

2005

The Reliability and Validity of Two Ambulatory Monitoring Actigraphs

Warren W. Tryon
Fordham University

Follow this and additional works at: https://fordham.bepress.com/psych_facultypubs

Part of the [Psychology Commons](#)

Recommended Citation

Tryon, Warren W., "The Reliability and Validity of Two Ambulatory Monitoring Actigraphs" (2005). *Psychology Faculty Publications*. 23.

https://fordham.bepress.com/psych_facultypubs/23

This Article is brought to you for free and open access by the Psychology at DigitalResearch@Fordham. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of DigitalResearch@Fordham. For more information, please contact considine@fordham.edu.

The reliability and validity of two Ambulatory Monitoring actigraphs

WARREN W. TRYON

Fordham University, New York, New York

Evidence for the reliability and validity of two models of Ambulatory Monitoring, Inc. actigraphs was obtained by testing four instruments of each kind 10 times each on a precision pendulum. Correlation and coefficient of variation methods were used to analyze the data. Reliability coefficients of .98 were obtained for both models. Coefficient of variation methods yielded reliability coefficients of 92% for the MotionLogger model and 97% for the BuzzBee model. Validity coefficients of .99 were obtained for both models. However, MotionLogger means were found to differ by up to 36% from one instrument to another, whereas BuzzBee means differed by 10% at most. Issues of standardization and measurement units were discussed.

Actigraphy is a general term that refers to the instrumented measurement of level of activity. A wide variety of devices is currently available with which to obtain longitudinal ambulatory activity level measures from the natural environment (Bouten, Westerterp, Verduin, & Janssen, 1994; Freedson, 1991; LaPorte, Montoye, & Caspersen, 1985; Meijer, Westerterp, Verhoeven, Koper, & ten Hoor, 1991; Melanson & Freedson, 1996; Montoye, Kemper, Saris, & Washburn, 1996, pp. 72–96; Tryon, 1985, 1991; Tryon & Williams, 1996; U.S. Department of Health and Human Services, 1996). Actigraphs are computerized instruments that use a microcontroller to measure movement and store data at user-selected intervals.

Instrument reliability differs from clinical repeatability (Tryon, 1991, pp. 9–15). Instrument reliability in this case concerns the extent to which actigraphs provide the same measurements when repeatedly moved in the same way. Such studies are best conducted under laboratory conditions in which test conditions can be carefully controlled and repeated with very little variation. All residual variation in how the testing apparatus moves the actigraph will be attributed to instrument error and will, therefore, underestimate actigraph reliability and validity. Clinical repeatability concerns the extent to which people produce similar actigraph readings from one testing to the next. This depends on several factors. Apart from instrument unreliability, variation in how the actigraph is attached and variation in how the person behaves—that is,

how they move the site of attachment (wrist or waist)—are two major reasons why actigraphy results may differ across tests. Placing one device on each hip questionably presumes symmetrical biomechanics. The aim of this research was to estimate the instrument reliability of an older and newer actigraph from a major vendor who has been selling actigraphs since their inception over 25 years ago and to compare these results with previously published clinical repeatability findings.

Unlike psychological tests, whose reliability and validity can be investigated only by administering them to people, actigraphs are physical instruments whose functional properties can be studied under laboratory conditions. Relatively few studies of instrument reliability have been reported. Foster, McPartland, and Kupfer (1978) investigated the reliability of the LSI sensor by transporting three of them through a 14° arc at a radius of 0.08 m, corresponding to 14° of rotation and a 0.0195-m sinusoidal movement. The correlation between Devices 1 and 2 was $r = .99$; the correlation between Devices 1 and 3 was also $r = .99$. Eaton, McKeen, and Lam (1988) investigated the reliability of the Kaulins and Willis Model 101 motion recorder (actometer) by attaching 27 of them to a chemical shaker bath for 5-, 10-, and 15-min trials. The effect of trials was significant and substantial [$F(2,52) = 28,544.87, p < .0001$], whereas the Actometer effect was not [$F(26,52) = 0.62, p < .90$]. The intraclass correlation of .99 supports the interchangeability of actometers. Tryon and Williams (1996) used two laboratory devices to standardize movement: a large pendulum and a spinner device. Three pendulum runs of fifty 10-sec epochs with three CSA (now MTI) actigraphs were conducted. The largest discrepancy among the actigraphs consisted of a 2.5% error, whereas the average error was approximately 0.6% (reliability of 97.5%–99.4%).

