Journal for the Measurement of Physical Behaviour, (Ahead of Print)
https://doi.org/10.1123/jmpb.2019-0016
© 2019 Human Kinetics, Inc.

Human Kinetics
ORIGINAL RESEARCH

Step Counts From Satellites: Methods for Integrating Accelerometer and GPS Data for More Accurate Measures of Pedestrian Travel

Brian M. Wood
University of California,
Los Angeles and
Max Planck Institute for Evolutionary Anthropology

Herman Pontzer
Duke University

Jacob A. Harris
University of California,
Los Angeles

Audax Z.P. Mabulla
University of Dar es Salaam

Marc Hamilton and Ted Zderic
University of Houston

Bret A. Beheim
Max Planck Institute for Evolutionary Anthropology

David A. Raichlen
University of Southern California

The rapid adoption of lightweight activity tracking sensors demonstrates that precise measures of physical activity hold great value for a wide variety of applications. The corresponding growth of physical activity data creates an urgent need for methods to integrate such data. In this paper, we demonstrate methods for 1) synchronizing accelerometer and Global Positioning System (GPS) data with optimal corrections for device-related time drift, and 2) producing principled estimates of step counts from GPS data. These methods improve the accuracy of time-resolved physical activity measures and permit pedestrian travel from either sensor to be expressed in terms of a common currency, step counts. We show that sensor-based estimates of step length correspond well with expectations based on independent measures, and functional relationships between step length, height, and movement speed expected from biomechanical models. Using 123 person-days of data in which Hadza hunter-gatherers wore both GPS devices and accelerometers, we find that GPS-based estimates of daily step counts have a good correspondence with accelerometer-recorded values. A multivariate linear model predicting daily step counts from distance walked, mean movement speed, and height has an R^2 value of 0.96 and a mean absolute percent error of 16.8% (mean absolute error = 1,354 steps; mean steps per day = 15,800; $n = 123$). To best represent step count estimation error, we fit a Bayesian model and plot the distributions of step count estimates it generates. Our methods more accurately situate accelerometer-based measures of physical activity in space and time, and provide new avenues for comparative research in biomechanics and human movement ecology.

Keywords: Hadza, hunter-gatherer, pedometer, time drift

Behavioral research is increasingly carried out with the aid of lightweight activity tracking sensors. The quality of such research depends upon the validation of individual sensors, and consideration of how data streams from different sensors can best be integrated. Fine scale measures of pedestrian travel can be gathered with relative ease using either accelerometers or Global Positioning System (GPS) devices (Chen, Janz, Zhu, & Brychta, 2012; Terrier & Schutz, 2005). Ideally, data from these sources could be integrated to enable greater comparative research and build on the specific advantages of each type of sensor (Duncan, Badland, & Mummery, 2009). The methods described here were developed with such a goal in mind; that is, to permit us to synchronize accelerometers with GPS devices, and estimate step counts from GPS data.

As part of our ongoing anthropological research with Hadza hunter-gatherers of northern Tanzania, we have collected a large amount (>2,000 person days) of GPS data recording the daily pedestrian travel of research participants. A central aim of our work is to compare the Hadza's patterns of physical activity, energy use, and space use to that found in other societies (Pontzer et al., 2012; Raichlen et al., 2016). In the largest global database of pedestrian travel that we are aware of (Althoff, Hicks, King, Delp, & Leskovec, 2017), daily step counts representing the Q2

Wood and Harris are with the Department of Anthropology, University of California, Los Angeles. Wood is also with, and Beheim is with, the Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Pontzer is with the Department of Evolutionary Anthropology, Global Health Institute, Duke University, Durham, NC. Mabulla is with the Department of Archaeology, University of Dar es Salaam, Dar es Salaam, Tanzania. Hamilton and Zderic are with the Texas Obesity Research Center and Health and Human Performance, University of Houston, Houston, TX. Raichlen is with the Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA. Wood (brianwood@anthro.ucla.edu) is corresponding author. Q1

AUTHOR PROOF

pedestrian travel of 717,527 people in 111 nations are recorded. To build a comparison between this large dataset and the Hadza, we were motivated to build a method for estimating step counts from our database of GPS tracks. To develop this method, we analyze here data collected during a seven-day period in 2015, during which Hadza research participants simultaneously wore both GPS devices and accelerometers. These data provide us with a valuable opportunity to develop and evaluate a model of step count estimation using GPS tracks. Our method for translating GPS data into estimates of step counts has wide applicability beyond our research, and we are aware of no prior study that has developed such a method.

While developing our step count estimation method, a major problem we encountered is the phenomenon of accelerometer time drift. Time drift reduces the accuracy of accelerometer time stamps and creates synchronization errors if accelerometer data are merged with GPS data using the devices' raw time stamps. To enable more accurate synchronization of data, we have developed software to estimate and correct for accelerometer timestamp error.

Methods for accurately synchronizing GPS with accelerometer data are generally useful for placing physical activity patterns into spatial context, as is done in research in public health, transportation planning, geography, and anthropology (e.g. Chen et al., 2012; Gordon-Larsen, Nelson, Page, & Popkin, 2006; Handy, Boarnet, Ewing, & Killingsworth, 2002). Some methods for combing GPS and accelerometer data sources have been reported (Almanza, Jerrett, Dunton, Seto, & Pentz, 2012; Mackett, Brown, Gong, Kitazawa, & Paskins, 2007; Oliver, Badland, Mavoa, Duncan, & Duncan, 2010; Rodríguez, Brown, & Troped, 2005; Terrier, Ladetto, Merminod, & Schutz, 2000; Troped et al., 2008). The majority of systems we reviewed ignored time drift error, and a minority corrected it using visually aided hand re-alignments of plotted data. Synchronizing sensor data by shifting, stretching, or compressing plots of data by hand is not ideal; with large datasets this could require a great deal of analyst time, and such workflows make replicating analyses challenging. For these reasons, we have developed an automated software approach, and distribute it freely to aid future research.

To learn the best way to estimate step counts from GPS data, it is worth considering biomechanical influences upon step length and step counts. The distance that a person walks during the course of a day should be positively and strongly correlated with the number of steps a person takes. In addition, biomechanical models predict that walking speed and subject height should also be important, because faster walking and taller subjects take longer steps (Alexander, 1984, Weyand, Smith, Puyau, & Butte, 2010). Therefore, given a fixed walked distance, height and walking speed should negatively correlate with step counts. Step count estimates based on GPS data might improve if they included, along with total distances walked, measures of movement speed and subject height. The influence of these variables is clear under controlled experimental conditions, but their inclusion in a model may not have a major impact on results given the imperfect nature of GPS data. Noise in GPS signals might overshadow any signal of the influence of subject height and walking speed upon step lengths or step counts. Alternatively, these biomechanical influences might be detectable, and therefore worth including in models estimating step counts. We therefore analyze our synchronized data to determine whether the predicted influences of subject height and movement speed are detectable in multivariate regression models predicting step length and step counts.

