Ethnic Differences in Physical Activity and Metabolic Risk: The Dallas Heart Study

Susan G. Lakoski, MD MS[1], Julia Kozlitina, PhD[2]

[1]Department of Internal Medicine, University of Vermont, Colchester, VT 05446

[2]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75390

Supplemental Digital Content 1: Accelerometer Data Processing and Quality Control Procedures

Data Processing and Quality Control

Visual representations of participants' activity levels, called actograms, were generated using Actical software (Version 2.1). A trained reader inspected all activity profiles to determine the start and end of monitor wear and check for device malfunction. The start of monitor wear was defined as the beginning of the first full hour of activity, and the end was defined as the last period of activity lasting more than 2 hours, followed by no valid activity. Actical manufacturer does not provide a commercially built calibrator or specify a valid range of count values for the monitor. However, no record in our data exceeded 20,000 counts per minute (CPM), suggesting that this may be the saturation value for the device. In a separate experiment involving volunteer subjects running on a treadmill at ~6 mph, a wrist-worn Actical generated output averaging >17,000 CPM (unpublished observations). Therefore, files that contained an implausibly large proportion of high-intensity values (>15,000 CPM) and those, in which the count per minute value never returned to zero after the device was removed, but remained at a nearly constant nonzero level (indicative of a baseline drift), were thought to come from malfunctioning monitors and were excluded from analysis (n = 63).

Study participants were asked to wear the monitor for 7 full days, however some did not comply with the protocol. Individuals who provided less than 1 full day of monitor wear (n = 84) were excluded from analysis. In addition, many profiles contained only partial data on the first and last days of monitor wear, but more than 24 hours of data on both days combined. To maximize the length of the monitoring period for each participant and the number of individuals available for analysis, we imputed the missing data on the last day of monitor wear with available data from the first day (provided that both days were either week- or weekend days). This has increased the number of participants with 4 or more full days of monitor wear from 2,498 to 2,628. Physical activity levels were compared between days containing full data and imputed data. No systematic differences were detected. Missing values in individual files were replaced by an average of count values immediately before and after the missing data interval (<0.1% of epochs in <0.5% of files).

Activity Count Normalization

Actical devices used in the current study came from two different production batches, as identified by the serial number beginning with a letter B or C. Although the two batches were identical in their technical specifications (correspondence with the manufacturer), preliminary analysis revealed that series C devices recorded systematically higher activity counts than those of the B series. Table S1 presents activity summaries for individuals who wore series B and C monitors. We restricted the analysis to those participants who provided at least 6 days of monitor wear in order to eliminate the potential bias resulting from differences in the length of the observation period. The two groups were similar in demographic and anthropometric characteristics. Mean activity counts were 1.35 higher, on average, for those wearing series C monitors. However, the differences between the two batches appeared to be intensity-dependent. Individuals who used series C devices were estimated to spend nearly twice as much time in moderate activity as those using series B devices, but were not substantially different in the estimates of time spent in vigorous activity, especially when looking at ≥10-minute bouts.

Others have found similar differences between Actical devices. Evans et al.(2) reported that one of the Actical versions used in their study recorded only about 0.55 total counts compared to other versions, and scaled the output from that version (i.e., divided by 0.55) so that all monitors had the same average count value. The simple scaling method works well when the output from several batches differs by a constant proportionality factor, but may not be suitable when there are non-linear relationships between batches.

For example, when applied to our data the scaling method (i.e., dividing the output from series C devices by a constant factor, 1.35) successfully removed the difference in mean counts per minute, but seemed to overcorrect higher-intensity values, so that estimates of time spent in moderate-to-vigorous activity were now significantly lower for the C series (Table S1).

The non-linear differences were confirmed in a laboratory experiment (performed as part of a separate reliability study), in which all available monitors were placed on an orbital shaker and simultaneously rotated in a horizontal plane at two speeds: low (90 rpm) and high (170 rpm) for several runs lasting 20 minutes. We found a significant difference in mean counts per minute between series B and C devices at the low speed (228.5 versus 434.5 CPM, respectively, $P = 5.2 \times 10^{-20}$), but not at the high speed (4207 versus 4194 CPM, $P = 0.75$).

