Published in final edited form as:

Am J Prev Med. 2017 April ; 52(4): 541–548. doi:10.1016/j.amepre.2016.11.010.

Reproducibility of Accelerometer-Assessed Physical Activity and Sedentary Time

Sarah Kozey Keadle, PhD, MPH[1,2], **Eric J. Shiroma, ScD**[3], **Masamitsu Kamada, PhD**[4,5], **Charles E. Matthews, PhD**[1], **Tamara Harris, MD**[3], and **I-Min Lee, MBBS, ScD**[4,6]

[1]Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland [2]Cancer Prevention Fellowship Program, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland [3]Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute of Aging, Bethesda, Maryland [4]Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts [5]National Institute of Health and Nutrition, Japan [6]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston Massachusetts

Abstract

Introduction—Accelerometers are used increasingly in large epidemiologic studies, but most studies are restricted to a single, 7-day accelerometer monitoring period given logistic and cost constraints. It is unknown how well a 7-day accelerometer monitoring period estimates longer-term patterns of behavior, which is critical for interpreting, and potentially improving, disease risk estimates in etiologic studies.

Methods—A subset of participants from the Women's Health Study (N=209; mean age, 70.6 [SD=5.7] years) completed at least two 7-day accelerometer administrations (ActiGraph GT3X+) within a period of 2–3 years. Monitor output was translated into total counts, steps, and time spent in sedentary, light-intensity, and moderate to vigorous–intensity activity (MVPA) and bouted-MVPA (i.e., 10-minute bouts). For each metric, intra-class correlations (ICCs) and 95% CIs were calculated using linear-mixed models and adjusted for wear time, age, BMI, and season. The data were collected in 2011–2015 and analyzed in 2015–2016.

Results—The ICCs ranged from 0.67 (95% CI=0.60, 0.73) for bouted-MVPA to 0.82 (95% CI=0.77, 0.85) for total daily counts and were similar across age, BMI, and for less and more active women. For all metrics, classification accuracy within 1 quartile was >90%.

Conclusions—These data provide reassurance that a 7-day accelerometer-assessment protocol provides a reproducible (and practical) measure of physical activity and sedentary time. However, ICCs varied by metric; therefore, future prospective studies of chronic diseases might benefit from existing methods to adjust risk estimates for within-person variability in activity to get a better estimate of the true strength of association.

Address correspondence to: Sarah Kozey Keadle, PhD, MPH, Kinesiology Department, California Polytechnic State University, San Luis Obispo CA 93407. skeadle@calpoly.edu.

No financial disclosures were reported by the authors of this paper.

INTRODUCTION

Physical activity lowers risk for chronic diseases and improves longevity.[1] Much of this evidence comes from prospective epidemiologic studies, which have historically relied on self-report questionnaires designed to estimate habitual physical activity over a relatively long time interval (e.g., past year).[2] Accelerometers provide several advantages over self-report questionnaires, including improved accuracy and precision, particularly for lower-intensity activities.[3] Recently, accelerometer-assessed physical activity has been linked to premature mortality in the National Health and Examination Survey.[4–6] Several other cohort studies have incorporated accelerometer assessments; however, economic and logistic constraints limit data collection protocols typically to a single 7-day period.[3,7,8] It is unknown whether data obtained from a 7-day monitoring protocol is representative of usual activity levels over a longer-time period, which is the primary exposure of interest in epidemiologic studies.

It is expected there will be some behavioral variation in physical activity over time.[9] However, if a 7-day accelerometer-assessment period is not sufficient to estimate longer-term physical activity owing to high amounts of individual behavioral variation, then etiologic studies using these data will underestimate disease risk estimates.[10,11] Furthermore, random within-person behavioral variation is a form of measurement error that can be accounted for in statistical models to improve risk estimates in etiologic studies.[12] Therefore, quantifying how well a single 7-day accelerometer monitoring period estimates longer-term patterns of behavior is a critical component for interpreting, and potentially improving, understanding of physical activity and disease risk. To the authors' knowledge, this has not been investigated previously. The purpose of this paper was to examine the reproducibility of physical activity among older women over 2–3 years.

METHODS

Study Sample

The Women's Health Study is a completed randomized trial (1992–2004) among 39,876 healthy women aged >45 years.[13–15] When the trial ended, 33,682 women (89% of those alive) consented to continue with observational follow-up, reporting on their health habits and medical history annually on questionnaires. During 2011–2015, an ancillary study was conducted to collect 7 days of accelerometer data. In total, 26,679 women responded (90% of invited). Of those who responded, 1,456 indicated they could not walk outside unassisted, making them ineligible, and 6,931 declined, leaving 18,289 willing and eligible women who were sent monitors (69.7% of respondents). There were 581 monitors lost or not returned, leaving a sample of 17,708 women (96.8% of willing and eligible, 66.4% of respondents) who wore and returned their devices. The study was approved by the Brigham and Women's Hospital's IRB committee and all women signed consent to participate.

For the present study, to minimize accelerometer loss and data collection delays, sampling was restricted to those women who had worn and returned an accelerometer and a physical activity questionnaire within 30 days of being mailed a device. Among these women, 283 were randomly selected and invited to repeat the physical activity assessment at two time periods within the next 3 years (approximately 1 and 2 years later, on a rolling basis). The aim was to obtain a sample of 100 women with complete data for the intraclass correlation (ICC) analysis. Because it was unclear what the response rate would be at the beginning of the study and the Women's Health Study data were collected on a rolling basis,[3] it was estimated that 40 monitors per season over the data collection period would ensure a representative and sufficient sample.

A flow chart of participant eligibility and enrollment is shown in Appendix Figure1. Women were invited to complete the third administration regardless of their response to the second administration. For the primary analysis, all women who wore and returned the monitor at least twice (i.e., baseline and one of the additional administrations; N=209, 93% of those willing and eligible) were included. A sensitivity analyses was conducted that restricted to those with valid data at all three time points (baseline, Time 2, and Time 3; N=129, 58% of those willing and eligible).