In the methods section below, we first describe how our GPS, accelerometer, and anthropometric data were collected. We then detail the problem of time drift, and our novel synchronization algorithm. We also provide more details of our statistical analyses. The data, variables, and structure of all statistical models are also listed in Table S1 of the Supplementary Materials (available online). In the results section, we first illustrate the patterns of time drift that exist in the accelerometer data. We then fit and report details of statistical models to estimate step lengths and step counts. Finally, we plot the estimates of a Bayesian model predicting daily step counts from measures of distance traveled, average movement speed, and subject height. In the discussion section, we describe limitations and extensions of our methods.

Methods

Study Location and Data Collection

The data in this paper were collected as part of anthropological study of Hadza hunter-gatherers of northern Tanzania (Wood & Marlowe, 2014; Pontzer et al., 2015; Raichlen et al., 2014). Rich descriptions of the Hadza people and their hunting and gathering lifestyle can be found in several monographs (Blurton Jones, 2016; Marlowe, 2010). Our research protocol was approved by human subject committees at Hunter College, the University of Arizona, Arizona State University, and Yale University. Our research also had approval from the Tanzanian Commission for Science and Technology (COSTECH) and the National Institute for Medical Research (NIMR). Participants gave informed consent prior to participation.

The field research was carried out in July of 2015 in a Hadza camp in a remote area of northern Tanzania. The sample collected here was gathered within the purview of an ongoing study of space use and physical activity, which has produced a large corpus of GPS tracks describing individual movement (Raichlen et al., 2014). During the period of data collection reported here, ActivPal Professional accelerometers (Pal Technologies, LTD, Scotland) were placed on 19 research participants, each of whom gave their informed consent to wear such devices for a seven-day study period. While a seven-day study period was used, it is important to note that unlike market-integrated societies, the Hadza do not follow a weekly cycle of work, and the term "weekend" has no meaning in Hadza culture or have any influence on how they schedule daily life. ActivPal accelerometers have been validated against video measures and pedometers to accurately measure step counts (Ryan, Grant, Tigbe, & Granat, 2006). Q3

The accelerometers were worn continuously, while GPS devices were placed on subjects in the early morning (~7 AM) and removed in the early evening (~7 PM). The ActivPal devices were encased in waterproof tape and affixed to the wearer's left thigh so as to permit the accurate measurement of posture and movement. GPS devices were attached to carabiners which were looped onto belts or fabric that were worn by the research participants. Research participants carried out their normal hunting and gathering activities during the course of the study period. Men and women foraged daily for meat, honey, and plant foods, collected water, collected firewood, and engaged in a variety of domestic activities in camp. The sensor data from the day in which the accelerometers were applied was not analyzed, because the transport and handling of the devices reduced the data quality. We were unable to compile a complete dataset that included all 19 subjects wearing both GPS and accelerometers on all seven days of the study. On four occasions, participants left camp in the morning prior to receiving their GPS devices. On one occasion, a

AUTHOR PROOF

GPS device malfunctioned and did not record data. In the resulting 128 person-days of free-living data, five were excluded from analysis, because they record a day in which five participants were driven to an important village meeting using the research vehicle.

All the GPS-recorded travel in the analyzed sample is thus pedestrian travel. The final dataset consists of 123 person days (1511.6 hours) in which GPS and accelerometers were worn simultaneously. In the analyzed dataset, two of the 19 participants contributed five days of GPS and accelerometer data, six participants contributed six days of data, and 11 contributed a full seven days of GPS and accelerometer data. When exported, the accelerometer data enumerated step counts per 15-second epoch. The GPS data were set to record locations of their wearer every five seconds. For privacy concerns, no personally identifying information about our research participants is shared.

The Problem of Time Drift

To understand best practices for synchronizing accelerometer and GPS data, it is useful to review each sensor's time keeping system. Time drift, known to occur in accelerometers (see Barreira, Zderic, Schuna, & Hamilton, 2015), arises because these sensors store a single timestamp at the moment of initialization, and other internally stored timestamps are generated as offsets from this master timestamp. Owing to slight imprecision in internal timekeeping, more distal timestamps lose absolute accuracy. Two sources of error affect accelerometer time stamps: whether the computer time used at device initialization is correct, and additionally, the time drift that distorts timestamps subsequent to device initialization. Our method estimates and corrects for both sources of asynchrony. GPS data are not subject to either source of error. This is because GPS devices continually receive timestamps directly from the atomic clocks onboard GPS satellites (Hofmann-Wellenhof, Lichtenegger, & Collins, 2012).

Our technique for optimally estimating and correcting for error in accelerometer time stamps is based on a simple premise. As a starting point, we surmise that if no time error existed in accelerometry data, then accelerometer-measured step counts and GPS-measured distances traveled should be highly correlated across time for a single pedestrian traveler. Therefore, a linear regression model predicting distances traveled (GPS) from steps taken (accelerometer) should have a high goodness-of-fit, or R^2 value. As time drift injects error into the accelerometer timestamps, the goodness-of-fit of such models using uncorrected data is expected to degrade. If this simple principle holds, then R^2 values of linear regression models incorporating varying degrees of time drift correction can be used to search for the optimal time drift correction, enabling accelerometer data to be synched to GPS / atomic time. The software implementation of our method enables fast processing of data and replication of analyses. We share our source code freely to encourage others to use or improve upon our method.

Data Synchronization Algorithm

The software (written in R, version 3.3.3) we developed to Q4 optimally synchronize GPS and accelerometer data can be downloaded here: https://osf.io/5z8q2. The algorithm is described in pseudocode (see Table 1).

A few notes about this workflow may add clarity. Regarding step 3, the range of candidate drift offsets to explore is study-dependent. In our case, according to Pal Technologies, devices could be expected to lose about 3 seconds of absolute accuracy per day during use, and that time drift could vary across devices. Based on this information, we estimated that over the course of a seven-day study, we could expect approx. 20 seconds of drift. To be conservative, in our testing, we explored a range of drift offsets from −100 to 100 seconds.

Regarding steps 5–7, here our method summarizes distances walked within time segments whose duration matches the recording epoch of the accelerometers (15 seconds). The step counts within each epoch are never changed, but the times that define the epoch are systematically adjusted across a range of candidate synchronizations, offset from the raw accelerometer time stamps. The process of summarizing the GPS data to produce a measure of distance traveled within any 15-second epoch is shown in Figure 1.

