



Published in final edited form as:

*Med Sci Sports Exerc.* 2015 December ; 47(12): 2558–2562. doi:10.1249/MSS.0000000000000725.

## Visual Inspection for Determining Days when Accelerometer is Worn: Is this Valid?

Eric J. Shiroma<sup>1,2</sup>, Masamitsu Kamada<sup>1,3</sup>, Colby Smith<sup>1</sup>, Tamara B. Harris<sup>2</sup>, and I-Min Lee<sup>1</sup>

<sup>1</sup>Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>2</sup>National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

<sup>3</sup>Department of Health Promotion and Exercise, National Institute of Health and Nutrition, Shinjuku-ku, Tokyo, Japan

### Abstract

**Purpose**—Logs are traditionally used for ascertaining accelerometer wear days in mail study designs but not all participants complete logs. Visual inspection of accelerometer output may supplement missing logs; however, no data are available on the validity.

**Methods**—We compared visual inspection with participant logs in 197 women (mean age = 71.0 years). Women were mailed an accelerometer to be worn during waking hours for 7 days, marking each wear day on a log, before returning the accelerometer by mail. For every participant, we created a series of graphs of accelerometer counts by time of day (1 chart for each day with accelerometer output, including mail days). Two raters, masked to log wear status, independently inspected these graphs and scored each day as “worn” or “not worn”.

**Results**—The median (interquartile range) number of valid wear days using either visual inspection or log was 7 (7, 7). For rater 1, the sensitivity of visual inspection was 99.7% (95% confidence interval: 99.2%, 99.9%) and specificity 97.2% (95.2%, 98.6%); rater 2, 99.7% (99.2%, 99.9%) and 97.0% (94.9%, 98.4%). Inter-rater agreement was 99.5%.

**Conclusion**—Visual inspection of accelerometer data is a valid alternative for missing participant wear logs when determining wear days in mail study designs.

### Keywords

physical activity; epidemiology; wear time; older adults

## INTRODUCTION

Accelerometers are devices that allow for the objective assessment of physical activity, capturing movement in up to 3 axes in real time with minimal participant burden.(13, 15)

CORRESPONDING AUTHOR: Eric J Shiroma, ScD, 7201 Wisconsin Ave, Gateway Bldg, Suite 3C309, Bethesda, MD 20814, P: 301-496-1178, F: 301-496-4006, Eric.shiroma@nih.gov.

Conflict of interest: None to report.

The results of the present study do not constitute endorsement by the American College of Sports Medicine.

One challenge in processing accelerometer data is correctly identifying when the monitor was worn.(6, 7) Traditionally, researchers put accelerometers on participants at a first visit and remove them at a second, or provide them at a first visit with instructions to begin wear the next day for a given number of days before returning the monitor. Thus, the days of wear are known.(2, 8, 12) Using statistical algorithms to identify and eliminate additional non-wear time (e.g. if the monitor was removed during showering or sleep) further reduces bias from incorrectly classified wear time.(3, 4, 12, 14, 15)

As accelerometer use becomes more common in research due to its decreasing cost, direct mail methods may be necessary to study large numbers of participants for logistical reasons. Such a protocol may result in accelerometers recording movements when in the mail (i.e., not true participant physical activity). While automated statistical algorithms exist to determine non-wear time(3, 4, 12), these algorithms were developed based on protocols with a face-to-face encounter between researcher and participant. Thus, solely using these algorithms for direct mail designs can overestimate wear time and bias physical activity measures, and restricting algorithm use only to participant reported wear days is recommended.(6)

However, requiring logs imposes an additional burden on participants, and not all will complete them. For example, in the Women's Health Study, an ancillary investigation using a direct mail design is collecting 7 days of accelerometer-assessed physical activity in ~18,000 women. About 10% of women who wore and returned their accelerometers did not return a wear log, resulting in meaningful missing data if logs are required to classify wear days.(6)

For participants with missing wear logs, visual inspection of the accelerometer output by raters may provide a means to use their data. However, there are no data on whether this is valid. In this study, we aim to compare determination of wear days and accelerometer-assessed physical activity when using visual inspection by raters and participant-reported wear logs.