Several clinical repeatability reports have been published. Foster et al. (1978) reported test–retest correlations ranging from $r(5) = .78, p < .05$, to $r(5) = .94, p <$

The author thanks Thomas Kazlauskys and Ambulatory Monitoring, Inc. for their support in conducting this research. The author is one of the named patent holders for the BuzzBee actigraph. Correspondence concerning this article should be sent to W. W. Tryon, Department of Psychology, 441 East Fordham Road, New York, NY 10458-5198 (e-mail: wtryon@fordham.edu).

*Note—This article was accepted by the previous editor,
Jonathan Vaughan.*

.001. Washburn, Chin, and Montoye (1980) reported test–retest reliability coefficients ranging from $-.13$ to 0.81 . The negative correlation of $-.13$ indicates either a major equipment malfunction or substantial behavioral inconsistency by the participant; the latter appears to have been the case. Sallis, Buono, Roby, Carlson, and Nelson (1990) reported a left–right hip correlation of $r(10) = .96, p < .01$. In Study 2, they had 6 girls and 9 boys wear one device on each hip and walked/ran for 10 min at 3, 4, and 5 miles per hour. The correlation across 40 pairs of measurements was $r(38) = .89, p < .01$. Pambianco, Wing, and Robertson (1990) reported an average left–right hip correlation of $r(18) = .94, p < .01$. Nichols, Patterson, and Early (1992) reported a waist–back correlation of $r(26) = .95, p < .01$ for the younger participants and $r(26) = .98, p < .01$ for the older participants. Patterson et al. (1993) reported that average test–retest reliability was $.98$ across 12 situations. Leidy, Abbott, and Fedenko (1997) reported test–retest correlations of $r(16) = .70, p < .001$ for light activity, $r(16) = .66, p < .01$ for moderate activity, and $r(16) = .79, p < .001$ for heavy activity for the zero-crossing mode (visit <http://www.ambulatory-monitoring.com/modes.html> for details). The test–retest correlations for the time-above-threshold mode were $r(16) = .46, p < .06$ for light activity, $r(16) = .69, p < .001$ for moderate activity, and $r(16) = .69, p < .001$ for heavy activity.

METHOD

Instruments

Actigraphs. Four Ambulatory Monitoring, Inc. (AMI) MotionLogger actigraphs and four BuzzBee actigraphs were used. AMI actigraphs have been used by many investigators. The present findings pertain to the substantial published literature generated by these investigators.

Precision pendulum. The behavior of the pendulum is well known and highly repeatable. A precision pendulum constructed by Precision Control Design was used for all reliability and validity tests reported here. It stands 0.71 m high and has a movable arm 0.43 m long. The initial angle of displacement is 15° (0.26 radians) from the vertical plane.

Procedure

Each actigraph was attached to the interface portion of the precision pendulum and initialized to operate in proportional integrating measure (PIM) mode (visit <http://www.ambulatory-monitoring.com/modes.html> for details), causing the actigraph to integrate the filtered and rectified voltage signal from the its piezoelectric transducer and to record cumulative activity at the end of every 5 sec. The pendulum was displaced from the vertical plane by 15° , recording was initiated, and 10 sec later the pendulum was allowed to swing freely until coming to rest. Nine additional trials were run. This procedure was repeated for all eight actigraphs, resulting in 10 trials for each of the eight actigraphs. The different weights of the two actigraph

models created two different times for the pendulum to come to rest. The BuzzBee actigraphs weighed 2.2 oz (62.4 g) and took sixty-three 5-sec epochs to come to rest, whereas the MotionLogger actigraphs weighed 1.4 oz (39.7 g) and took eighty-four 5-sec epochs to come to rest. Their weight ratio is 1.57 , and their epoch ratio is $84/63 = 1.33$. All testing was conducted at the Ambulatory Monitoring facilities in Ardsley, NY.

