with the use of neighborhood parks in 2 cities. *J Public Health Manag Pract*. 2012; 18(5):440–444. Epub 2012/07/28. PubMed PMID-22836535. [PubMed: 22836535]
19. Tester J, Baker R. Making the playfields even: evaluating the impact of an environmental intervention on park use and physical activity. *Prev Med*. 2009; 48(4):316–320. [PubMed: 19463491]
20. Veitch J, Ball K, Crawford D, Abbott GR, Salmon J. Park Improvements and Park Activity A Natural Experiment. *Am J Prev Med*. 2012; 42(6):616–619. PubMed PMID: ISI: 000304090900011. [PubMed: 22608379]

Figure 1.
Screen shots of the 12-button counter: counting mode (upper) and result showing mode (lower).

Table 1

Means and standard deviations of total METs by gender and age groups in the study sample for inter-rater reliability.

	Total METs		METs in MVPA	
	Female	Male	Female	Male
Child	2.1 (6.3)	3.4 (9.3)	1.1 (4.9)	2.3 (8.1)
Teenager	0.6 (2.9)	1.7 (6.2)	0.3 (1.8)	1.1 (5.3)
Adult	3.0 (7.2)	6.3 (12.9)	1.1 (5.0)	3.2 (9.9)
Senior	0.2 (1.6)	0.3 (1.1)	0.1 (1.4)	0.1 (0.8)
Total	17.6 (27.3)			

Table 2

Inter-rater reliability measures by age and gender categories.

Genders	Age groups	Correlation			ICC		Agreement probability		
		Total METs	METs in MVPA	METs in MVPA	Total METs	METs in MVPA	Total METs	METs in MVPA	METs in MVPA
Female	Child	0.97	0.92	0.97	0.97	0.92	0.90	0.90	0.90
	Teenager	0.97	0.91	0.97	0.97	0.90	0.90	0.90	0.82
	Adult	0.94	0.86	0.93	0.93	0.81	0.91	0.91	0.88
	Senior	0.95	0.94	0.95	0.95	0.93	0.95	0.95	0.91
Male	Child	0.94	0.92	0.94	0.94	0.92	0.85	0.85	0.83
	Teenager	0.88	0.86	0.88	0.88	0.85	0.88	0.88	0.82
	Adult	0.97	0.96	0.97	0.97	0.96	0.90	0.90	0.86
	Senior	0.77	0.70	0.74	0.74	0.64	0.86	0.86	0.85