## METHODS

### Study sample

Participants were drawn from an ancillary study of the Women's Health Study investigating accelerometer-assessed physical activity in relation to health outcomes. The study population and detailed methods have been previously described elsewhere.(7, 11) Briefly, women who agreed to participate and were able to walk outside the home without assistance (inclusion criteria) were mailed an accelerometer (ActiGraph GT3X+) and asked to wear it on their hip for 7 days during waking hours. In addition, women were instructed to indicate on a wear log the days when they wore the monitor (Figure 1). Women then returned the monitor and log by mail. The study was approved by the institutional review board of Brigham and Women's Hospital.

Accelerometer data were screened for periods of wear using a standard algorithm by Choi et al.(3, 4) Briefly, non-wear time was defined as 90 consecutive minutes of zero counts across

all 3 axes, with an allowance of 2-minutes of nonzero counts provided there were 30-minute consecutive zero count windows up and downstream. As of 3 August 2012, 7650 women had returned an accelerometer and a wear log. For the present study, we randomly selected 200 women of these women. We then excluded 3 women who had incomplete logs or did not have at least 1 day of at least 10 hours of wear time (conventional standard for a valid wear day) based on the Choi algorithm. This resulted in a final analysis sample of 197 women (mean age = 71.0 (SD = 5.8) years).

### Identification of wear days

Potentially valid accelerometer wear days were days of at least 10 hours of wear time based on the Choi algorithm. We then classified each of these potential wear days as a valid wear day or non-wear day using 2 separate methods: participant-reported wear logs or visual inspection by rater. The participant-reported wear logs served as the “gold standard”. We used only accelerometer data from days indicated as wear days on the participant log to estimate measures of physical activity.

For the visual inspection method, we created a series of graphs for each participant using the SGPANEL procedure in SAS 9.3, with each graph plotting a day’s accelerometer counts on the vertical axis and the time of the day on the horizontal axis. All accelerometer data, including from days in the mail, were charted (one graph per day up to an observed maximum of 28 days) (Figure 2). Two independent raters (MK and CS) who were masked to the log assessment of wear versus non-wear inspected these graphs to score each panel, which represents one day, as “worn” or “not worn”. Raters were minimally trained: 1) they first reviewed 2 or 3 participants’ data from a separate sample, with logs alongside that indicated which days were “worn” days, and 2) they were aware that participants were instructed to wear the monitor for 7 consecutive days. For this method, valid wear days required both visual inspection by rater and sufficient wear time by the Choi algorithm.

### Statistical analyses

We calculated three classification measures: 1) sensitivity (probability of scoring a day as “worn” by visual inspection of accelerometer output, given this is a wear day based on the participant’s log) for each rater; 2) specificity (probability of scoring a day as “not worn”, given this is a non-wear day based on the participant’s log) for each rater; and 3) percent agreement between the raters.

To compare physical activity measures estimated from wear days identified using visual inspection or wear logs, we calculated the median (25<sup>th</sup> percentile, 75<sup>th</sup> percentile) of the participants’ daily average of wear time; time in sedentary behavior, light-intensity physical activity, moderate-to-vigorous physical activity; and number of steps. Physical activity intensity was classified according to the number of vector magnitude counts per minute: sedentary behavior: <200; light activity: 200 to <2690; and moderate-to-vigorous activity: 2690 counts per minute.(1, 10) We tested for differences between estimates using visual inspection or wear logs with Wilcoxon Signed-Rank tests.

## RESULTS

The average number of days of collected accelerometer data, which included both days the participant wore the monitor as well as days in the mail, was 18.0 (SD = 4.0) days. After applying the Choi algorithm, the mean number of potential wear days was 9.0 (SD = 1.4). When restricting the potential wear days to only those indicated by participants as worn on the wear log, the mean decreased to 6.9 (SD = 0.5) days, consistent with study instructions. Using visual inspection in place of wear logs resulted in a similar average number of valid wear days (mean = 6.9 (SD = 0.6)).

Individual raters spent between 2.5 and 3 hours each to inspect daily graphs of the 197 women. The median (25<sup>th</sup> percentile, 75<sup>th</sup> percentile) number of valid wear days whether identified using visual inspection or participant logs was 7 (7, 7). For rater 1, the sensitivity of identifying wear versus non-wear days using participant-reported wear logs as the gold standard was 99.7% (95% confidence interval (95% CI): 99.2%, 99.9%) and the specificity 97.2% (95.2%, 98.6%). For rater 2, the sensitivity was 99.7% (99.2%, 99.9%) and the specificity was 97.0% (94.9%, 98.4%). The inter-rater percent agreement was 99.5%.