The exponential decay function specified in Equation 1 was used as the gold standard for pendulum decay:

$$Y = ae^{-bt} \quad (1)$$

This equation was selected, not derived, on the basis of the following considerations: (1) Pendulums decay at an exponential rate, (2) this equation enabled experimental matching of the initial and the ending average actigraph values by fitting the constants a and b to the average result of the first and last trials of the MotionLogger and BuzzBee data, and (3) correlations are invariant over linear transformation and, therefore, a constant term, whether added, subtracted, multiplied, or divided, will not change the observed correlations between actigraph measures and this mathematical description of pendulum decay.

RESULTS

Reliability Analyses

Measurement reliability in this context concerns the psychometric equivalence across the 10 repeated pendulum trials for each actigraph. Two forms of a test are considered psychometrically equivalent if they are parallel (Honaker, 1988), and the following three requirements must be met to have parallel forms (Ghiselli, 1964, p. 227; Ghiselli, Campbell, & Zedeck, 1981, pp. 192–227): They must have (1) equal means and (2) equal variances, and (3) they must correlate to the same degree with other variables. If two test forms meet Criteria 2 and 3 but have different means, they can be made parallel by adding or subtracting a constant. Systematic average differences among actigraphs can be removed in this way, or the devices can be modified through calibration procedures until such differences disappear. If two test forms meet Criteria 1 and 3 but have different variances, they can be made parallel through an equipercentile transformation (Anastasi & Urbina, 1997, pp. 70–71). The third requirement—that the two tests, or forms of a test, be highly correlated with each other—is crucial and is a necessary condition for psychometric equivalence.

MotionLogger actigraphs. A total of eighty-four 5-sec epochs (7.0 min) were recorded during each trial before these actigraphs came to rest. Table 1 presents the means and standard deviations across all 84 measurements obtained during each trial, along with the correlation¹ between the 84 actigraph measurements and the 84 temporally corresponding Equation 1 values. Reliability was assessed by calculating all 45 independent test–retest correlation coefficients from a 10 trial \times 10 trial correlation matrix for each of the four MotionLogger

Table 1
Means, Standard Deviations, and Correlations With Simulated Pendulum Performance for
10 Trials of Eighty-Four 5-Sec Activity Measurements for Four MotionLogger Actigraphs

Trial	MotionLogger Actigraph 1			MotionLogger Actigraph 2			MotionLogger Actigraph 3			MotionLogger Actigraph 4		
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>
1	169.58	109.65	.992	204.17	116.636	.991	189.24	107.93	.988	251.20	153.73	.994
2	205.98	124.82	.994	285.33	161.076	.995	253.24	143.38	.993	311.52	179.92	.995
3	204.83	122.69	.995	295.73	163.254	.994	249.20	144.64	.994	307.65	176.45	.994
4	212.60	121.92	.991	296.38	161.987	.994	250.08	142.55	.992	301.57	175.55	.995
5	210.38	122.39	.993	292.96	165.644	.995	249.82	140.32	.992	310.11	175.57	.994
6	212.38	123.85	.994	297.56	169.916	.996	253.43	142.15	.992	310.30	178.00	.995
7	201.57	127.08	.993	292.90	166.826	.995	248.94	148.72	.994	304.25	177.69	.995
8	208.36	123.38	.994	284.39	162.787	.995	239.33	137.34	.992	296.39	172.34	.995
9	213.49	121.85	.993	286.13	165.165	.995	252.05	141.10	.993	294.38	174.89	.995
10	206.67	123.10	.993	285.81	168.191	.995	251.88	140.67	.992	288.86	169.27	.995
Average	204.58	122.07	.993	282.14	160.15	.995	243.72	138.88	.992	297.62	173.34	.995
<i>SD</i>	12.88	4.64	.001	27.85	15.54	.001	19.56	11.28	.002	17.99	7.51	.001
<i>CV</i>	6.30	3.80	0.11	9.87	9.71	0.14	8.03	8.12	0.17	6.05	4.33	0.05