In our case, GPS devices recorded locations at five-second intervals (Figure 1A) while accelerometer data provided step counts summaries within 15-second epochs (Figure 1B), and the GPS epochs were not subperiodic with the accelerometer epochs. In order to create summaries of GPS-recorded movement within accelerometer epochs, we first injected pseudo-trackpoints into the GPS data at the starting times of each 15-second accelerometer epoch, using linear interpolation (Figure 1C). It is important to note that the raw GPS data were never downsampled; prior to being summarized, the GPS dataset included both the original trackpoints and the interpolated points. Within each 15-second track segment, we then summarized how many meters of travel occurred (Figure 1D). The resulting intermediate dataset, or candidate synchronization, then listed, for every accelerometer epoch, a count of steps taken and a measure of distance walked. Each of these candidate synchronizations was then fed into a linear regression model in which step count was estimated from distance traveled, and the R^2 value of each model was saved. In Figure 2, we plot the R^2 values that resulted from linear regression models upon a range of candidate

Table 1 Pseudocode of Algorithm to Synchronize GPS With Accelerometer Data

Q5 1. For each person-day of data:
    2. Get the uncorrected time series defining the start of each accelerometer epoch.
    3. Add a range of drift offsets (e.g., from −100 to 100 seconds) to these start times, forming a set of vectors of adjusted epoch times.
    4. For each vector of adjusted epoch times:
        5. Inject interpolated trackpoints into the GPS data at each adjusted epoch time, forming track segments.
        6. For each track segment:
            7. Sum the distance traveled, forming a candidate synchronization.
    8. Perform linear regression using the candidate synchronization, estimating step counts from distances traveled, and save the R^2 value of the fit model.
9. Select the candidate synchronization that produced the highest R^2 value.

AUTHOR PROOF

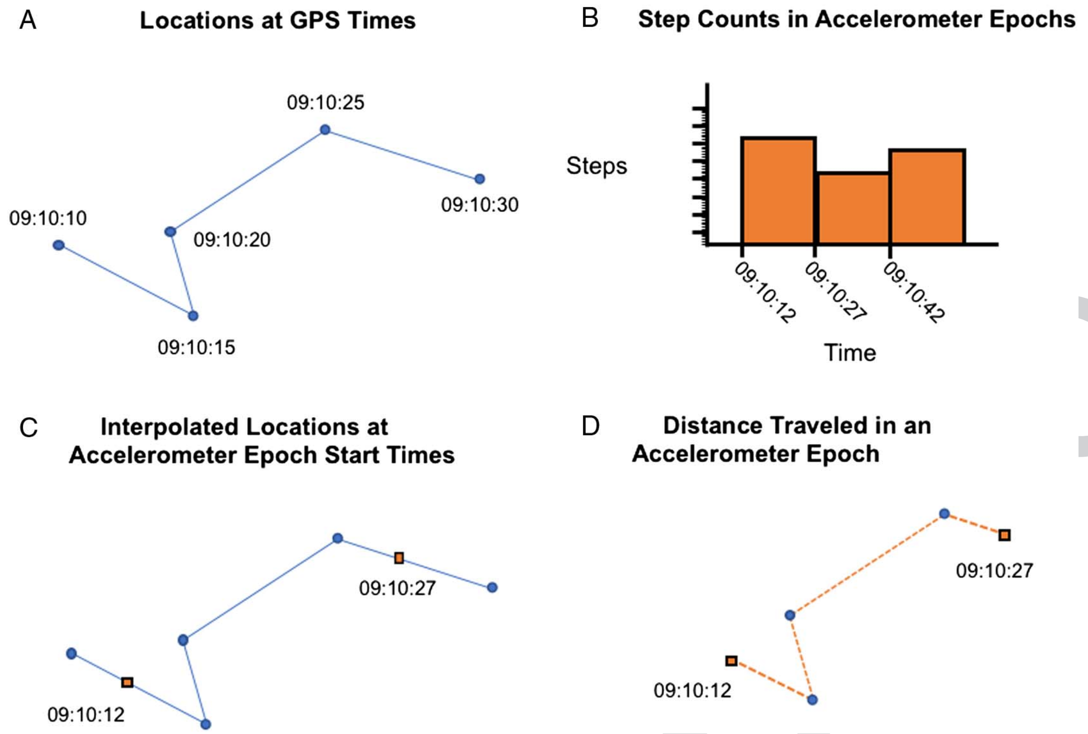

Figure 1 — How distances traveled within accelerometer epochs (raw or adjusted) were summarized. A) GPS locations were recorded every five seconds; B) step counts were summarized into 15-second epochs; C) interpolated locations were calculated at the start time of each accelerometer epoch; D) distances traveled within each accelerometer epoch were summarized and joined with the step counts.