Accelerometer Protocol

The overview of data processing and cleaning procedures for the ancillary study has been reported previously.[16] Briefly, women were mailed an ActiGraph GT3X+ (ActiGraph, Inc., Pensacola, FL) accelerometer and asked to wear the monitor, secured with an adjustable belt on the hip, for 7-days during waking hours and return the monitor and wear-log by mail. To be included in the analysis, women had to have valid data on at least 4 of the 7 days of wear, as is conventional.[16] Accelerometer data were expressed as counts per minute (cpm). The Choi algorithm[17,18] plus dates from the wear-log were used to identify valid days of wear (≥600 minutes per day).[16] The files were summarized into total daily vector magnitude (VM) counts (an index of total physical activity), steps per day, and time in sedentary (0–199 cpm),[19] light (200–2,690 cpm), and moderate to vigorous (≥2,691 cpm) physical activity (MVPA).[20] Bouted-MVPA was defined as 10 consecutive minutes above the 2,691-cpm threshold, allowing for up to 2 minutes below the threshold.[7] Accelerometer data processing was done using R, version 3.3.1 (www.R-project.com) and the Choi algorithm was implemented using the PhysicalActivity package (https://cran.r-project.org/web/packages/PhysicalActivity/PhysicalActivity.pdf).

Measures

As part of the parent study, women completed annual questionnaires on sociodemographic characteristics, lifestyle habits, and medical history. Participant characteristics for women in the present analysis were compared to all women in the accelerometer study using questionnaire data from baseline of the accelerometer study (i.e., 2011).

Statistical Analysis

Times spent sedentary, light-intensity activity, steps, and VM counts were approximately normally distributed and the ICC values with natural log transformation were virtually unchanged, thus untransformed values were used for all analyses. Time spent in MVPA and MVPA-bouts were not normally distributed and natural log transformed values were used in analyses. For each activity metric, the weekly average was calculated and Spearman correlations between baseline and the second and third administrations were estimated.[21]

Repeated measures linear-mixed models were used to estimate whether the accelerometer metric changed over time, adjusting for age and monitor wear time.

The ICCs were used to estimate the random measurement error for each accelerometer metric. Between-person and within-period variances were estimated using random effects models that included average weekly value for each metric at each of the three administrations. To calculate ICCs, the between-person variance was divided by the sum of within- and between-person variances. ICCs and 95% CIs were calculated using the macro described by Hankinson et al.[22] (http://www.hsph.harvard.edu/donna-spiegelman/software/icc9/). Generally, ICC values ≥0.75 are considered excellent; 0.4–0.75, fair to good; and <0.4, poor reproducibility.[23] ICCs are presented with and without adjustment for age, BMI, season of the year, and monitor wear time. ICCs were stratified by age (median, 70.3 years), BMI (normal weight [<25 kg/m^2] versus overweight and obese [≥25 kg/m^2]) using questionnaire data from baseline of the accelerometer study, and activity status based on median VM counts.

To assess the utility of a single 7-day administration for classifying longer-term activity, women were cross-classified according to the quartile distribution from the baseline administration and the quartile distribution of the average of the second and third administrations (i.e., longer-term behavior, as has been done previously with blood biomarkers).[21] There were 80 participants who did not have both Time 2 and Time 3; for those individuals, a single timepoint was used. The quartile cut points were defined separately for each period, and thus were not identical at baseline and Times 2 and 3. A high proportion of women did not have any bouted-MVPA, so tertiles were examined to ensure approximately equal distribution across groups in analyses of this metric.

To examine the impact of different sample size at the three time points, sensitivity analyses were conducted for those with complete data at all three timepoints (n=129), those with baseline and Time 2 only (n=172), and those with baseline and Time 3 only (n=166) using VM counts, a metric of overall activity. In the primary analysis, quartiles were recalculated at each timepoint. As a sensitivity analysis, percentage agreement was examined when the baseline quartile cut points for VM counts were applied to the later timepoints. SAS, version 9.3 was used for all analyses.

RESULTS

The demographic characteristics for women in the present study (N=209) were similar to those for the total sample from the accelerometer study (N=17,708) (Appendix Table1). The median (25th, 75th percentile) number of days between baseline and Time 2 was 392 (332, 432), and between baseline and Time 3 was 838 (767, 1,004) days (Table 1). The median values for each accelerometer metric are shown in Table 1. For total VM counts, the Spearman correlation between baseline and the average of Times 2 and 3 was 0.83; baseline and Time 2, 0.81; and baseline and Time 3, 0.81. There were modest, but statistically significant declines in total activity over time (2.4% and 3.7% lower at Times 2 and 3, respectively). Steps, MVPA, and bouted-MVPA also declined over time, whereas sedentary time and light intensity activity did not (Table 1).

The adjusted ICCs (95% CIs) ranged from 0.67 (0.60, 0.73) for bouted-MVPA to 0.82 (0.77, 0.85) for total daily VM counts (Table 2). ICC values were similar with and without adjustment for age, BMI, season, and wear time (Table 2). In the stratified analysis, the ICCs were similar for all metrics between younger and older, and normal weight and overweight/obese women (Table 3). The ICCs for MVPA and bouted-MVPA were lower for less active (0.51 and 0.45, respectively), compared with more active (0.73 and 0.64, respectively) (Table 3). In the less active group, the median (interquartile range) for bouted-MVPA was 0 (0, 4.5) minutes. The ICCs for sedentary time were higher for less active than more active (0.75 vs 0.64) women (Table 3).

