Classifying Gait Behavior in 6-20 Year Olds: A Logistic Regression Approach Using Cadence to Predict the Walk-to-Run Transition

Scott W. Ducharme1, Dusty S. Turner2, James D. Pleuss2, Aston K. McCullough1, Christopher C. Moore1, John M. Schuna Jr3, Diana M. Thomas2, Catrine Tudor-Locke1\*, & Elroy J. Aguiar1

*1Physical Activity and Health Laboratory, Department of Kinesiology, University of Massachusetts Amherst, Amherst, Massachusetts*

*2Department of Mathematical Sciences, United States Military Academy-West Point, West Point, New York*

*3School of Biological and Population Health Sciences, Oregon State University,*

*Corvallis, Oregon*

**\*Corresponding author:**

Catrine Tudor-Locke, PhD, FNAK, FACSM

Department of Kinesiology

University of Massachusetts Amherst

160A Totman Building

30 Eastman Lane

Amherst, MA, 01003

Phone: +1 413 577 4702

Email: [ctudorlocke@umass.edu](mailto:ctudorlocke@umass.edu)

# Abstract

**Background:** The transition from walking to running occurs in adults at around 140 steps/minute. It is currently unknown when this transition occurs in children and adolescents, or to what extent individual characteristics, such as age or height, impact this threshold. Understanding which cadences correspond to walking versus running better informs physical activity research using wearable sensors.

**Purpose:** To develop a model to predict the cadences at which individuals 6-20 years old are either walking or running using age- and anthropometry-specific information.

**Methods:** Sixty-nine children and adolescents 6 to 20 years of age performed sequentially faster 5-min treadmill walking trials, starting at 0.22 m/s (i.e., 0.5 mph) and increasing by 0.22 m/s until completion of the trial during which participants freely chose to run. Steps taken during each trial were manually counted (hand tally) and converted to cadence (steps/minute). After identifying the best subset of parameters to inform this transition, a logistic regression model was developed.

**Results:** The logistic regression analysis produced a simple mathematical equation that can be used to estimate the preferred transition cadence using age, sex, height, and BMIz-score. This transition cadence ranged from 136 to 161 steps/min for the 6 to 20 year olds in this study.

**Conclusions:** The preferred transition cadence represents a simple and practical measure to characterize gait behavior from activity monitors in children, adolescents, and young adults. Moreover, herein we provide an equation and an open access online R-Shiny app that practitioners or clinicians can use to obtain an estimation of individual-specific preferred transition cadence.

**Keywords:** preferred transition speed, step frequency, locomotion, physical activity, preferred transition cadence

# Introduction

During upright locomotion, individuals generally choose to walk at relatively slow speeds (i.e., < 2.0 m/s) and run at faster speeds (i.e., > 2.0 m/s) (Alexander, 2002). When individuals progressively increase their locomotor speeds, the transition from walking to running appears spontaneous. Numerous studies have attempted to explain this preferred transition speed (PTS) phenomenon. For example, the transition to running may occur because, compared to running at a given speed, fast walking at that same speed correlates to reduced stability (Diedrich & Warren, 1995; Li, 2000), greater metabolic cost of walking (Alexander, 2002; Diedrich & Warren, 1995; Minetti, Ardigo, & Saibene, 1994), greater perceived effort (Hreljac, 1993; Minetti et al., 1994; Noble et al., 1973; Prilutsky & Gregor, 2001), or suboptimal energy substrate use (Ganley, Stock, Herman, Santello, & Willis, 2011). Mathematically, the Froude number (i.e., Fr = v2/(g\*l), where v = velocity, g = acceleration due to gravity, and l = leg length)), which models gait as an inverted pendulum and thus incorporates leg length, suggests that walking cannot occur when the centrifugal acceleration forces exceed the centripetal forces due to gravity, thereby requiring a flight phase, and thus provides a prediction for this transition (Alexander, 1989; Usherwood, 2005).

Irrespective of the mechanism, determining the threshold for this transition is important because it would allow for a more precise classification of gait behavior. Wearable sensors are becoming increasingly popular in laboratory and free-living research, as well as personal use by consumers. Many sensors provide minute-by-minute step data (i.e., cadence [steps/minute]). Whereas determining the PTS requires precise speed information, which is not available in most wearable sensors, cadence may provide a more accessible measure of gait behavior, i.e., walking versus running. Thus, the preferred transition cadence (PTC) may be a more practical and accessible outcome measure for researchers or clinicians aiming to quantify physical activity behavior, specifically, minutes per day of running.