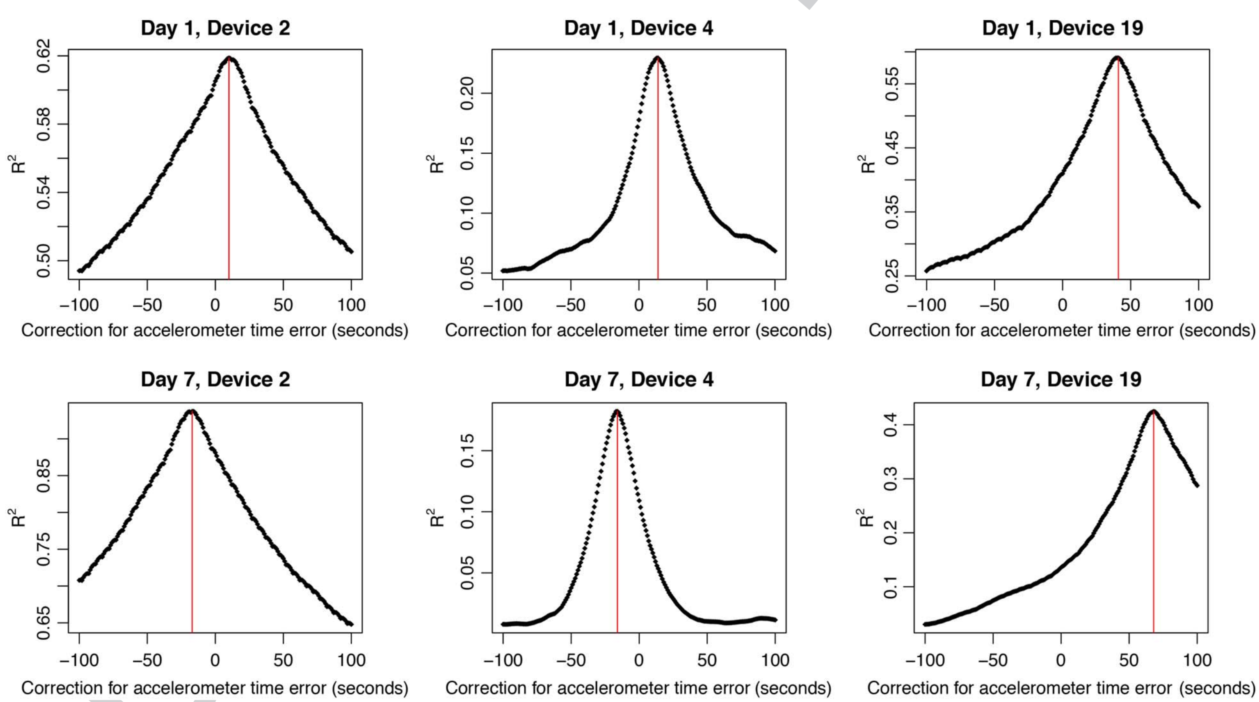

Figure 2 — R^2 values derived from candidate synchronizations of 6 person-days of data, ranging from –100 to 100 seconds of error correction. The correction on each day with the highest R^2 value is shown in red. Note that the correction needed for the same device varies across the study, owing to accelerometer time drift.

synchronizations on six person-days of data. Identifying optimal corrections for accelerometer time error is relatively straightforward using this method.

As seen in Figure 2, the R^2 values in these plots display clearly defined peaks across candidate synchronizations. The unimodal distributions and clear peaks give us confidence that our estimation process is indeed finding the most accurate synchronization possible.

We applied the algorithm illustrated in Figure 1, Figure 2, and Table 1 to the entire dataset, thus identifying the time drift that affected each device across the days of our study. In the results section, we plot the time drift of each device across the study and calculate rates of drift for each device using linear regression. To evaluate our synchronization method, we also compare the goodness of fit of regression models predicting step counts from distances traveled using raw accelerometer data to those using synchronized data.

Analysis of Synchronized GPS-Accelerometer Dataset

Here, we explain details of our analyses to examine how well the synchronized dataset conforms to theoretical and empirical expectations, and our method for estimating step counts from GPS data. Because our data include repeated measures of individuals, we adopted a multilevel modeling approach with varying effect terms for each individual. All models include varying intercepts by individual. Where possible, we also specified varying slopes by individual, but problems with model convergence did not permit such a structure in all cases (see Supplemental Materials, Table S1 [available online] for details). The R^2 values of models in this study are computed
Q6 following the method of Raudenbush and Bryk (2002), as implemented in the R package 'mitml' (Grund, Robitzsch, & Luedtke, 2019), and only consider the variance accounted for by fixed effects (i.e. variance at level 1). The *p*-values of model coefficients are estimated using Satterwaite's method as implemented in the R package 'lmerTest' (Kuznetsova, Brockhoff, & Christensen, 2017).

We first examine the univariate relationship between daily distance traveled and steps taken, across 123 person-days of data, using a multilevel regression model (M1). Next, we examine whether, as biomechanical models predict, height (cm) and walking speed (meters / second) are positively associated with calculated step length (m). To do so, we fit a multilevel model (M2) to the most detailed data representation, that is, with the data summarized into 362,784 optimally synchronized 15-second epochs. For each epoch, we calculated the average step length by dividing the distance walked by the numbers of steps logged. In epochs with zero movement, step length was scored as zero.

We check the validity of the synchronized sensor data by estimating the linear relationship between movement speed and step length in the sensor data (M3) and comparing the result of this analysis to a model (M4) fit to experimental data collected under controlled conditions, during a prior study of Hadza locomotion (Pontzer et al., 2014). In that prior study, we experimentally measured step lengths for men and women ($n = 199$ trials, 54 subjects, 36 men, 19 women) walking on level ground at speeds ranging from 0.84 to 2.08 meters per second. In order to make the analyses maximally comparable, we fit M3 to epochs of movement within this same range of walking speeds.

We next fit a multilevel regression model for the prediction of daily step counts from GPS and anthropometric data (M5). The predictor variables of this model include: 1) distance traveled per day (meters); 2) average movement speed (meters / second); and 3) height of the GPS-wearing research participant (cm). To evaluate the accuracy of this model, we calculate its R^2 value, mean average error (MAE) and mean average percent error (MAPE).

In order to best represent model estimation uncertainty, we then adopt a Bayesian approach, and fit a multilevel model (M6) using the R package 'brms' (Bürkner, 2017), using uniform Q7
priors. In this model, varying slopes and intercepts were fit to each individual. In terms of fixed effects, this model includes the same predictor variables as M5 just described above. The parameter estimates were calculated using 4,000 post-warmup samples. Using this fit model, we plot posterior distributions of model estimates, and their relationship to distance traveled, mean speed, and height.

Results

We first examine patterns of error identified in the timestamps of accelerometer data. For each accelerometer examined, the rate of drift (seconds / day) was consistent across the seven-day wear period (Figure 3). However, we found that the direction of drift differed across devices. On day 1 of the study, all devices started with baseline timestamps that were earlier than GPS time. This is due to a difference between GPS time and clock time on the computer used to initialize the devices. Thereafter, some of the device's clocks were slower than GPS clocks (devices 10–19), and thus require increasingly positive drift corrections, while other device clocks (1–9) were faster than GPS clocks.

Correspondence of GPS and Accelerometer-Based Measures of Movement

Correcting for accelerometer time drift led to a notable increase in the correspondence of accelerometer and GPS measures of movement. The average R^2 value of 123 within-day linear models estimating step counts from distances traveled was 0.58 ($SD = 0.28$) before correcting for time drift and 0.78 (SD = 0.22) after correcting for it. Using the synchronized data, we found that there was a tight relationship, across days, between total distance traveled and total step counts:

Figure 4 plots the relationship between daily distance traveled and steps taken, across 123 person-days. The multilevel model fit to these data estimates that across the sample, 1,445 steps were taken per kilometer, or 0.69 meters per step, which is a biologically reasonable estimate for a short-statured human population (Pontzer et al., 2014).

Movement Speed, Height, and Step Length Estimates

Experimental studies and biomechanical models show that subject height and walking speed are positively associated with step length (Alexander, 1984; Weyand et al., 2010). Here, we investigate whether these relationships are detectable in our synchronized GPS-accelerometer dataset. The results are shown in Table 2.

Below, using linear regression, we plot the univariate relationship between walking speed and step length estimated from an experimental dataset collected during a prior study (Pontzer et al., 2014) and the synchronized sensor dataset of the current study.