To compare consistency of activity classification, women were cross-classified by quartile distributions at baseline compared to Times 2 and 3 (Table 4). For total VM counts, an index of total physical activity, there were 117/209 women (56.0%) who were classified in the same quartile at both timepoints, an additional 81/209 women (38.8%) were within 1 quartile, and 11/209 (5.3%) of women were misclassified by ≥2 quartiles (Table 4). For sedentary time, 50.2% were classified in the same quartile at both timepoints, an additional 42.6% were within 1 quartile, and 7.2% were misclassified by ≥2 quartiles (Table 4). The percentage of women who were classified within 1 quartile was 95.2% for light-intensity activity, 93.3% for MVPA (Table 4), and 93.3% for steps (Appendix Table 2). Bouted-MVPA was examined in tertiles, and the percentage classified in the same tertile at both timepoints was 64.1% (Appendix Table 2).

A sensitivity analysis using VM counts was conducted to examine the impact of the different sample sizes at Times 2 and 3 on the ICCs and quartile classification. Appendix Table 3 shows slightly higher ICC values for the 129 participants with complete data at all three timepoints (Appendix Figure 1 shows flow chart of enrollment), compared with the 209 used in the main analysis who had data at baseline and Time 2, Time 3, or both. The percentage of women classified in the same quartile using VM counts for the full sample was 56%, whereas slightly more women were consistently classified (62%) for the 129 with complete data at all three timepoints (Appendix Table 4). When comparing baseline to Time 2 alone, 60% were consistently classified, and for baseline and Time 3 alone the value was 52% (Appendix Table 4). When separate cut points were used at each timepoint, 56% were classified in the same quartile at both timepoints and 94.7% were within 1 quartile. Similarly, when the baseline VM count quartile cut point was applied to Times 2 and 3, a total of 57.4% were classified in the same quartile at both timepoints and 95.2% were within 1 quartile (Appendix Table 5).

DISCUSSION

This study examined the reliability of accelerometer-assessed physical activity and sedentary time over 2–3 years among older women. The ICCs were good to excellent and were similar among older and younger, normal weight and overweight/obese women. There was some indication that the ICCs were lower among less active individuals, particularly for MVPA, likely due to the high number of zeroes in this group (i.e., approximately 25% had <1 minute of MVPA). For all activity-related metrics, quartile rankings were fairly well preserved, with 90% of women classified in the same quartile or within 1 quartile over time. Collectively, these results suggest that a single 7-day accelerometer administration reasonably estimates longer-term physical activity and sedentary time.

The ICCs in the present study, using a single 7-day accelerometer administration, were similar to or higher than reproducibility for self-report questionnaires that assess longer-term activity. For total activity (VM counts), the ICCs were 0.82 for three administrations 2–3 years apart (Table 2). Reproducibility of past-year physical activity has been examined in studies that administered questionnaires twice, approximately 1 year apart. The reported ICCs for total activity using self-reported questionnaires was 0.69 among older women,[24] 0.70 among Chinese women,[25] 0.82 among college-educated women,[26] and 0.69 for men and women.[27] Two previous studies included a longer follow-up similar to the present design and reported r=0.59[28] for total activity and ICC=0.52[29] for vigorous activity. Reliability of household and lower-intensity activities as assessed by questionnaire tends to be lower than total activity.[24–26,29,30] In addition, the ICC values observed in the present study (0.67–0.82) are similar to those for other important biological measures including cholesterol, blood pressure, glucose, and C-reactive protein over 4 years (0.43–0.81)[31] and plasma steroid hormones and insulin-like growth factor over 2–3 years (0.22–0.94).[21]

A commonly used method in epidemiologic studies is to rank individuals by ordinal groups (e.g., quartiles) of the exposure of interest. To examine how well a single accelerometer administration categorizes relative levels of physical activity over time, the authors compared baseline values to the average of Times 2 and 3, and found that approximately 50%–56% of women were classified in the same relative quartile and 90% of women were classified within 1 quartile. Few previous studies have assessed the reliability of physical activity as a categorical variable. Jurj and colleagues[32] reported 66% were classified in the same quartile for self-reported total activity over a 1-year period, which is similar to the 61% over 1 year in the present study (Appendix Table 4), and 56% over the 3-year period (Table 4). The percentage agreement for activity metrics observed in the present study was similar to other epidemiologic exposures of interest, such as testosterone (56%) and estradiol (41%).[21,22] The primary analysis evaluated whether the quartiles rankings were preserved over time, thus the present results may not reflect classification agreement on an absolute basis. However, differences in the quartile cut points for the metrics did not vary greatly over time, and the sensitivity analysis using baseline quartile cut points at baseline and at Times 2 and 3 revealed similar results (Appendix Table 5), suggesting that classification agreement for the same cut points over time would also be fairly reproducible.

Although the present results indicate activity levels among older adult women are fairly stable over time, even modest amounts of within-person random error will result in attenuation of true relative risk and an underestimation of the actual benefits of physical activity or harms associated with sedentary time.[12] Methods have been developed to correct relative risk estimates for random within-person variation and could be applied in future studies.[9] To provide some insight into the magnitude of attenuation of risk estimates for even modest amounts behavioral variability estimate the degree of attenuation was estimated by multiplying the ICC by the natural log of the true relative risk and exponentiated.[12] For a true relative risk of 2.5, a metric with ICC=0.82 would be observed (with error) as a relative risk of 2.1, whereas a metric with an ICC=0.67 would have an observed relative risk of only

1.8. Future prospective studies of physical activity and disease or mortality should consider correcting relative risk estimates even for modest amounts of random error due to behavioral variation, and the current results can inform such efforts.[12] It should also be noted that activity declined slightly during the 3-year period for the sample as a whole; however, the approximately equal proportion of individuals classified in higher versus lower quartiles over time supports that it was likely largely attributable random to random measurement error due to within-person variability.