To understand the nature of the differences between the two batches better and to remove systematic variation between monitors, we next compared empirical quantiles of counts obtained from series B and C monitors. The motivating idea behind the approach is that, while some individuals are more active than others, the overall *distribution* of counts should be similar between batches. We then adapted the methods used in microarray analysis(1) to develop a normalization algorithm based on quantile alignment. To account for differences in the length of monitor wear between individuals, and to reduce the total amount of data, we calculated a fixed number of quantiles ($p = 0, 0.1, 0.2, \ldots, 1$) for each individual, then averaged these quantiles across individuals. As the raw distribution of counts is skewed, we took square roots of count values prior to calculating the quantiles

(one would obtain the same result by taking the square root after the calculation, since quantiles depend on the ranking of observations only). The square root transformation improves visual comparisons of the distributions between batches and acts as a variance-stabilizing transformation for count data. Figure S1A shows mean quantiles of square roots of counts obtained from series B and series C monitors, along with 95% ranges. Mean quantiles for series C monitors tend to be higher than those for series B over the entire range of intensities; but the relationship is non-linear, as illustrated by the quantile-quantile plot in the figure inset. In particular, there is a large difference at a moderate-intensity range, but small difference at the low and high intensities. To equalize the distribution of counts between series B and C monitors, we estimated the relationship between mean quantiles with a polynomial regression (including a square and a cubic term), and used the estimated coefficients to correct count per minute values for series C devices, as follows:

if $CPM_C > 0$ and $CPM_C < 5000$,

$$CPM_C^* = (-0.02202 + 0.7035 \times CPM_C^{(1/2)} + 0.0057 \times CPM_C^{(2/2)} - 0.00002 \times CPM_C^{(3/2)})^2$$

else, $CPM_C^* = CPM_C$

(We did not correct count values above 5000 CPM, because there was not enough data in that range to estimate the relationship reliably.) Figure S1B displays empirical quantiles after normalization, showing that the distribution of counts for the two batches is now the same on average, although there is still a lot of variability in individual quantiles (see the 95% bands), which is partly due to individual differences in physical activity levels. The bottom section of Table S1 confirms that activity summaries did not differ significantly between batches after the normalization.

We have tried to model the relationship between quantiles using more flexible methods (e.g., natural splines) but the resulting correction did not offer a substantial improvement over the regression-based method shown here, therefore we used the simpler approach. To ensure that the empirical normalization algorithm did not introduce bias into our data, we repeated the primary association analysis for a subset of participants who used series B monitors. The results in this subset were consistent with those obtained using the full population (Table S2). We report the findings for the full population to increase power.

References

1. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-93.
2. Evans DS, Snitker S, Wu SH et al. Habitual sleep/wake patterns in the Old Order Amish: heritability and association with non-genetic factors. *Sleep*. 2011;34(5):661-9.

Table S1. Subject Characteristics and Physical Activity Outcomes for Series B and C Monitors

	Series B ($n_{subject}$=1444)	Series C ($n_{subject}$=804)	P-value
Age (yr)	50.4 (10.8)	50.8 (11.3)	0.38
Male (%)	38.5	36.2	0.30
Race/ethnicity (%)			
Non-Hispanic Black	53.1	55.2	
Non-Hispanic White	31.7	29.1	0.43
Hispanic	15.2	15.7	
Body mass index ($kg \cdot m^{-2}$)	31.7 (7.5)	31.5 (7.6)	0.61
I. Raw data:			
Mean CPM	253.4 (111.6)	342.8 (128.6)	5.0×10^{-61}
Moderate activity, $min \cdot d^{-1}$	35.9 (34.0)	67.9 (52.1)	1.5×10^{-63}
Moderate-to-vigorous activity, $min \cdot d^{-1}$	37.4 (35.6)	69.6 (53.5)	8.5×10^{-61}
in ≥10-min bouts	9.7 (17.3)	20.9 (30.7)	2.8×10^{-33}
Vigorous activity, $min \cdot d^{-1}$	1.5 (3.7)	1.8 (3.9)	2.4×10^{-4}
in ≥10-min bouts	0.6 (2.9)	0.6 (3.3)	0.42
II. After scaling ($\times 1.35^{-1}$):			
Mean CPM	253.4 (111.6)	250.2 (93.9)	0.69
Moderate activity, $min \cdot d^{-1}$	35.9 (34.0)	25.6 (25.4)	8.1×10^{-15}
Moderate-to-vigorous activity, $min \cdot d^{-1}$	37.4 (35.6)	26.2 (26.1)	1.3×10^{-15}
in ≥10-min bouts	9.7 (17.3)	5.7 (11.9)	1.5×10^{-14}
Vigorous activity, $min \cdot d^{-1}$	1.5 (3.7)	0.7 (2.9)	3.5×10^{-37}
in ≥10-min bouts	0.6 (2.9)	0.4 (3.0)	2.5×10^{-6}
III. After quantile normalization:			
Mean CPM	253.4 (111.6)	254.7 (104.7)	0.47
Moderate activity, $min \cdot d^{-1}$	35.9 (34.0)	36.5 (33.2)	0.32
Moderate-to-vigorous activity, $min \cdot d^{-1}$	37.4 (35.6)	38.0 (34.5)	0.36
in ≥10-min bouts	9.7 (17.3)	9.1 (16.8)	0.28
Vigorous activity, $min \cdot d^{-1}$	1.5 (3.7)	1.4 (3.6)	0.98
in ≥10-min bouts	0.6 (2.9)	0.5 (3.2)	0.082