After using the visual inspection protocol to identify valid wear days, we reduced accelerometer data from these days to produce the following summary measures. The median daily wear time was 900.0 minutes; number of vector magnitude counts in thousands was 470.5; minutes/day of sedentary behavior was 493.3; minutes/day of light-intensity physical activity was 368.0; minutes/day of moderate-to-vigorous physical activity was 20.7; and daily number of steps was 5173.6 (Table 1). Summary measures using participant-reported wear days were: median daily wear time, 898.7 minutes; number of vector magnitude counts in thousands, 468.0; minutes/day of sedentary behavior, 493.3; minutes/day of light-intensity physical activity, 368.0; minutes/day of moderate-to-vigorous physical activity, 20.0; and daily number of steps, 5084.8. There were no significant differences ( $P > 0.05$ ) between any of these summary measures estimated from wear days identified by visual inspection or wear logs (Table 1).

## DISCUSSION

To our knowledge, this is the first study to examine if in conjunction with statistical wear time algorithms visual inspection can be used, instead of participant wear logs, to validly classify the days on which an accelerometer is worn in a direct mail study design. We observed that two independent raters, masked to participant wear log status, classified wear versus non-wear days with sensitivity of >99% and specificity >97%, and with inter-rater agreement of >99%. The practical significance of these findings is that in large studies using a direct mail design: (1) accelerometer data among participants without wear logs (~10% in the Women's Health Study(7)) can be salvaged, and (2) only one rater is needed to view accelerometer output to score wear days for participants with missing logs.

Several studies with face-to-face contact with participants have examined how to classify wear time by accelerometer data, wear-logs, or a combination of the two. Peeters et al showed that statistical algorithms examining accelerometer data reliably and accurately

determined wear time compared to wear-logs.(9) Two such algorithms, Choi et al and Troiano et al, have been validated against direct observation and 24-hr wear protocols, removing the need for wear diaries.(3, 4, 12) However, these algorithms have not been adapted for direct mail studies.

Direct mail studies, while logistically most feasible for large-scale studies, have a potential for bias because artifactual movements of the monitor in the mailing process can be improperly included as wear time by conventional algorithms.(3, 4, 12) In the Women's Health Study, we have previously shown use of conventional automated algorithms over-estimate the median number of wear days by 2 (~30 hours of additional wear time), if not restricted to days when participants indicated they wore the monitor.(6) However, participant-reported wear logs are not always available or complete; resulting in missing data. Visual inspection may supplement wear logs to maximally use the data collected, but this had not previously been tested for reliability or validity.

In this study, raters were able to validly classify wear versus non-wear days at a rate of approximately 75 participants per hour. While manually keying the data from the logs took longer (~40 participants/hour), the logs provided more information than simply if the monitor was worn that day, such as time the participant woke up or went to bed, time the monitor was removed during the day, or comments on types of activities engaged in during the day. Thus, logs may still be needed if investigators want such additional information.

The raters were intentionally given minimal instructions by which to classify days as “worn” or “not worn”. In short, they were aware of the instructions provided to the participants requesting the monitor be worn for 7 days during waking hours. Upon the conclusion of the visual inspection of the accelerometer data, rater 1 commented it became apparent that most participants wore the monitor for 7 days and generally had similar waking/bed times across the wear days. The rater utilized this daily pattern to help distinguish wear-days from the preceding mailing (non-wear) day. In addition, rater 1 highlighted a potential flaw of a day-level analysis defined from midnight to midnight. For a few individuals, it was clear they were wearing the monitor past midnight. This may result in loss of the post-midnight wear data if the next day was classified as not worn. However, the potential data loss would only include the last day of wear if the rater identifies a sequential series of wear days. This misclassification is also inherent in available statistical algorithms for determining wear time in non-direct mail studies, as they conventionally use a “midnight to midnight” definition for each day of wear.