As seen in Figure 5, the linear relationship between walking speed and step length is remarkably similar when estimated from either the experimental data or the sensor data, suggesting that analyses of synchronized sensors can accurately detect biomechanical influences upon gait. While the parameter estimates are similar, there is greater noise in the free-living, sensor-recorded data. The R^2 value of the model fit to experimental data (M3) is 0.84 and that of the sensor data (M4) is 0.11. In both models, speed is significantly

AUTHOR PROOF

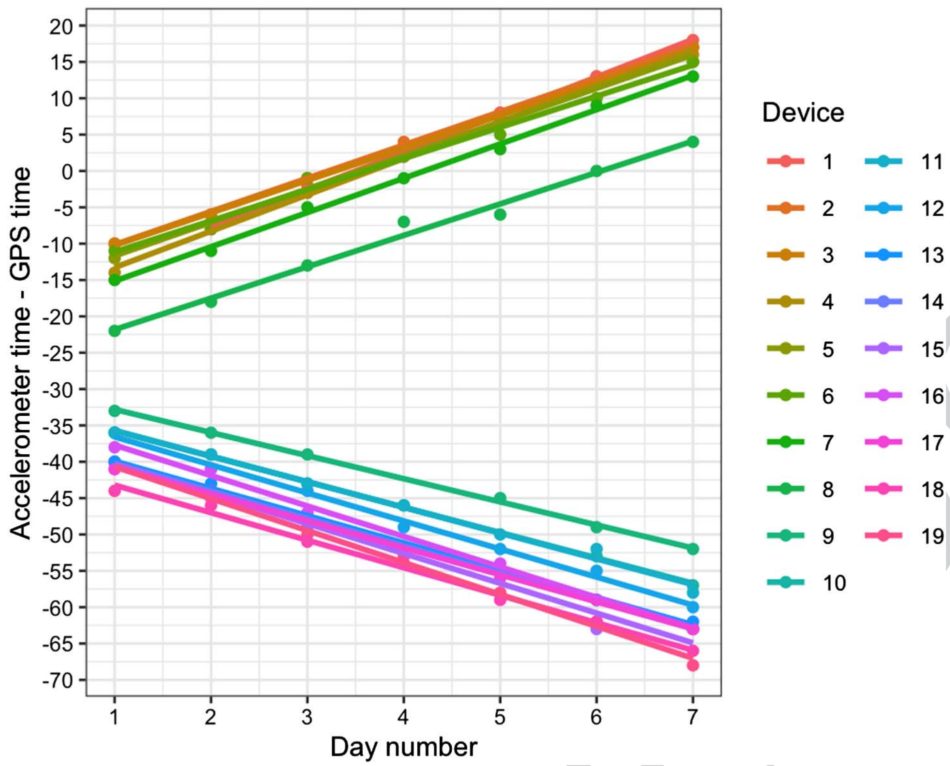

Figure 3 — Differences between accelerometer time codes and GPS time codes, as estimated by our algorithm. Lines represent linear regression models fit to the data of each accelerometer, across days of the study. The R^2 values of these linear models average 0.995 (range 0.980–.999). On day 1, the asynchrony of times is owing to differences between GPS time and the computer time used to initialize the accelerometers. Thereafter, further shifts are owing to accelerometer time drift.

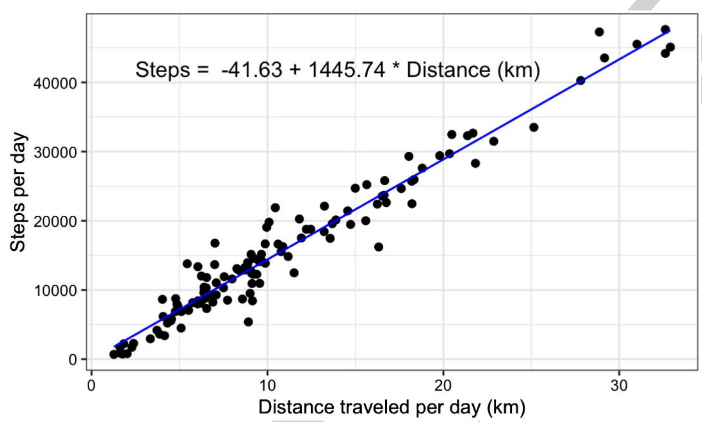

Figure 4 — Total distance traveled (from GPS) and steps logged per day (from accelerometers). The fit linear relationship from the multilevel model is shown in blue. The R^2 value of this model is 0.95.

and positively associated with step length ($\beta_{experimental} = 0.3$, $p < .001$; $\beta_{sensors} = 0.31$, $p < .001$).

Estimation of Step Counts From GPS Data

We describe here details of our multivariate model that estimates daily step counts from GPS data and height, and it's correspondence to accelerometer-recorded step counts. In this dataset, research participants traveled 10,955 meters per day on average ($SD = 7{,}176$, n = 123) at a mean speed of 0.26 meters per second ($SD = 0.17$, $n = 123$).

Table 2 Fixed Effects From a Multilevel Regression Model (M2) of Step Length Fit to the Synchronized GPS and Accelerometry Data

Term	Estimate	Std. Error	*p*-value
(Intercept)	–0.0353	0.0237	.1537
Height (cm)	0.0003	0.0002	.0512
Walking speed (m/s)	0.5281	0.0097	<.001

Note. In this model, step length is estimated as a function of height and walking speed of subjects. Observations: 362,784 15-second intervals, nested within 19 research participants. $R^2 = 0.68$.

The model estimating step counts from subject height, distance traveled, and mean speed has a very high goodness of fit ($R^2 = 0.96$). The mean absolute difference between model predictions and accelerometer-recorded step counts is 1,354, which is 8.6% of the average steps per day (15,800). The mean absolute percent error of the model is 16.8%. The parameter estimates of this model show that distance traveled and mean speed have strong influences on estimated step counts, as expected from biomechanical models. The parameter estimate for influence of height is in the direction expected, but is not statistically significant. It is important to note that the sample analyzed here includes only adult subjects (Table 3), and therefore height variation is limited relative to what would be found in a more complete demographic sample including both children and adults.

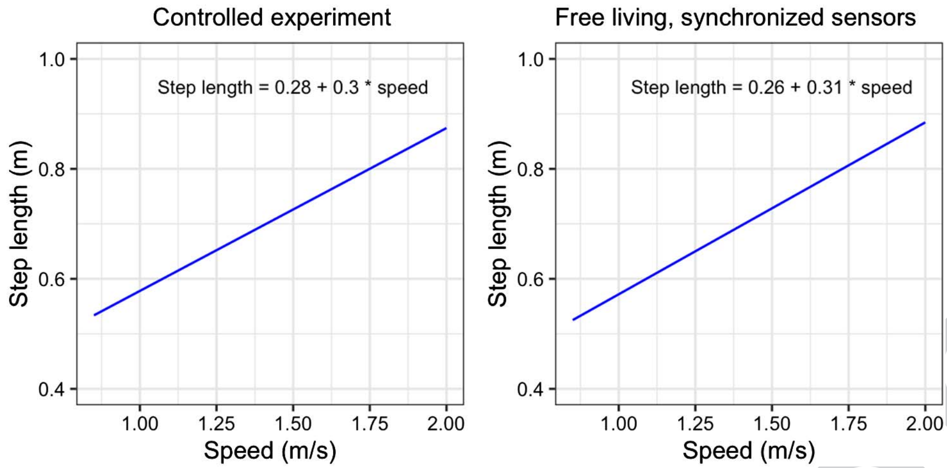

Figure 5 — The relationship between walking speed and step length estimated from A) experimental measures and B) the synchronized accelerometer-GPS dataset. The regression equation and parameter estimates of each fit model (M3, M4) are displayed on the plot.