Continuous characteristics were compared using Wilcoxon rank-sum test and categorical using chi-square tests. Intensity cut-points: 1500 CPM for moderate- and 4000 CPM for vigorous activity. Abbreviations: CPM – count per minute.

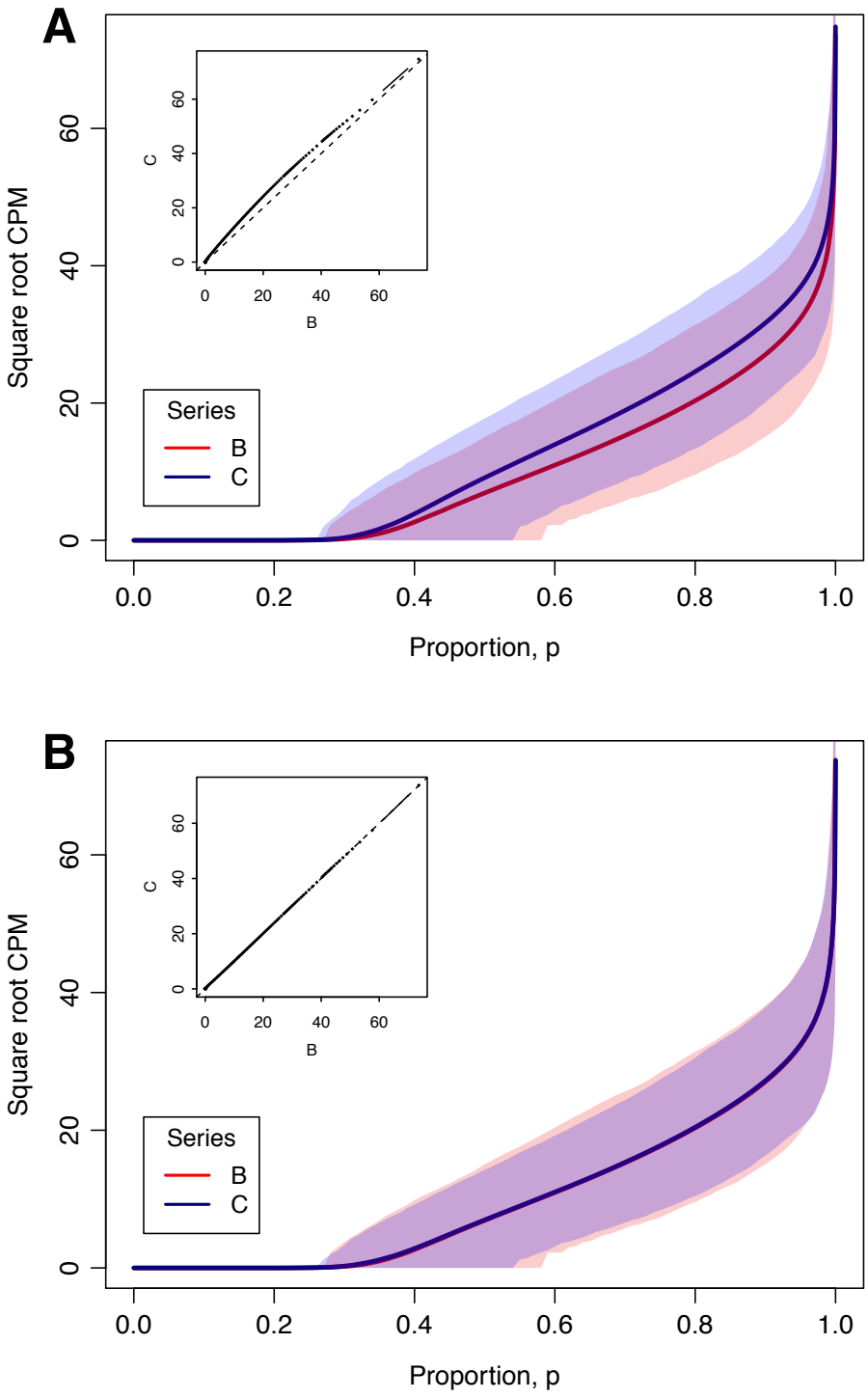

Figure 5. (A) Empirical quantiles of the square roots of counts per minute obtained from series B and C monitors. Thick curves show average quantiles for each device batch, and the colored bands represent point-wise 95% ranges. The inset displays a quantile-quantile plot of CPM for the two batches. The dashed line is the reference line, x=y. (B) Quantiles of counts after normalization.