This study does have several limitations. We did not have a true gold standard for objectively determining wear time in the participants, such as heart rate or direct observation. We considered participant wear logs to be the gold standard, but participants could have incorrectly logged their wear days. Keadle et al found that while wear logs have missing data (particularly noting the time monitors were put on or taken off), documentation of whether the device was worn or not on a particular day was quite complete ( 2.2% missing data).(6) However, there is little literature examining how well logs are filled out compared to other standards. In addition, visual inspection was performed for each participant for the entire potential wear period using day “blocks” within one graph (Figure

2); thus individual, ordered days were correlated within a participant but analyses did not adjust for this correlation. Furthermore, visual inspection is a subjective assessment method and subject to rater variability. However, we observed inter-rater agreement to be >99%.

Women from the WHS were very compliant with the study protocol, with an average of 6.9 wear days. This results in high sensitivity and specificity, which may not be seen among participants with few wear days. However, even in the general population, compliance is reasonably high among those who agree to participate and wear an accelerometer; thus, the limitation may not be major. For example, among comparably aged women in NHANES who wore the monitor for at least one day, 80% had at least 5 days of valid wear.(12) It is also possible that women who returned a log (and thus were eligible for this study) are different from those who did not. However, when comparing wear time as assessed using the Choi algorithm for women who returned a log versus those who did not, the times were similar. Finally, these data were collected from older, mostly White women with higher socio-economic status than the general population, and who were somewhat more active than nationally-representative samples of similarly aged women,(5, 11) which may limit the generalizability of findings. Future studies that incorporate other technology such as body sensors may remove the need for logs or visual inspection.

In conclusion, in direct mail study designs, visual inspection of accelerometer data by a single rater with minimal training is a valid and reliable alternative to participant wear logs for determining the days on which an accelerometer is worn if participants have missing logs, allowing maximal use of data collected.

## Acknowledgments

Financial support

This research was supported by research grants CA154647 and CA047988 from the National Institutes of Health. EJS and TBH were supported in part by and the Intramural Research Program of the National Institutes of Health, National Institute on Aging.

Thank you

We are grateful to the staff of the Women's Health Study (Brigham and Women's Hospital), particularly Ara Sarkissian, MM; Bonnie Church, BA; and Jane Jones, MEd.

## Reference List

1. Aguilar-Farias N, Brown WJ, Peeters GM. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. *J Sci Med Sport*. 2014; 17(3):293–9. [PubMed: 23932934]
2. Arnardottir NY, Koster A, Van Domelen DR, et al. Objective measurements of daily physical activity patterns and sedentary behaviour in older adults: Age, Gene/Environment Susceptibility-Reykjavik Study. *Age Ageing*. 2013; 42(2):222–9. [PubMed: 23117467]
3. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc*. 2011; 43(2):357–64. [PubMed: 20581716]
4. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. *Med Sci Sports Exerc*. 2012; 44(10):2009–16. [PubMed: 22525772]

5. Hagstromer M, Troiano RP, Sjostrom M, Berrigan D. Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. *Am J Epidemiol.* 2010; 171(10):1055–64. [PubMed: 20406758]
6. Keadle SK, Shiroma EJ, Freedson PS, Lee IM. Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. *BMC Public Health.* 2014; 14:1210. [PubMed: 25421941]
7. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med.* 2014; 48(3):197–201. [PubMed: 24297837]
8. Murphy SL. Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. *Prev Med.* 2009; 48(2):108–14. [PubMed: 19111780]
9. Peeters G, van Gellecum Y, Ryde G, Farias NA, Brown WJ. Is the pain of activity log-books worth the gain in precision when distinguishing wear and non-wear time for tri-axial accelerometers? *J Sci Med Sport.* 2013; 16(6):515–9. [PubMed: 23294696]
10. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport.* 2011; 14(5):411–6. [PubMed: 21616714]
11. Shiroma EJ, Freedson PS, Trost SG, Lee IM. Patterns of accelerometer-assessed sedentary behavior in older women. *JAMA.* 2013; 310(23):2562–3. [PubMed: 24346993]
12. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008; 40(1):181–8. [PubMed: 18091006]
13. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med.* 2014; 48(13):1019–23. [PubMed: 24782483]
14. Tudor-Locke C, Camhi SM, Troiano RP. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003–2006. *Prev Chronic Dis.* 2012; 9:E113. [PubMed: 22698174]
15. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc.* 2005; 37(11 Suppl):S582–8. [PubMed: 16294121]