actigraphs. All correlations equaled or exceeded $r(82) = .980, p < .0001$. Reliable measurement implies that the same mean and standard deviation result when the same phenomenon—pendulum decay, in this case—is repeatedly measured. The coefficient of variation [$CV = (SD/M) \times 100$] was used to quantify the degree of variability, relative to the magnitude of movement detected. The CVs for the four MotionLogger means were 6.30%, 9.87%, 8.03%, and 6.05% or 7.56%, on average. Hence, MotionLogger means differed by approximately 8% when repeatedly tested. The CVs for the four MotionLogger standard deviations were 3.80%, 9.71%, 8.12%, and 4.33% or 6.49%, on average. Hence, MotionLogger standard deviations varied by approximately 7% when repeatedly tested.

Instruments are expected to be comparable. MotionLogger 1 gave the lowest mean of 204.58 activity units² whereas MotionLogger 4 gave the highest mean of 297.62. The average of the four MotionLogger means was 257.015. The difference of 93.04 between the highest and the lowest means divided by the average equaled

36.2%. MotionLogger actigraph means varied by approximately 36% across devices.

BuzzBee actigraphs. A total of sixty-three 5-sec epochs (5.25 min) were recorded during each trial before these actigraphs came to rest. Table 2 presents the means and standard deviations across all 63 measurements obtained during each trial, along with the correlation between the 63 actigraph measurements and the 63 temporally corresponding Equation 1 values. Reliability was assessed by calculating all 45 independent test–retest correlation coefficients from a 10 trial \times 10 trial correlation matrix for each of the four BuzzBee actigraphs. All the correlations equaled or exceeded $r(61) = .978, p < .0001$. Reliable measurement implies that the same mean and standard deviation result when the same phenomenon (pendulum decay) is repeatedly measured. The CV was used to quantify the degree of variability relative to the magnitude of movement detected. The CVs for the four BuzzBee means were 1.79%, 1.51%, 5.83%, and 2.28% or 2.85%, on average. Hence, BuzzBee means

Table 2
Means, Standard Deviations, and Correlations With Simulated Pendulum Performance for
10 Trials of Sixty-Three 5-Sec Activity Measurements for Four BuzzBee Actigraphs

Trial	BuzzBee Actigraph 1			BuzzBee Actigraph 2			BuzzBee Actigraph 3			BuzzBee Actigraph 4		
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i>
1	246.76	100.88	.990	241.65	108.14	.991	260.95	121.11	.994	266.92	118.36	.994
2	253.35	108.64	.992	252.32	112.87	.994	257.17	118.49	.993	273.06	117.18	.991
3	249.00	105.48	.992	255.33	109.97	.993	244.65	128.25	.994	273.37	117.40	.993
4	248.94	106.16	.994	253.02	109.00	.992	250.05	124.69	.993	256.73	111.86	.992
5	243.59	102.91	.990	253.70	106.68	.991	246.59	124.81	.994	273.52	117.92	.994
6	243.43	108.36	.993	250.46	107.56	.992	212.11	115.90	.992	272.19	116.27	.992
7	240.19	102.72	.991	248.60	103.29	.990	253.79	117.66	.991	274.95	119.12	.993
8	246.84	104.47	.991	251.24	107.41	.992	258.62	120.02	.990	260.59	113.75	.994
9	238.86	103.93	.990	252.68	105.89	.991	260.51	120.59	.994	268.68	116.93	.994
10	243.60	102.69	.992	249.98	105.91	.993	258.60	120.24	.994	272.38	117.78	.994
Average	245.46	104.62	.992	250.90	107.67	.992	250.30	121.18	.993	269.24	116.66	.993
<i>SD</i>	4.38	2.54	.001	3.79	2.60	.001	14.60	3.74	.001	6.13	2.12	.001
<i>CV</i>	1.79	2.43	0.14	1.51	2.41	0.12	5.83	3.08	0.15	2.28	1.90	0.11

differed by approximately 3% when repeatedly tested. The CVs for the four BuzzBee standard deviations were 2.43%, 2.41%, 3.08%, and 1.90% or 2.46%, on average. Hence, MotionLogger standard deviations varied by approximately 3% when repeatedly tested.