Limitations

Strengths of this study include measurement over a period of 2–3 years in a large sample of women. Generalizability is the primary limitation; this sample only includes older women, the majority of whom are highly educated and white. The women were also very compliant, with an average wear time of 15 hours per day. Almost all women wore the monitor for least 6 days (97.6% baseline, 97.1% Time 2, and 95.2% Time 3). This analysis assumed, according to usual convention, that 4 or more valid days of wear are sufficient to achieve an adequate estimate of activity within a single monitoring period.[33–35] The number of days needed for different activity metrics has been extensively studied and some have suggested a higher number of consecutive days are needed to provide a reliable estimate of activity,[36] and the number of days required is highly dependent on the target "goal" that is set by investigators (e.g., ICC=0.8 or 0.9).[36] These results investigate the reproducibility of 4–7 days' worth of information from a single monitor administration over several years. Although the authors considered season of the year in the analysis, the study design did not capture each season for each participant, so the adjustment for the average seasonal effects may not be complete and that residual effects of season are reflected in the ICCs. The present analyses do not address other forms of systematic or random error that may result from processing accelerometer data into activity metrics. Measurement error correction models that address these other sources of error are also needed in studies of physical activity and disease.[37–39]

CONCLUSIONS

The present results suggest that a 7-day accelerometer assessment protocol provides a reproducible (and practical) measure of physical activity and sedentary time among older women. These data can inform interpretation of past and future studies investigating associations between physical activity and disease risk.

ACKNOWLEDGMENTS

We are grateful to the staff and participants of the Women's Health Study (Brigham and Women's Hospital).

This research was supported by research grants CA154647 and CA047988 from NIH and the Intramural Research Program at the National Cancer Institute and National Institute of Aging.

Appendix

Appendix Table 1

Descriptive Comparison of Women in the Repeat-Assessment Subset to All Women in the Full Accelerometer Sample

	Repeat subset N=209		Full accelerometer sample N=17,708	
Characteristics	**Mean**	**SD**	**Mean**	**SD**
Age (yrs)	70.6	5.7	70.3	5.7
BMI (kg/m^2)	25.7	4.8	26.3	5.0
	Median	**IQR**	**Median**	**IQR**
Days of wear	7	(7, 7)	7	(7, 7)
Wear time, min	904.4	(861.6, 946)	894.4	(846.0, 937.9)
Steps[a]	5,423	(3,901, 7,038)	5,070	(3,606, 6,917)
Sedentary time, min[b]	509.3	(455.1, 579.7)	502.7	(434.3, 570.8)
Light activity time, min[c]	351.9	(296.6, 412.4)	350.7	(292.0, 410.6)
MVPA time, min[d]	21.2	(8.9, 46.3)	27.7	(12.7, 49.4)
	N	**Percent (%)**	**N**	**Percent (%)**
Education (college degree)	95	45.9	8,745	50.2
Race (white)	202	97.1	16,863	96.0
Health status (very good or excellent)	170	84.2	13,457	78.2
Smoking status (current)	10	4.8	602	3.4

Notes: Demographic variables are self-reported from 2011 questionnaire. Values are daily averages based on the number of days of wear.

MVPA, moderate to vigorous physical activity; VM, vector magnitude; cpm, counts per minute

[a]Summation of vector magnitude counts per minute and step count over the daily wear period.

[b]Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200[14]

[c]Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2,690.[15]

[d]MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691,[15] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold.

Appendix Table 2

Cross-Classification Matrices for Steps and Bouted-MVPA at Baseline and the Average of Times 2 and 3

	Quartiles for mean of Times 2 and 3					
Quartiles at baseline	**1**	**2**	**3**	**4**	**Total**	**Percent agreement**
Steps						
1	**36**	12	4	0	52	
2	10	**23**	16	3	52	
3	5	16	**20**	12	53	
4	1	1	13	**37**	52	
Total	52	52	53	52	209	**55.5%**

	Tertiles for mean of Times 2 and 3				
Tertiles at baseline	**1**	**2**	**3**	**Total**	**Percent agreement**
Bouted-MVPA time					
1	**52**	22	6	80	
2	15	**31**	13	59	
3	2	17	**51**	70	
Total	69	70	70	209	**64.1%**

Notes: Bouted-MVPA expressed at tertiles to ensure approximately equal distribution due to high amounts of 0's. Values are daily averages based on the number of days of wear. Bolded values indicate classified in the same quartile at both-time.

MVPA, moderate to vigorous physical activity; VM, vector magnitude; cpm, counts per minute

[a]Summation of step count over the daily wear period. Baseline: Q1 <3,900.9; Q2: 3,900.9 to <5,423.42 Q3: 5,423.43 to 7,052 Q4: ≥7,052. Time 2/3: Q1 <3,728.4; Q2: 3,728.4 to <4,939 Q3: 4,939 to 6,785 Q4: ≥6,785.

[b]MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm ≥2,691,[15] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold. Baseline: Q1 <0.3 min; Q2: 0.3 to <14 min; Q3: ≥14 min. Time 2/3: Q1 <0.6 min; Q2: 0.6 to <11.7 min; Q3: ≥11.7 min.

Appendix Table 3

Intraclass Correlation (ICC) Values for Different Subsets Based on Missing Data

Variables	All time points[a]	Subset with no missing data[b]	Subset with baseline and Time 2 data	Subset with baseline and Time 3 data
N	209	129	172	166
VM counts[c]	0.82 (0.77, 0.85)	0.84 (0.79, 0.88)	0.83 (0.78, 0.88)	0.85 (0.80, 0.89)
Steps[c]	0.78 (0.73, 0.83)	0.81 (0.75, 0.85)	0.85 (0.80, 0.89)	0.79 (0.73, 0.85)
Sedentary time, min[d]	0.75 (0.69, 0.80)	0.74 (0.67, 0.80)	0.73 (0.66, 0.80)	0.75 (0.67, 0.82)
Light activity time, min[e]	0.70 (0.64, 0.76)	0.69 (0.61, 0.76)	0.68 (0.60, 0.76)	0.70 (0.61, 0.78)
MVPA time, min[b]	0.70 (0.64, 0.76)	0.73 (0.66, 0.79)	0.74 (0.67, 0.81)	0.70 (0.61, 0.78)
Bouted-MVPA time, min[f]	0.67 (0.60, 0.73)	0.73 (0.66, 0.79)	0.78 (0.71, 0.83)	0.72 (0.63, 0.79)

Notes: ICCs are adjusted for age, BMI, monitor wear time, and season when the data were collected. Values are daily averages based on the number of days of wear.