Diedrich & Warren (1995) reported the PTC was, on average, 142.8 steps/minute in young, healthy adults 18-31 years old. A more recent study provided concurring evidence that the PTC can be accurately predicted using a stride frequency of 70.8, i.e., 141.6 steps/minute in young adults (Hansen, Kristensen, Nielsen, Voigt, & Madeleine, 2017). However, while a PTC of ~140 steps/min has been determined in adults, the PTC of children and adolescents remains unclear. Notably, in the study by Hansen et al.(2017), leg length did not affect the observed PTC. However, in children and adolescents, the range of leg lengths is much larger than in adult populations based developmental stage. Thus, accurate prediction of the PTC may necessitate precise measures of leg length, as well as other anthropometric values such as height and weight. Therefore, the purpose of this study was to develop a model to predict age- and anthropometry-specific PTCs in individuals 6-20 years old (i.e., across the developmental lifespan). We hypothesized that cadence and anthropometric measures would classify gait behavior reasonably, i.e., with a prediction accuracy > 0.80.

# Methods

## Study design and regulatory information

This is a secondary analysis of data from the CADENCE-Kids study (Clinical Trials.gov - NCT01989104). A full description of the study design and participant characteristics are reported in Tudor-Locke et al. (2018).

## Participants

One hundred twenty-three children, adolescents, and young adults aged 6 to 20 years volunteered to participate. For participants 6-17 years of age, informed parental/legal guardian permission and child assent was required. All participants aged 18-20 years provided informed consent. Because of the considerable changes to body anthropometrics across this developmental age-span, and to ensure a uniform age distribution, study enrollment was set up such that at least 4 boys and 4 girls from each age year would be included. All participants were able to ambulate without an external device, free from mental illness within the past 5 years, not pregnant, and not taking any medication that would affect heart rate response to exercise.

## Anthropometric Measures

Barefoot standing height was measured via a stadiometer (Harpenden; Holtain Ltd., Crosswell, Crymych, Pembrokeshire, UK). Seated height was also measured with the stadiometer, whereby participants sat on a table with legs freely hanging. Leg length was quantified as standing height minus seated height. Weight and body fat percentage were measured using a Tanita digital scale (Tanita SC-240; Tanita corporation, Tokyo, Japan). This scale uses bioelectrical impedance technology to predict body fat percentage. Waist circumference was measured using a standard tape measure and identified as the narrowest circumference between the hips and ribs. Height and waist circumference measurement precision was to the nearest 0.1 cm, while weight was to the nearest 0.1 kg. All measurements were performed twice. If the height or waist circumference measurements differed by > 0.5 cm, or weight by > 0.5 kg, a third measurement was taken, and the average of the two closest measurements were used. BMI was calculated as weight (kg) divided by height squared (m2), and reported in kg/m2. However, because BMI is difficult to standardize in children, adolescents, and young adults, a BMI z-score (BMIz) was calculated using reference data from the Centers for Disease Control and Prevention (Kuczmarski et al., 2000). BMIz provides age and sex adjusted measures of the height-weight relationship. Finally, BMI percentile (BMI%) was used to define each participant as underweight (BMI < 5th percentile), normal weight (5th ≤ BMI < 85th percentile), overweight (85th ≤ BMI < 95th percentile), or obese (BMI ≥ 95th percentile).

**Cadence Measures**

Steps taken during each 5-minute trial were directly observed and manually counted (hand tally). Cadence (steps/min) was then computed as the total number of steps divided by 5. A video recording (GoPro Hero 3, GoPro Inc., San Mateo, CA, USA) of each participant’s lower body provided a redundant record for step verification purposes in the event of miscounting or ambiguous data.

## Protocols

Participants performed sequentially faster treadmill walking trials, starting at 0.22 m/s (i.e., 0.5 mph) and increasing in 0.22 m/s increments until: 1) completion of the fastest speed (2.23 m/s; 5.0 mph) for the protocol 2) completion of the trial during which participants freely chose to run, or 3) researcher or participant volition.

**Model Development**

Prior to model development, we first determined if any of the potential variables (see Table 1 for full list of variables) were highly correlated (i.e., > 90%). In the event that there were two correlated variables, and in an effort to provide the most feasible model for clinicians to use, we selected the easiest variable to obtain. For example, if height and leg length are highly correlated, we would select height because it requires a single measurement, and it does not require the further steps of measuring seated height, and then subtracting stool height from seated height, and then from standing height.