Instruments are expected to be comparable. BuzzBee 1 gave the lowest mean of 245.46, whereas BuzzBee 4 gave the highest mean of 269.24. The average of the four BuzzBee means was 253.975. The difference of 23.78 between the highest and the lowest means divided by the average equaled 9.4%. BuzzBee actigraph means varied by approximately 10% across devices.

Validity Analyses

Each actigraph model was validated against the simulated pendulum decay function specified by Equation 1. The repeated measurements taken during each of the 10 pendulum trials were correlated with expected values at corresponding times, using Equation 1.

MotionLogger actigraphs. Table 1 provides the Pearson product moment correlations between measured and expected values for each of the four MotionLogger actigraphs across all 84 recording epochs obtained during each trial. Results are reported to three decimals in, because all the values exceeded .99. These data indicate that the validity of the four MotionLogger actigraphs studied exceeded .99.

BuzzBee actigraphs. Table 2 provides the Pearson product moment correlations between measured and expected values for each of the four BuzzBee actigraphs across all 63 recording epochs obtained during each trial. Results are again reported to three decimals in, because all the values exceeded .99. These data indicate that the validity of the four BuzzBee actigraphs studied exceeded .99.

DISCUSSION

Controlled laboratory analysis revealed that the MotionLogger and BuzzBee actigraphs examined appear to be highly reliable and valid. Test–retest correlation coefficients equaled or exceeded .98 for both devices, and validity coefficients exceeded .99 for both devices. More stringent CV analysis revealed 8% measurement error in the means and standard deviations of the MotionLogger actigraphs (92% reliable) and 3% measurement error in the means and standard deviations of the BuzzBee actigraphs (97% reliable). Validity coefficients of at least .99 were obtained for both models. These findings are similar to those obtained by Foster et al. (1978), Eaton et al. (1988), and Tryon and Williams (1996). MotionLogger means were found to differ by up to 36% from one instrument to another, whereas BuzzBee means differed by 10% at most. The most discrepant result was for MotionLogger 1 on Trial 1. The fact that all nine subsequent trials were both higher and more consistent suggests that the considerably lower values recorded throughout Trial 1 might have been due to a warm-up effect. However, no other evidence of warm-up effects was found for any of the

other seven actigraphs. BuzzBee actigraphs appear to be considerably more similar to one another than are MotionLogger actigraphs.

It is important to make between-device differences as small as possible, because between-device variance will inflate within-group error variance and, consequently, reduce statistical power, making it more likely that investigators will fail to find hypothesized effects. Between-device variability can be reduced through calibration procedures that alter the properties of the electronic circuits that govern the activity measurement values recorded in such a way that each device matches a laboratory standard as closely as possible. This is the preferred method for making devices comparable. Alternatively, between-device variability can be reduced by determining an empirical constant for each device that, when multiplied by each reading, increases the similarity between devices. Data from instruments that give below-average results can be increased by multiplying each data point by the ratio of the mean reading across all instruments used in a study by the mean reading of the specific instrument in question. For example, the mean reading of the MotionLogger with the lowest readings was 204.58. The mean of the means for all four MotionLogger actigraphs was 257.015. Hence, a conversion factor for this low-reading MotionLogger actigraph is $257.015/204.62 = 1.2563$. Data from instruments that give above-average results can be decreased by multiplying each data point by the reverse ratio. For example, the mean reading of the MotionLogger with the highest readings was 297.62. The mean of the means for all four MotionLogger actigraphs was 257.015. Hence, a conversion factor for this high-reading MotionLogger actigraph is $257.015/297.62 = 0.8636$. Alternatively, an equipercentile transformation (Anastasi & Urbina, 1997, pp. 70–71) can be used to reduce between-device variability. Percentile scores from 1 to 100 can be computed for each device on the basis of laboratory testing. Activity measurements are converted into percentile scores, and these are recorded, instead of the activity units. Actigraph vendors could include these calculations into actigraph firmware.