MVPA, moderate to vigorous physical activity; VM, vector magnitude; cpm, counts per minute

[a]Have data at baseline and either Time 2, Time 3, or both.

[b]Have complete data at baseline, Time 2, and Time 3

[c]Summation of vector magnitude counts per minute and step count over the daily wear period.

[d]Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200[19]

[e]Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2,690[20]

[f]MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691,[20] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold.

Appendix Table 4

Cross-Classification Matrix by Quartile for Total VM Counts at Baseline and the Longer-Term Estimates for Sub-Sets Based on Missing Data

	Quartiles for mean of Times 2 and 3					
Quartiles at baseline	**1**	**2**	**3**	**4**	**Total**	**Percent agreement**
Full sample (n=209)						
1	**35**	*14*	2	1	52	
2	*14*	**23**	*12*	3	52	
3	3	*13*	**24**	*13*	53	
4	0	2	*15*	**35**	52	
Total	52	52	53	52	209	**56%**
Subset with both complete data for both Time 2 and Time 3 (n=129)						
1	**25**	*6*	1	0	32	
2	*5*	**17**	*9*	1	32	
3	2	*8*	**15**	*8*	33	
4	0	1	*8*	**23**	32	
Total	32	32	33	32	129	**62%**
	Quartiles for Time 2 only					
Baseline and Year 2 (n=172)						
1	**32**	*8*	2	2	44	
2	*9*	**18**	*10*	2	39	
3	2	*13*	**23**	*8*	46	
4	0	4	*8*	**31**	43	
Total	43	43	43	43	172	**60%**
	Quartile for Time 3 only					
Baseline and Year 3 (n=166)						
1	**26**	*11*	2	0	39	
2	*10*	**18**	*12*	1	41	
3	5	*11*	**13**	*11*	40	
4	0	2	*15*	**29**	46	
Total	41	42	42	41	166	**52%**

Notes: Total VM counts is the summation of vector magnitude counts per minute over the daily wear period. Values are daily averages based on the number of days of wear. Bolded values indicate classified in the same quartile at both-time points and italicized are within one quartile.

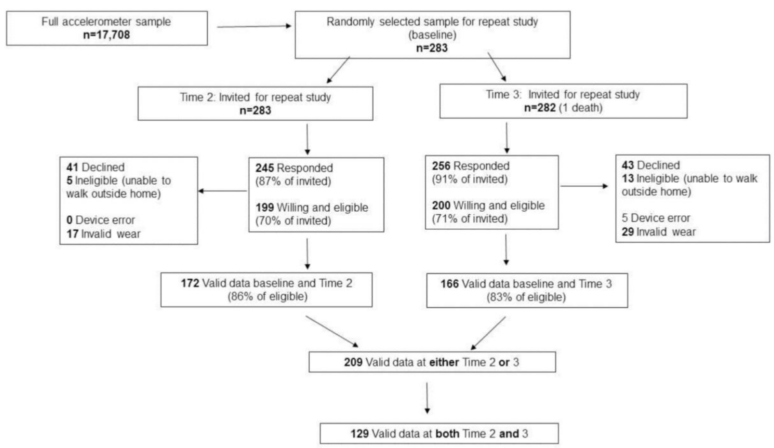
VM, vector magnitude

Appendix Table5

Comparison of Using Same Cut-Point at Baseline Versus Using Different Cut-Points Over Time to Define Quintiles

		Quartiles for mean of Times 2 and 3					
Quartile at baseline	**Values**	**1**	**2**	**3**	**4**	**Total**	**Percent agreement**
Cut-points computed separately at each time-point for vector magnitude counts/1,000							
Values		<362.5	362.5-449.3	449.3-557.8	>557.8		
1	<356.6	**35**	14	2	1	52	
2	356.6-480.9	14	**23**	12	3	52	
3	480.9-581.9	3	13	**24**	13	53	
4	>581.9	0	2	15	**35**	52	
Total		52	52	53	52	209	**56%**
Cut-points from T1 as applied to Times 2 and 3 for vector magnitude counts/1,000							
Values		<356.6	356.6-480.9	480.9-581.9	>581.9		
1	<356.6	**34**	12	3	0	49	
2	356.6-480.9	16	**32**	22	4	74	
3	480.9-581.9	1	7	**20**	14	42	
4	>581.9	1	1	8	**34**	44	
Total		52	52	53	52	209	**57.4%**

Notes: Bolded values indicate classified in the same quartile at both-time points.



Appendix Figure 1.

Flow-chart of participant eligibility and enrollment.