Table 3 Research Participants in Study, Who Produced 123 Person Days of Data in Which Both GPS Devices (Garmin E-trex) and Accelerometers (ActivPal) Were Worn Q9

Gender	Age *M* (*SD*)	Height, cm *M* (*SD*)	*N* Participants	*N* Days	Hours per day *M* (*SD*)
Female	33 (15.6)	149.6 (6.9)	10	63	12.4 (0.3)
Male	35 (13.5)	162.7 (6.4)	9	60	12.2 (0.3)

Note. Hours per day refers to hours both devices were worn per day. *M* (*SD*) = Mean (Standard Deviation).

Table 4 Multilevel Model Estimating Step Counts From GPS and Anthropometric Data*

Term	Estimate	Std. Error	*p*-value
(Intercept)	4294.218	5516.5	.447
Distance walked	2004.553	119.457	<.001
Mean speed	−24004.574	5010.346	<.001
height	−27.412	35.47	.450
$R^2 = 0.96$	Mean absolute error (MAE) = 1,354	Mean absolute percent error (MAPE) = 16.8%	

* Fixed effects from Model 5, estimating step counts from GPS measures and height.

To visualize model uncertainty, we now plot posterior predictions of a Bayesian model that includes the same predictor variables as those listed in Table 4. In the plots below, we represent the stand-alone effects of subject height, distance traveled, and average speed, when statistically controlling for the other variables, by holding them at their mean values.

Posterior predictions of the Bayesian model illustrate the strong effects of distance walked and mean movement speed, and the weaker and more variable effect of subject height on step count.

Discussion and Conclusion

In this study, we demonstrated that complex patterns of time error arise in accelerometry data, and we have provided an automated method for optimally estimating and correcting such error. The simple method we have described here will hopefully enable the assessment and correction of accelerometer time drift in future studies. Given the observed differences in time drift error among accelerometers (see Figures 2 and 3), it would not be advisable for researchers to apply a single time drift correction across devices, nor to assume that device clocks will be invariably faster or slower than absolute time. These inter-device differences underscore the importance of using automated methods to estimate and correct for accelerometer time drift in a device-specific manner. Our synchronization method holds promise for spatial research in public health. For example, Zhao, Kwan, and Zhou (2018) recently advocated for greater spatial precision in studies examining neighborhood effects or other geographic influences upon physical activity.

Our analysis of synchronized GPS and accelerometer data shows a strong relationship between daily distance walked and step counts, resulting in a grand average step length of 0.69 meters (Figure 4). However, step length varies across individuals and contexts. Our multivariate model (M2, Table 2) shows that faster walking speeds and taller subjects both generated longer step lengths. Thus, measures of step length derived from our synchronized GPS and accelerometer data conform to the expectations of biomechanical models (Alexander, 1984; Weyand et al., 2010). The relationship between walking speed and step length is quite pronounced, as expected. The relationship between height and step length is in the direction expected, but it is a very weak effect. Our analyses also show that quite similar estimates of the relationship between step length and walking speed arise when based either on carefully controlled experimental observations, or synchronized sensor data recording free-living subjects (Figure 5). While the best fit linear relationship between these variables has a remarkably

AUTHOR PROOF

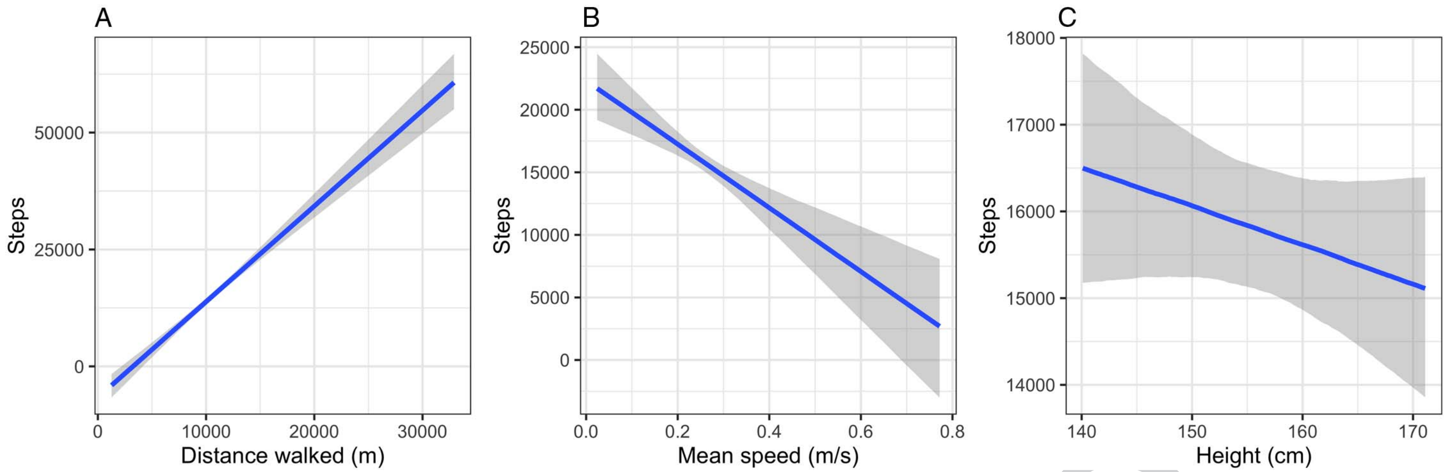

Figure 6 — The relationship between a) distance walked, b) mean speed, and c) subject height on total step counts per day, when conditioning on the other variables being at their mean values. The grey areas represent the 95% credible intervals of the Bayesian model (M6) posterior predictions, and the Q10 blue line represents the mean model prediction.

similar slope in both datasets, there is considerably more noise in the sensor dataset, and the goodness of fit of the sensor-based model is much lower.