Table S2. Association Between Metabolic Risk Factors and Moderate-to-Vigorous Activity Duration for DHS Participants Who Used Series B Monitors

Trait	N	Beta (SE) (per 10 min · d^{-1})	P-value	P (interaction)
BMI (kg · m^{-2})				
Non-Hispanic Black	835	-0.043 (0.011)	3.9×10^{-5}	
Non-Hispanic White	494	-0.056 (0.014)	1.2×10^{-4}	
Hispanic	244	-0.032 (0.013)	0.013	
Total	1573	-0.043 (0.007)	1.8×10^{-9}	0.65
Waist circumference (cm)				
Non-Hispanic Black	826	-0.053 (0.010)	3.9×10^{-7}	
Non-Hispanic White	492	-0.060 (0.015)	4.9×10^{-5}	
Hispanic	244	-0.035 (0.013)	0.0065	
Total	1562	-0.050 (0.007)	3.7×10^{-12}	0.50
Glucose (mg · dL^{-1}) [a]				
Non-Hispanic Black	647	-0.018 (0.011)	0.11	
Non-Hispanic White	433	-0.013 (0.015)	0.39	
Hispanic	204	0.009 (0.013)	0.48	
Total	1284	-0.010 (0.008)	0.19	0.03
HOMA-IR (U)				
Non-Hispanic Black	810	-0.040 (0.009)	1.4×10^{-5}	
Non-Hispanic White	484	-0.026 (0.013)	0.041	
Hispanic	238	-0.011 (0.012)	0.35	
Total	1532	-0.030 (0.006)	4.1×10^{-6}	0.24
Systolic BP (mm Hg)				
Non-Hispanic Black	833	0.011 (0.010)	0.28	
Non-Hispanic White	493	0.035 (0.013)	0.0092	
Hispanic	244	0.002 (0.012)	0.86	
Total	1570	0.013 (0.007)	0.052	0.22
Diastolic BP (mm Hg)				
Non-Hispanic Black	833	0.006 (0.010)	0.55	
Non-Hispanic White	493	0.031 (0.014)	0.031	
Hispanic	244	-0.009 (0.013)	0.51	
Total	1570	0.007 (0.007)	0.32	0.044
Heart Rate (bpm)				
Non-Hispanic Black	833	-0.047 (0.010)	3.3×10^{-6}	
Non-Hispanic White	493	-0.018 (0.016)	0.26	
Hispanic	244	-0.009 (0.014)	0.52	
Total	1570	-0.031 (0.007)	2.8×10^{-5}	0.016
Triglycerides (mg · dL^{-1}) [b]				
Non-Hispanic Black	825	0.039 (0.009)	2.8×10^{-5}	
Non-Hispanic White	491	0.019 (0.014)	0.17	
Hispanic	241	-0.002 (0.012)	0.86	
Total	1557	0.024 (0.007)	3.8×10^{-4}	0.49
HDL-C (mg · dL^{-1})				
Non-Hispanic Black	825	-0.018 (0.010)	0.068	
Non-Hispanic White	491	-0.032 (0.016)	0.042	
Hispanic	241	-0.012 (0.014)	0.37	
Total	1557	-0.020 (0.007)	0.0048	0.041

LDL-C ($mg \cdot dL^{-1}$)				
Non-Hispanic Black	825	0.012 (0.011)	0.29	
Non-Hispanic White	491	0.002 (0.015)	0.88	
Hispanic	241	-0.011 (0.014)	0.44	
Total	1557	0.003 (0.007)	0.70	0.46

Associations are reported for total duration of moderate-to-vigorous activity (i.e., including every minute above threshold). The reported beta coefficients represent a difference in the response (in standard deviation units) associated with a 10-min difference in the duration of PA. The analyses are adjusted for age, gender, ethnicity, BMI (except where BMI and waist circumference were the response) and diabetes as indicated. All response variables except glucose and LDL-cholesterol were logarithm transformed prior to analysis. Abbreviations: BMI – body-mass index; BP – blood pressure; HDL-C – high-density lipoprotein cholesterol; LDL-C – low-density lipoprotein cholesterol. [a] Excludes diabetic individuals; [b] Adjusted for diabetes status in addition to other covariates