REFERENCES

1. Physical Activity Guidelines Committee Report. U.S. DHHS; Washington, DC: 2008. www.health.gov/paguidelines/report

2. Haskell WL. Physical Activity by Self-Report: A Brief History and Future Issues. J Phys Act Health. 2012; 9:S5–S10. [PubMed: 22287448]

3. 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. http://dx.doi.org/10.1136/bjsports-2013-093154. [PubMed: 24297837]

4. Fishman EI, Steeves JA, Zipunnikov V, et al. Association between Objectively Measured Physical Activity and Mortality in NHANES. Med Sci Sports Exerc. 2016; 48(7):1303–1311. http://dx.doi.org/10.1249/MSS.0000000000000885. [PubMed: 26848889]

5. Loprinzi, PD. Light-Intensity Physical Activity and All-Cause Mortality. Am J Health Promot. In press. Online January 5, 2016. http://dx.doi.org/10.4278/ajhp.150515-ARB-882

6. Schmid D, Ricci C, Leitzmann MF. Associations of Objectively Assessed Physical Activity and Sedentary Time with All-Cause Mortality in U.S. Adults: The NHANES Study. PLoS One. 2015; 10(3):e0119591. http://dx.doi.org/10.1371/journal.pone.0119591. [PubMed: 25768112]

7. 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):181–188. http://dx.doi.org/10.1249/mss.0b013e31815a51b3. [PubMed: 18091006]

8. Hooker SP, Hutto B, Zhu W, et al. Accelerometer measured sedentary behavior and physical activity in white and black adults: The REGARDS study. J Sci Med Sport. 2015; 19(4):336–341. http://dx.doi.org/10.1016/j.jsams.2015.04.006. [PubMed: 25937313]

9. Kelly P, Fitzsimons C, Baker G. Should we reframe how we think about physical activity and sedentary behaviour measurement? Validity and reliability reconsidered. Int J Behav Nutr Phys Act. 2016; 13:32. http://dx.doi.org/10.1186/s12966-016-0351-4. [PubMed: 26931142]

10. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. BMJ. 2010; 340:c2289. http://dx.doi.org/10.1136/bmj.c2289. [PubMed: 20573762]

11. Knuiman MW, Divitini ML, Buzas JS, Fitzgerald PEB. Adjustment for regression dilution in epidemiological regression analyses. Ann Epidemiol. 1998; 8(1):56–63. http://dx.doi.org/10.1016/S1047-2797(97)00107-5. [PubMed: 9465995]

12. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol. 1992; 136(11):1400–1413. [PubMed: 1488967]

13. Cook NR, Lee IM, Gaziano JM, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. JAMA. 2005; 294(1):47–55. http://dx.doi.org/10.1001/jama.294.1.47. [PubMed: 15998890]

14. Lee IM, Cook NR, Gaziano JM, et al. Vitamin E in the primary prevention of cardiovascular disease and cancer - The Women's Health Study: A randomized controlled trial. JAMA. 2005; 294(1):56–65. http://dx.doi.org/10.1001/jama.294.1.56. [PubMed: 15998891]

15. Ridker PM. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in 39,876 women: The Women's Health Study. Am Heart J. 2005; 149(6):1010. http://dx.doi.org/10.1016/j.ahj.2005.04.001.

16. 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. http://dx.doi.org/10.1186/1471-2458-14-1210. [PubMed: 25421941]

17. Choi L, Liu ZW, Matthews CE, Buchowski MS. Validation of Accelerometer Wear and Nonwear Time Classification Algorithm. Med Sci Sports Exerc. 2011; 43(2):357–364. http://dx.doi.org/10.1249/MSS.0b013e3181ed61a3. [PubMed: 20581716]

18. 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):2009–2016. http://dx.doi.org/10.1249/MSS.0b013e318258cb36. [PubMed: 22525772]

19. 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. 2013; 17(3):293–299. http://dx.doi.org/10.1016/j.jsams.2013.07.002. [PubMed: 23932934]

20. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. 2011; 14(5):411–416. http://dx.doi.org/10.1016/j.jsams.2011.04.003. [PubMed: 21616714]

21. Missmer SA, Spiegelman D, Bertone-Johnson ER, Barbieri RL, Pollak MN, Hankinson SE. Reproducibility of plasma steroid hormones, prolactin, and insulin-like growth factor levels among

premenopausal women over a 2-to 3-year period. Cancer Epidemiol Biomarkers Prev. 2006; 15(5): 972–978. http://dx.doi.org/10.1158/1055-9965.EPI-05-0848. [PubMed: 16702379]