Our multivariate model to estimate step counts using GPS data and height (M5) has a very high goodness of fit (0.96) and a moderate mean absolute percent error (16.8%). While this model does not detect a statistically significant influence of subject height, we recommend that subject height still be considered a likely influence upon step counts, especially in samples that include both adults and sub-adults.

One weakness of our study is that we do not examine the accuracy of the step count measures produced by the accelerometers in our field setting. Though all wearable accelerometers generate some misclassification error, ActivPal devices have been shown to produce accurate step count measures when subjects are walking on treadmills and overground (Grant, Dall, Mitchell, & Granat, 2008; Treacy et al., 2017). However one study has found that their accuracy declines when research subjects simulate daily life across a range of physical activities, and other devices on the market might provide more accurate measures of step counts in free-living subjects (Hickey, John, Sasaki, Mavilia, & Freedson, 2016). As a part of our larger research efforts, part of which focuses on posture, we were motivated to use an accelerometer that could measure leg position (horizontal or vertical), which is a unique feature provided by the ActivPal.

Our study was carried out in a remote rural area, where people spend most of their waking hours outdoors, and nearly all travel is done on foot. Observation conditions would be much different in an urban setting, where GPS signals are often impeded by structures, people work indoors, and where pedestrian travel is overall less frequent. These limitations do not preclude our method being applied in such contexts, but it is likely that additional data processing and analyses procedures would be called for. We share the source code of our software so that other researchers could extend or modify our method as suited.

We can imagine our method being extended to permit more accurate synchronization of audio or video data to GPS or accelerometers, where footfalls are likely to leave identifiable audio and video cues. Once coded as step counts, such data could thereafter be synchronized with GPS data or accelerometer data using the methods we have outlined. In addition, this method may be used to compare data collected from different accelerometers to determine the accuracy of new devices or wear locations.

Our findings gives us confidence that synchronized GPS and accelerometer data are reliable enough to characterize functional relationships between anatomy and behavior that structure pedestrian travel. We also hope that our sensor synchronization method, or extensions of it, will be of value to researchers investigating spatial correlates of physical activity and health.

Acknowledgments

We thank the Hadza participants for their involvement in this study. We also thank Carla Wood, Christian and Nani Schmelling, Daudi Peterson, Richard McElreath, and James Holland Jones. Funding was provided by the National Science Foundation [1440867 (DAR), 1440841 (HP), 1440671 (BMW)], The L.S.B. Leakey Foundation (BMW), the University of Arizona Bio5 Institute (DAR), and the John Templeton Foundation to the Institute of Human Origins at Arizona State University.

References

Alexander, R.M. (1984). Stride length and speed for adults, children, and fossil hominids. *American Journal of Physical Anthropology, 63*(1), 23–27. PubMed ID: 6422766 doi:10.1002/ajpa.1330630105 Q11

Almanza, E., Jerrett, M., Dunton, G., Seto, E., & Pentz, M.A. (2012). A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health & Place, 18*(1), 46–54. PubMed ID: 22243906 doi:10.1016/j.healthplace.2011.09.003

Althoff, T., Hicks, J.L., King, A.C., Delp, S.L., & Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature, 547*(7663), 336–339. PubMed ID: 28693034 doi:10.1038/nature23018

Barreira, T.V., Zderic, T.W., Schuna, J.M., & Hamilton, M.T. (2015). Free-living activity counts-derived breaks in sedentary time: Are they real transitions from sitting to standing? *Gait & Posture, 42,* 70–72. doi:10.1016/j.gaitpost.2015.04.008

Blurton Jones, N. (2016). *Demography and evolutionary ecology of Hadza hunter-gatherers*. Cambridge University Press. Q12

Bürkner, P.C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

AUTHOR PROOF

Chen, K.Y., Janz, K.F., Zhu, W., & Brychta, R.J. (2012). Re-defining the roles of sensors in objective physical activity monitoring. *Medicine & Science in Sports & Exercise. 44*(1): S13–S23. doi:10.1249/MSS.0b013e3182399bc8

Duncan, M.J., Badland, H.M., & Mummery, W.K. (2009). Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport, 12,* 549–556. PubMed ID: 19237315 doi:10.1016/j.jsams.2008.10.010

Gabry, J., & Mahr, T. (2018). *bayesplot: Plotting for Bayesian Models. R package version 1.6.0*. Retrieved from https://CRAN.R-project.org/package=bayesplot

Gatesy, S.M., & Biewener, A.A. (1991). Bipedal locomotion: effects of speed, size and limb posture in birds and humans. *Journal of Zoology, 224*(1), 127–147. doi:10.1111/j.1469-7998.1991.tb04794.x

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). *Bayesian data analysis*. London, UK: Chapman & Hall/CRC Press.

Gordon-Larsen, P., Nelson, M.C., Page, P., & Popkin, B.M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics, 117*(2), 417–424. PubMed ID: 16452361 doi:10.1542/peds.2005-0058

Grant, P.M., Dall, P.M., Mitchell, S.L., & Granat, M.H. (2008). Activity-monitor accuracy in measuring step number and cadence in community-dwelling older adults. *Journal of Aging and Physical Activity, 16*(2), 201–214. PubMed ID: 18483442 doi:10.1123/japa.16.2.201

Grund, S., Robitzsch, A., & Luedtke, O. (2019). *mitml: Tools for Multiple Imputation in Multilevel Modeling. R package version 0.3-7*. Retrieved from https://CRAN.R-project.org/package=mitml

Handy, S.L., Boarnet, M.G., Ewing, R., & Killingsworth, R.E. (2002). How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine, 23*(2), 64–73. doi:10.1016/S0749-3797(02)00475-0

Hickey, A., John, D., Sasaki, J.E., Mavilia, M., & Freedson, P. (2016). Validity of activity monitor step detection is related to movement patterns. *Journal of Physical Activity and Health, 13*(2), 145–153. PubMed ID: 26107045 doi:10.1123/jpah.2015-0203

Hofmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (2012). *Global positioning system: theory and practice*. New York, NY: Springer Science & Business Media.

Hooten, M.B., & Hobbs, N.T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs, 85*(1), 3–28. doi:10.1890/14-0661.1

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. doi:10.18637/jss.v082.i13

Mackett, R., Brown, B., Gong, Y., Kitazawa, K., & Paskins, J., (2007). Children's independent movement in the local environment. *Built Environment, 33*(4), 454–468. doi:10.2148/benv.33.4.454

Marlowe, F. (2010). *The Hadza: Hunter-gatherers of Tanzania*. Berkeley, CA: University of California Press.