**Study ID:**

**Activity Monitor Log**

Study ID Barcode

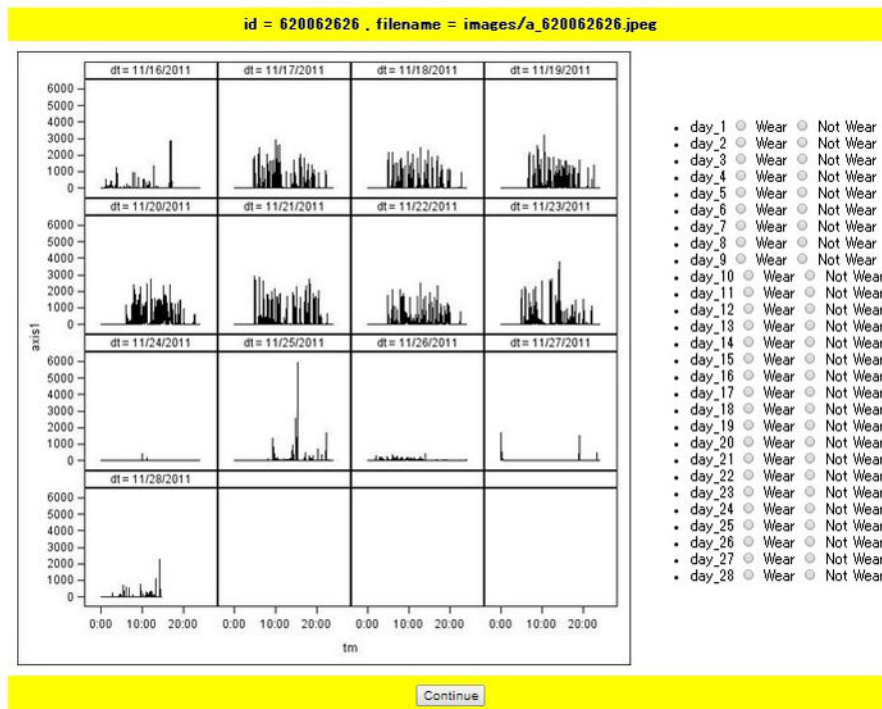
Day	Circle the Day of the Week	Date monitor put on	Time out of bed in the morning	Time monitor put on in the morning	Time monitor removed at night	Time into bed for the night	List times during the day the monitor was not worn and state reason for not wearing it (e.g., showering). Any additional comments?
1	M T W TH F SA SU	__/__/__ mm/dd/yy	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
2	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
3	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
4	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
5	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
6	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____
7	M T W TH F SA SU	__/__/__	__ : __ am pm	__ : __ am pm	__ : __ am pm	__ : __ am pm	_____

Thank you for completing this form. Please return it with the activity monitor.

OFFICE USE ONLY     1    2    3

**Figure 1.** Activity monitor wear log, Women’s Health Study. Participants were instructed to indicate on the log when they wore the monitor.





**Figure 2.** Screenshot of the visual inspection interface for one participant. Graphs display counts plotted against time of day, sequentially for each day. The maximum number of days with accelerometer output was 28 days in this study. For this particular participant, there were 13 days of output data.

**Table 1**

Wear time and physical activity according to visual inspection or wear log protocol in 197 women, Women's Health Study

	Visual Inspection	Wear Log	<i>P</i> Difference
Wear time (min/d)	900.0 (848.3, 950.3)	898.7 (846.9, 947.4)	0.90
Total counts per day (in 1000s)	470.5 (372.5, 578.1)	468.0 (372.5, 574.7)	0.46
Sedentary behavior (min/d)	493.3 (443.4, 574.1)	493.3 (442.3, 574.1)	0.30
Light activity (min/d)	368.0 (312.7, 417.6)	368.0 (312.7, 415.6)	0.43
Moderate-to-vigorous activity (min/d)	20.7 (8.9, 37.4)	20.0 (8.9, 37.4)	0.95
Steps per day	5173.6 (3852.1, 7331.9)	5084.8 (3836.1, 7384.6)	0.86

Estimates were calculated using accelerometer data from wear days, as determined by the particular protocol. Data from wear days, as identified under each protocol, were processed using the Choi wear time algorithm.(3, 4) Estimates are medians (25th percentile, 75th percentile) unless otherwise specified. P-value for difference between visual inspection and wear log was calculated by Wilcoxon Signed-Rank test.