The reliability and validity estimates reported here are considerably greater than those in some of the previously published studies in which people, rather than laboratory instruments, have been used to assess the measurement properties of actigraphs. Clinical repeatability confounds several sources of variance with instrument unreliability. Whereas all actigraphs were attached to a pendulum of fixed length, actigraphs are frequently attached to the wrists of arms of varying lengths, and limb length influences activity readings (Tryon, 1984). Whereas all actigraphs maintain the same orientation during testing, people move their wrists through various angles while behaving. Whereas the precision pendulum behaves the same way from trial to trial, people rarely behave as consistently. The test–retest correlation coefficients of at least .978, obtained under the laboratory conditions used in this study, were found to account for 95.6% of the

variance. Person-based reliability coefficients as low as .46, reported by other investigators, account for 21.2% of the variance (cf. Leidy et al., 1997) and severely underestimate actigraph reliability by erroneously attributing variance to instrument unreliability, rather than to other sources. The negative test-retest correlation coefficient of $-.13$ reported by Washburn et al. (1980) indicates that participants can behave very inconsistently from one session to the next.

Phantom Measurement

Phantom measurement refers to the situation in which reliability coefficients are substantially greater than validity coefficients (Tryon & Bernstein, 2002). For example, if a test has a reliability coefficient of .9 and a validity coefficient of .3, it reliably measures 81% of something, but only 9% of the reference criterion, indicating that it is measuring 72% of something else. It is noteworthy that the validity coefficients reported here for the two models of actigraphs studied are essentially equivalent to their reliability coefficients, which means that no phantom measurement is present. This finding supports the conclusion that actigraphs measure movement, activity, and nothing else. Activity ratings and self-reports have the potential of reflecting factors other than movement. For example, Tryon and Pinto (1994) reported that teachers sometimes rate children as hyperactive when objective measures of their activity level show that they are no more active than other children that the teachers rated as normally active. These inaccurate teachers' ratings clearly reflect information other than activity level. Phantom measurement is a primary reason to measure, rather than rate or self-report, activity levels.

Limitations

Several limitations of the present research are worth noting. Tryon and Williams (1996) reported small but measurable variations in pendulum behavior, using a laboratory grade accelerometer. It is likely that measurable variation existed in the pendulum runs used to test actigraphs in this study, which to this extent, underestimated the obtained reliability and validity coefficients. However, the magnitude of the reliability and validity coefficients reported above are sufficiently close to unity for all contemporary clinical and research applications.

The pendulum arm used by Tryon and Williams (1996) was 1.7 m (5 ft 7 in.) long versus the present pendulum, whose arm was 0.43 m long. The frequency with which a simple pendulum oscillates is given by Equation 2,

$$f = \frac{1}{2\pi\sqrt{\frac{L}{g}}}, \quad (2)$$

where $g = 9.81 \text{ msec}^{-2}$ and f is in hertz (cycles/second). Tryon and Williams's pendulum ($L = 1.7 \text{ m}$) oscillated at $f = 0.382 \text{ Hz}$, whereas the present pendulum ($L = 0.43 \text{ m}$) oscillated at $f = 0.760 \text{ Hz}$, or nearly twice as fast. It is possible that an actigraph could respond more

reliably at one frequency than at another. Pendulums operate at a single frequency. The magnitude of acceleration associated with the pendulum decays as per Equation 1 above, but the frequency remains approximately constant and proportional to the square root of the distance between the pivot and the actigraph, which is the basis of the pendulum clock (cf. Equation 2). Tryon and Williams also used a spinner device that sampled a range of movement frequencies, as well as amplitudes. This device was not available to test the AMI actigraphs. However, the spinner and pendulum results presented by Tryon and Williams were similar, suggesting that reliability assessment of the AMI actigraphs on a spinner might also be high.