Oliver, M., Badland, H., Mavoa, S., Duncan, M.J., & Duncan, S. (2010). Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. *Journal of Physical Activity and Health, 7*(1), 102–108. PubMed ID: 20231761 doi:10.1123/jpah.7.1.102

Pontzer, H., Raichlen, D., Wood, B., Emery Thompson, M., Racette, S., Mabulla, A., & Marlowe, F. (2015). Energy expenditure and activity among Hadza hunter-gatherers. *American Journal of Human Biology, 27*(5), 628–637. PubMed ID: 25824106 doi:10.1002/ajhb.22711

Pontzer, H., Suchman, K., Raichlen, D.A., Wood, B.M., Mabulla, A.Z.P., & Marlowe, F.W. (2014). Foot strike patterns and hind limb joint angles during running in Hadza hunter-gatherers. *Journal of Sport and Health Science, 3*(2), 95–101. doi:10.1016/j.jshs.2014.03.010

Raichlen, D., Pontzer, H., Harris, J., Mabulla, A., Marlowe, F., Snodgrass, J., . . . Wood, B. (2016). Physical activity patterns and biomarkers of cardiovascular disease risk in hunter-gatherers. *American Journal of Human Biology, 29*(2), e22919. doi:10.1002/ajhb.22919

Raichlen, D., Wood, B., Gordon, A., Mabulla, A., Marlowe, F., & Pontzer, H. (2014). Evidence of Lévy Walk Foraging Patterns in Human Hunter-Gatherers. *Proceedings of the National Academy of Sciences, 111*(2), 728–733. doi:10.1073/pnas.1318616111

Rodríguez, D.A., Brown, A.L., & Troped, P.J. (2005). Portable global positioning units to complement accelerometry-based physical activity monitors. *Medicine & Science in Sports & Exercise, 37*(11), S572–S581. doi:10.1249/01.mss.0000185297.72328.ce

Ryan, C.G., Grant, P.M., Tigbe, W.W., & Granat, M.H. (2006). The validity and reliability of a novel activity monitor as a measure of walking. *British Journal of Sports Medicine, 40*(9), 779–784. PubMed ID: 16825270 doi:10.1136/bjsm.2006.027276

Saelens, B.E., Sallis, J.F., Black, J.B., & Chen, D. (2003). Neighborhood-based differences in physical activity: an environment scale evaluation. *American Journal of Public Health, 93*(9), 1552–1558. PubMed ID: 12948979 doi:10.2105/AJPH.93.9.1552

Terrier, P., & Schutz, Y. (2005). How useful is satellite positioning system (GPS) to track gait parameters? A review. *Journal of NeuroEngineering and Rehabilitation, 2*(28), 1–11.

Terrier, P., Ladetto, Q., Merminod, B., & Schutz, Y. (2000). High-precision satellite positioning system as a new tool to study the biomechanics of human locomotion. *Journal of Biomechanics, 33*(12), 1717–1722. PubMed ID: 11006399 doi:10.1016/S0021-9290(00)00133-0

Treacy, D., Hassett, L., Schurr, K., Chagpar, S., Paul, S.S., & Sherrington, C. (2017). Validity of different activity monitors to count steps in an inpatient rehabilitation setting. *Physical Therapy, 97*(5), 581–588. PubMed ID: 28339904 doi:10.1093/ptj/pzx010

Troped, P.J., Oliveira, M.S., Matthews, C.E., Cromley, E.K., Melly, S.J., & Craig, B.A. (2008). Prediction of activity mode with global positioning system and accelerometer data. *Medicine & Science in Sports & Exercise, 40*(5), 972–978. PubMed ID: 18408598 doi:10.1249/MSS.0b013e318164c407

Weyand, P.G., Smith, B.R., Puyau, M.R., & Butte, N.F. (2010). The mass-specific energy cost of human walking is set by stature. *Journal of Experimental Biology, 213*(23), 3972–3979. doi:10.1242/jeb.048199

Wood, B., & Marlowe, F. (2014). Toward a reality-based understanding of Hadza men's work. *Human Nature, 25*(4), 620–630. PubMed ID: 25315040 doi:10.1007/s12110-014-9218-z

Zhao, P., Kwan, M.P., & Zhou, S. (2018). The uncertain geographic context problem in the analysis of the relationships between obesity and the built environment in Guangzhou. *International Journal of Environmental Research and Public Health, 15*(2), 308. doi:10.3390/ijerph15020308

AUTHOR PROOF

Queries

Q1. Please ensure author information is listed correctly here and within the byline.

Q2. Please provide complete details for the reference "Pontzer et al., 2012" to be included in the reference list.

Q3. Please also provide city location for Pal Technologies

Q4. Please provide manufacturer location for R software.

Q5. As per style, single column text should be converted to a Table. Hence, Figure 1 has been converted to Table 1 and the subsequent tables have been renumbered sequentially.

Q6. Please provide complete details for the reference "Raudenbush and Bryk (2002)" to be included in the reference list.

Q7. The year of publication for the reference citation "Bürkner, 2019" has been changed to "2017" to match with the reference in the list. Please verify.

Q8. In Table 2 footnote, “R^2 0.68” has been changed to “$R^2 = 0.68$”. Please confirm change.

Q9. Original Figure 1 has been converted to Table 1. Considering the order of citations in the text, the original Tables 1 and 3 have been converted to Tables 3 and 4. Please verify.

Q10. Please add in-text citation for Figure 6.

Q11. Please provide in-text citation for the reference "Gabry & Mahr (2018)", "Gatesy & Biewener (1991)", "Gelman et al. (2013)", "Hooten & Hobbs (2015)", and "Saelens et al. (2003)".

Q12. Please provide publisher location for reference "Blurton Jones (2016)".

A1. This is correct
A2. Pontzer, H., Raichlen, D.A., Wood, B.M., Mabulla, A.Z., Racette, S.B. and Marlowe, F.W., 2012. Hunter-gatherer energetics and human obesity. PloS one, 7(7), p.e40503.
A3. Glasgow, Scotland
A4. Vienna is the location for the foundation that oversees R development, but R is free software and it is not manufactured in any given city. It is written by teams of people and it is downloaded from a global network of computers. It would be unusual to list a city suggesting it was manufactured there. However, if you need to list a city for some reason, Vienna would be the choice. This is the standard citation for R:
R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
A5. OK
A6. Raudenbush, S.W. and Bryk, A.S., 2002. Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.
A7. OK.
A8. OK.
A9. OK.
A10. In the penultimate paragraph of Results, please replace the text “In the plots below” with “In figure 6”
A11. Please remove the entries in the references for Gabry & Mahr (2018); please remove the entry for Gatesy & Biewener (1991); please remove entry for Hooten and Hobbs; please remove entry for Gelman et al. 2013. Please add a citation to “Saelens et al. (2013)” to the list of citation that immediately follows the phrase “as is done in research in public health, transportation planning, geography, and anthropology”
A12. Cambridge, United Kingdom

AUTHOR PROOF