22. Hankinson SE, Manson JE, Spiegelman D, Willett WC, Longcope C, Speizer FE. Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. Cancer Epidemiol Biomarkers Prev. 1995; 4(6):649–654. [PubMed: 8547832]
23. Rosner, B. Fundamentals of biostatistics. 5th ed. Duxbury Press; Belmont, CA: 2005.
24. Orsini N, Bellocco R, Bottai M, Pagano M, Wolk A. Reproducibility of the past year and historical self-administered total physical activity questionnaire among older women. Eur J Epidemiol. 2007; 22(6):363–368. http://dx.doi.org/10.1007/s10654-006-9102-1. [PubMed: 17333471]
25. Matthews CE, Shu XO, Yang G, et al. Reproducibility and validity of the Shanghai Women's Health Study physical activity questionnaire. Am J Epidemiol. 2003; 158(11):1114–1122. http://dx.doi.org/10.1093/aje/kwg255. [PubMed: 14630608]
26. Chasan-Taber L, Erickson JB, McBride JW, Nasca PC, Chasan-Taber S, Freedson PS. Reproducibility of a self-administered lifetime physical activity questionnaire among female college alumnae. Am J Epidemiol. 2002; 155(3):282–289. http://dx.doi.org/10.1093/aje/155.3.282. [PubMed: 11821254]
27. Pols MA, Peeters PH, Ocke MC, Slimani N, Bueno-de-Mesquita HB, Collette HJ. Estimation of reproducibility and relative validity of the questions included in the EPIC Physical Activity Questionnaire. Int J Epidemiol. 1997; 26(Suppl 1):S181–189. [PubMed: 9126546]
28. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol. 1994; 23(5):991–999. http://dx.doi.org/10.1093/ije/23.5.991. [PubMed: 7860180]
29. Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. Epidemiology. 1996; 7(1):81–86. http://dx.doi.org/10.1097/00001648-199601000-00014. [PubMed: 8664406]
30. Friedenreich CM, Courneya KS, Neilson HK, et al. Reliability and validity of the Past Year Total Physical Activity Questionnaire. Am J Epidemiol. 2006; 163(10):959–970. http://dx.doi.org/10.1093/aje/kwj112. [PubMed: 16524954]
31. Jackson SE, van Jaarsveld CH, Beeken RJ, Gunter MJ, Steptoe A, Wardle J. Four-year stability of anthropometric and cardio-metabolic parameters in a prospective cohort of older adults. Biomark Med. 2015; 9(2):109–122. http://dx.doi.org/10.2217/bmm.14.78. [PubMed: 25689899]
32. Jurj AL, Wen W, Xiang YB, et al. Reproducibility and validity of the Shanghai Men's Health Study physical activity questionnaire. Am J Epidemiol. 2007; 165(10):1124–1133. http://dx.doi.org/10.1093/aje/kwk119. [PubMed: 17351289]
33. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. Br J Sports Med. 2014; 48(3): 278–282. http://dx.doi.org/10.1136/bjsports-2012-091410. [PubMed: 22936409]
34. 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. http://dx.doi.org/10.1186/1479-5868-8-62. [PubMed: 21679426]
35. Baranowski T, Masse LC, Ragan B, Welk G. How many days was that? We're still not sure, but we're asking the question better! Med Sci Sports Exerc. 2008; 40(7 suppl):S544–S549. http://dx.doi.org/10.1249/MSS.0b013e31817c6651. [PubMed: 18562972]
36. Togo F, Watanabe E, Park H, et al. How many days of pedometer use predict the annual activity of the elderly reliably? Med Sci Sports Exerc. 2008; 40(6):1058–1064. http://dx.doi.org/10.1249/MSS.0b013e318167469a. [PubMed: 18461001]
37. Tooze JA, Troiano RP, Carroll RJ, Moshfegh AJ, Freedman LS. A Measurement Error Model for Physical Activity Level as Measured by a Questionnaire With Application to the 1999-2006 NHANES Questionnaire. Am J Epidemiol. 2013; 177(11):1199–1208. http://dx.doi.org/10.1093/aje/kws379. [PubMed: 23595007]
38. Lim S, Wyker B, Bartley K, Eisenhower D. Measurement Error of Self-Reported Physical Activity Levels in New York City: Assessment and Correction. Am J Epidemiol. 2015; 181(9):648–655. http://dx.doi.org/10.1093/aje/kwu470. [PubMed: 25855646]

39. Arem H, Keadle SK, Matthews CE. Invited commentary: Meta-physical activity and the search for the truth. Am J Epidemiol. 2015; 181(9):656–658. http://dx.doi.org/10.1093/aje/kwu472. [PubMed: 25855648]

Table 1

Descriptive Statistics for Physical Activity and Sedentary Variables Among Older Women 2-3 Years

	Baseline N=209		Time 2 N=172		Time 3 N=166		
Variable	**Median**	**IQR**	**Median**	**IQR**	**Median**	**IQR**	***p*-value**
Days from baseline	-	-	392	(332, 432)	838	(767, 1,004)	-
Days of wear	7	(7, 7)	7	(7, 7)	7	(7, 7)	-
Wear time, min	904.4	(861.6, 946)	877.6	(829.0, 922.0)	886.0	(825.3, 940.0)	**<0.01**
VM counts[a]	483,187	(358,112, 581,932)	466,044	(362,703, 570,682)	441,423	(350,357, 580,432)	**<0.01**
Steps[a]	5,423	(3,901, 7,038)	5,168	(3,678, 6,897)	4,921	(3,775, 6,613)	**<0.01**
Sedentary time, min[b]	509.3	(455.1, 579.7)	494.6	(444.4, 557.7)	516.3	(437.9, 574.1)	0.09
Light activity time, min[c]	351.9	(296.6, 412.4)	351.9	(288.6, 404.5)	342.4	(287.4, 401.7)	0.60
MVPA time, min[d]	21.2	(8.9, 46.3)	22.1	(7.6, 40.2)	18.8	(7.6, 37.4)	**0.03**
Bouted-MVPA time, min[d]	4.5	(0, 22.3)	4.0	(0, 19.5)	3.0	(0, 17.1)	**0.03**

Notes: Boldface indicates statistical significance (p<0.05. p-values are from repeated measures linear mixed models that adjust for age and wear-time to test the differences in each metric over time. Values are daily averages based on the number of days of wear.

IQR, interquartile range; VM, vector magnitude; MVPA, moderate to vigorous physical activity; cpm, counts per minute

[a]Summation of vector magnitude counts per minute and step count over the daily wear period.

[b]Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200[19]

[c]Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2,690.[20]

[d]MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691,[20] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold.

Table 2

Intraclass Correlation (ICC) Values Among Older Women Over a 2-3 Year Period Among 209 Women

Variable	Model 1	Model 2	Model 3
VM counts[a]	0.83 (0.79, 0.87)	0.81 (0.77, 0.85)	0.82 (0.77, 0.85)
Steps[a]	0.82 (0.77, 0.85)	0.79 (0.74, 0.83)	0.78 (0.73, 0.83)
Sedentary time, min[b]	0.73 (0.68, 0.78)	0.71 (0.65, 0.77)	0.75 (0.69, 0.80)
Light activity time, min[c]	0.75 (0.69, 0.80)	0.69 (0.63, 0.75)	0.70 (0.64, 0.76)
MVPA time, min[d]	0.73 (0.67, 0.78)	0.70 (0.64, 0.76)	0.70 (0.64, 0.76)
Bouted-MVPA time, min[d]	0.69 (0.63, 0.75)	0.67 (0.60, 0.73)	0.67 (0.60, 0.73)

Notes: Model 1: unadjusted; Model 2: ICCs include adjustment for age, BMI, and season when the data were collected; Model 3: Adjusted for covariates in Model 2 and further adjusted for monitor wear-time. ICCs are estimated using linear mixed models with all three time points entered in the model.