Both the pendulum and the spinner test in a single plane, which is consistent with the fact that the AMI actigraph is a uniaxial device. Waist movement occurs primarily in the vertical planes, but human wrist movements occur in three dimensions. Although there are advantages to testing in all three dimensions, Redmond and Hegge (1985) reported that data recorded along three separate axes were correlated at around $r = .60$ when a 1-sec recording epoch was used and increased to $r = .98$ when a 256-sec recording epoch was used. A 5-sec recording epoch was used in this study, and a 60-sec recording epoch is standard. A 5-min epoch contains 300 sec and should render dimensional measurements correlated in excess of .98. The substantial cross-axis correlation associated with these recording epochs suggests that the findings of a single-axis assessment tend to generalize to a multiple-axis assessment, especially at longer recording epochs. Additional research is needed and better laboratory devices are required to more fully investigate the operational characteristics of actigraphs.

Measurement Units

Measurement units are associated with the instantaneous quantification performed by the actigraph and/or with average measurements made over a recording epoch. The PIM mode (<http://www.ambulatory-monitoring.com/modes.html>) integrates area under a time-varying voltage curve. The sensor used in the CSA actigraph, now sold by MTI (<http://www.mtiactigraph.com/>) responds to time varying acceleration in units of g/sec , where g is the rate at which bodies fall freely (9.81 msec^{-2}). The actigraph makes measurements 10 times each second. Integration over the resulting sampling time of 0.1 sec in combination with other details provided by Tryon and Williams (1996), results in units of 1.664 milli- $g/\text{activity count}$. One cannot work backward from counts to gs , because the frequency of the sensor beam remains unknown. A possible solution to this problem is to obtain simultaneous actigraph and calibrated instrument grade accelerometer readings over the full range of activity levels each device can accurately record and to estimate a conversion equation.

Alternatively, one can empirically convert average activity readings over a typical recording epoch, such as

1 min, into standard units. Freedson, Melanson, and Sirard (1998) have published a regression equation for converting average CSA actigraph, now sold by Manufacturing Technology, Inc. (MTI), activity counts over 1-min epochs to calories. Version 3.2.3 of MTI software implements this formula and can be downloaded from the users' area of their Web page (<http://www.mtiactigraph.com/software.asp>) Individual Monitoring (IM) Systems (<http://www.imsystems.net/>) provide equations for converting average activity level into average calories on the basis of empirically derived constants. Klippel and Heil (2003), Heil and Klippel (2003), and Puyau, Adolph, Vohra, and Butte (2002) have published regression equations for converting the Mini Mitter Co. actigraph (<http://www.minimitter.com/>) counts into calories. Puyau et al. also have published regression equations for converting MTI actigraph counts into calories.

REFERENCES

- ANASTASI, A., & URBINA, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- BOUTEN, C. V., WESTERTERP, K. R., VERDUIN, M., & JANSSEN, J. D. (1994). Assessment of energy expenditure for physical activity using a triaxial accelerometer. *Medicine & Science in Sports & Exercise*, **26**, 1516-1523.
- EATON, W. O., MCKEEN, N. A., & LAM, C. (1988). Instrumented motor activity measurement of the young infant in the home: Validity and reliability. *Infant Behavior & Development*, **11**, 375-378.
- FOSTER, F. G., MCPARTLAND, R. J., & KUPFER, D. J. (1978). Motion sensors in medicine: Pt. i. A report on reliability and validity. *Journal of Inter-American Medicine*, **3**, 4-8.
- FREEDSON, P. S. (1991). Electronic motion sensors and heart rate as measures of physical activity in children. *Journal of School Health*, **61**, 220-223.
- FREEDSON, P. S., MELANSON, E., & SIRARD, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise*, **30**, 777-781.
- GHSISELLI, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.
- GHSISELLI, E. E., CAMPBELL, J. P., & ZEDECK, S. (1981). *Measurement theory for the behavioral sciences*. New York: Freeman.
- HEIL, D. P., & KLIPPEL, N. J. (2003). Validation of energy expenditure prediction algorithms in adults using the Actical electronic activity monitor. *Medicine & Science in Sports & Exercise*, **35**, S285.
- HONAKER, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology*, **8**, 561-577.
- KLIPPEL, N. J., & HEIL, D. P. (2003). Validation of energy expenditure prediction algorithms in adolescents/teens using the Actical electronic activity monitor. *Medicine & Science in Sports & Exercise*, **35**, S282.
- LAPORTE, R. E., MONTROYE, H. J., & CASPERSEN, C. J. (1985). Assessment of physical activity in epidemiologic research: Problems and prospects. *Public Health Reports*, **100**, 131-146.
- LEIDY, N. K., ABBOTT, R. D., & FEDENKO, K. M. (1997). Sensitivity and reproducibility of the dual-mode actigraph under controlled levels of activity intensity. *Nursing Research*, **46**, 5-11.
- MEIJER, G. A. L., WESTERTERP, K. R., VERHOEVEN, F. M. H., KOPER, H. B. M., & TEN HOOR, F. (1991). Methods to assess physical activity with special reference to motion sensors and accelerometers. *IEEE Transactions on Biomedical Engineering*, **38**, 221-229.
- MELANSON, E. L., JR., & FREEDSON, P. S. (1996). Physical activity assessment: A review of methods. *Critical Reviews in Food Science & Nutrition*, **36**, 385-396.
- MONTROYE, H. J., KEMPER, H. C. G., SARIS, W. H. M., & WASHBURN, R. A. (1996). *Measuring physical activity and energy expenditure*. Champaign, IL: Human Kinetics.
- NICHOLS, J. F., PATTERSON, P., & EARLY, T. (1992). A validation of a physical activity monitor for young and older adults. *Canadian Journal of Sport Sciences*, **17**, 299-303.
- PAMBIANCO, G., WING, R. R., & ROBERTSON, R. (1990). Accuracy and reliability of the Caltrac accelerometer for estimating energy expenditure. *Medicine & Science in Sports & Exercise*, **22**, 858-862.
- PATTERSON, S. M., KRANTZ, D. S., MONTGOMERY, L. C., DEUSTER, P. A., HEDGES, S. M., & NEBEL, L. E. (1993). Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology*, **30**, 296-305.
- PUYAU, M. R., ADOLPH, A. L., VOHRA, F. A., & BUTTE, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, **10**, 150-157.
- REDMOND, D. P., & HEGGE, F. W. (1985). Observations on the design and specification of a wrist-worn human activity monitoring system. *Behavior Research Methods, Instruments, & Computers*, **17**, 659-669.
- SALLIS, J. F., BUONO, M. J., ROBY, J. J., CARLSON, D., & NELSON, J. A. (1990). The Caltrac accelerometer as a physical activity monitor for school-age children. *Medicine & Science in Sports & Exercise*, **22**, 698-703.
- TRYON, W. W. (1984). Measuring activity using actometers: A methodological study. *Journal of Behavioral Assessment*, **6**, 147-153.
- TRYON, W. W. (1985). The measurement of human activity. In W. W. Tryon (Ed.), *Behavioral assessment in behavioral medicine* (pp. 200-256). New York: Springer-Verlag.
- TRYON, W. W. (1991). *Activity measurement in psychology and medicine*. New York: Plenum.
- TRYON, W. W., & BERNSTEIN, D. (2002). Understanding measurement. In J. C. Thomas & M. Hersen (Eds.), *Understanding research in clinical and counseling psychology: A textbook* (pp. 27-68). Mahwah, NJ: Erlbaum.
- TRYON, W. W., & PINTO, L. P. (1994). Comparing activity measurements and ratings. *Behavior Modification*, **18**, 251-261.
- TRYON, W. W., & WILLIAMS, R. (1996). Fully proportional actigraphy: A new instrument. *Behavior Research Methods, Instruments, & Computers*, **28**, 392-403.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (1996). *Physical activity and health: A report of the surgeon general* (pp. 29-37). Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.
- WASHBURN, R., CHIN, M. K., & MONTROYE, H. J. (1980). Accuracy of pedometer in walking and running. *Research Quarterly for Exercise & Sport*, **51**, 695-702.

NOTES

1. The Pearson product moment correlation coefficient makes no assumption of independence of observations. The routine calculation of time-lagged autocorrelations to quantify serial dependency is proof of this.
2. Most actigraphs do not report standard units of measure. Measurement standards have not yet been developed. The issue of measurement units will be considered in the Discussion section.

Manuscript received February 13, 2004;
revision accepted for publication August 9, 2004.)