VM, vector magnitude; MVPA, moderate to vigorous physical activity; cpm, counts per minute

[a] Summation of VM cpm and step count over the daily wear period.

[b] Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200[19]

[c] Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2,690.[20]

[d] MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691,[20] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold.

Table 3

Intraclass Correlation Coefficients (ICC) for Activity Metrics Stratified by Age, BMI, and Activity Level Over 2-3 Years

Variables	Younger (N=106)	Older (N=103)	Normal weight (N=110)	Overweight/ obese (N=99)	Less active (N=94)	More active (N=115)
VM counts[a]	0.81 (0.74, 0.86)	0.82 (0.76, 0.87)	0.82 (0.76, 0.87)	0.81 (0.74, 0.86)	0.64 (0.53, 0.73)	0.74 (0.66, 0.81)
Steps[a]	0.78 (0.71, 0.84)	0.78 (0.71, 0.84)	0.78 (0.71, 0.84)	0.78 (0.71, 0.84)	0.58 (0.46, 0.68)	0.73 (0.65, 0.80)
Sedentary time, min[b]	0.74 (0.66, 0.81)	0.74 (0.65, 0.81)	0.73 (0.65, 0.80)	0.76 (0.68, 0.83)	0.75 (0.67, 0.82)	0.64 (0.54, 0.73)
Light activity time, min[c]	0.69 (0.59, 0.77)	0.70 (0.60, 0.77)	0.65 (0.55, 0.74)	0.75 (0.67, 0.82)	0.76 (0.68, 0.83)	0.62 (0.51, 0.72)
MVPA time, min[d]	0.68 (0.58, 0.76)	0.71 (0.62, 0.79)	0.65 (0.55, 0.77)	0.75 (0.67, 0.82)	0.51 (0.38, 0.63)	0.73 (0.64, 0.81)
Bouted-MVPA time, min[d]	0.68 (0.58, 0.76)	0.65 (0.55, 0.74)	0.67 (0.57, 0.75)	0.65 (0.55, 0.74)	0.45 (0.32, 0.59)	0.64 (0.53, 0.73)

Notes: Values are ICCs are adjusted for age, BMI, wear time, and season when the data were collected. Younger is <Median age (70.3y); Normal weight (BMI <25 kg/m^2), overweight/obese (BMI ≥25 kg/m^2); Less active (< Median vector magnitude counts per minute [VM cpm] 463,678 cpm).

VM, vector magnitude; MVPA, moderate to vigorous physical activity; cpm, counts per minute

[a] Summation of VM cpm and step count over the daily wear period.

[b] Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200.[19]

[c] Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2,690.[20]

[d] MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691,[20] MVPA-bouted uses same cpm threshold for activity accumulated in 10 consecutive minutes allowing 2-min below the threshold.

Table 4

Cross-Classification Matrix by Quartiles at Baseline and the Average Over 2-3 Years

	Quartiles for mean of Times 2 and 3					
Quartiles at baseline	**1**	**2**	**3**	**4**	**Total**	**Percent agreement**
Total counts						
1	**35**	*14*	2	1	52	
2	*14*	**23**	*12*	3	52	
3	3	*13*	**24**	*13*	53	
4	0	2	*15*	**35**	52	
Total	52	52	53	52	209	**56.0%**
Sedentary time						
1	**35**	*13*	3	1	52	
2	*13*	**15**	*21*	3	52	
3	4	*20*	**18**	*11*	53	
4	0	4	*11*	**37**	52	
Total	52	52	53	52	209	**50.2%**
Light intensity activity						
1	**37**	*12*	1	2	52	
2	*11*	**22**	*17*	2	52	
3	4	*17*	**20**	*12*	53	
4	0	1	*15*	**36**	52	
Total	52	52	53	52	209	**55.0%**
MVPA time						
1	**34**	*13*	5	1	53	
2	*15*	**23**	*12*	1	51	
3	3	*12*	**25**	*13*	53	
4	0	4	*11*	**37**	52	
Total	52	52	53	52	209	**56.9%**

Notes: Bolded values indicate classified in the same quartile at both-time points and italicized are within one quartile.

MVPA, moderate to vigorous physical activity; VM, vector magnitude; cpm, counts per minute

[a]Summation of VM cpm over the daily wear period. Baseline: Q1 <358,112; Q2 358,112 to <483,187; Q3: 483,187 to <582,059.8 Q4: ≥582,059.8. Time 2/3: Q1 <362,598; Q2 362,598 to <450,051.4; Q3: 450,051.4 to <559,695.8 Q4: ≥559,695.8.

[b]Sedentary time is defined as the daily sum of minutes during which the accelerometer registers VM cpm <200.[19] Baseline: Q1 <455.1 min; Q2: 455.1 to <509.2 min; Q3: 509.2 to 580.3 min Q4: ≥580.3 min. Time 2/3: Q1 <451 min; Q2: 451 to <506.3 min; Q3: 506.3 to 566.4 min Q4: ≥566.4 min.

[c]Light-intensity physical activity min is defined the daily sum of minutes during which the accelerometer registers VM cpm between 200 and 2690.[20] Baseline Q1: <296.6 min; Q2: 296.6 to <351.9 min; Q3: 351.9 to <414.0 min Q4: ≥414.0 min. Time 2/3 Q1: <290.1 min; Q2: 290.1 to <343.8 min; Q3: 343.8 to <396.6 min Q4: ≥396.6 min.

[d]MVPA time is defined the daily sum of minutes during which the accelerometer registers VM cpm >=2,691.[20] Baseline: Q1 <9.0 min; Q2: 9 to <21 min; Q3: 21 to <46.0 min Q4: ≥46.0 min. Time 2/3: Q1 <8.3 min; Q2: 8.3 to <19.1 min; Q3: 19.1 to <36.4 min Q4: ≥36.4 min.