Logistic regression models were developed using the final set of independent variables after linear dependencies were removed. The goal of logistic regression is to develop a model that accurately classifies an outcome into two groups. For this analysis, running and walking are the two outcomes to be classified and, as such, the dependent variable for these models. The model was built using the “purposeful selection” technique detailed in “Applied Logistics Regression (Hosmer, Lemeshow, & Sturdivant, 2013). This technique ensures minimal collinearity between potential covariates and removes individually insignificant variables prior to building the initial model. Variables are then removed sequentially based upon the highest p-value until all remaining variables are statistically significant (i.e. p-value < 0.05).

**Table 1:** List of independent variables used to develop models.

|  |  |
| --- | --- |
| **Independent Variable** | **Explanation** |
| Sex | Male or Female biological sex at birth |
| Age (years) | Age of participant |
| Height (cm) | Height of participant |
| Weight (kg) | Body mass |
| Waist (cm) | Waist circumference |
| BMI (kg/m2) | Body Mass Index |
| BMI percentile | Age- and sex-specific percentile of the participant’s BMI. 85th percentile BMI designated overweight, 95th percentile BMI designated as having obesity |
| BMI z-score | Age- and sex-specific standardized BMI score |
| % body fat | Measured using bioelectrical impedance |
| Cadence (steps per minute) | Accumulated step count in five minutes divided by 5 |

## Data and Statistical Analysis

To assess the prediction accuracy of the final model, we performed a k-fold cross-validation, with k=10. The purpose of cross-validation is to determine how well the model will perform on out-of-sample data. For this validation method, the data is partitioned into 10 “folds”. A model is built with nine of the folds and tested on the unused “holdout set”, saving the resulting accuracy. After replicating with each fold as the holdout set, the resultant accuracies are compiled. The resulting cross-validation accuracy for this model is 0.974, much better than our hypothesized prediction accuracy of 0.80.

**R Shiny App Development**

An R Shiny app was created in R (R Core Team, 2018) to provide users with the probability that an individual is walking or running across a range of cadences given their individual-specific parameters.

# Results

**Participant characteristics**

Participant demographic data are reported in Table 2. Of the 123 potential participants, 69 individuals transitioned to running for the duration of the trial. Thus, only these data were used for model development. The total sample of 69 individuals consisted of 37 male and 32 female participants.

**Table 2:** Participant characteristics and preferred transition cadences across age groups.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age (years)** | **6-8** | | **9-11** | | **12-14** | | **15-17** | | **18-20** | |
| **Sex** | *F* | *M* | *F* | *M* | *F* | *M* | *F* | *M* | *F* | *M* |
| **n** | 2 | 4 | 4 | 3 | 7 | 8 | 9 | 10 | 10 | 12 |
| **Height (cm)** | 137 ± 1.9 | 130 ± 7.1 | 145 ± 6.1 | 143 ± 1.5 | 156 ± 7.7 | 162 ± 7.8 | 161 ± 7.0 | 173 ± 6.3 | 164 ± 4.9 | 177 ± 6.8 |
| **Weight (kg)** | 34.2 ± 8.0 | 26.6 ± 2.0 | 46.0 ± 4.9 | 38 ± 6.1 | 47.8 ± 11.9 | 64.5 ± 17.2 | 64.1 ± 15.6 | 72.1 ± 25.6 | 61.7 ± 12.5 | 73.2 ± 12.0 |
| **Waist Circumference (cm)** | 61.6 ± 9.9 | 54.6 ± 5.9 | 69.7 ± 8.6 | 65.8 ± 12.4 | 64.7 ± 9.8 | 81.1 ± 16.1 | 72.5 ± 13.6 | 79.7 ± 17.8 | 73.5 ± 9.9 | 80.7 ± 10.1 |
| **BMI (kg/m2)** | 18.4 ± 4.8 | 15.8 ± 2.0 | 21.8 ± 2.6 | 18.5 ± 2.8 | 19.3 ± 3.4 | 24.5 ± 6.4 | 23.5 ± 5.9 | 24 ± 8.1 | 22.9 ± 4.5 | 23.4 ± 3.1 |
| **BMI %** | 62.2 ± 48.3 | 45.8 ± 38.9 | 68.7 ± 11.9 | 64.0 ± 36.8 | 47.5 ± 32.3 | 75.0 ± 27.0 | 60.9 ± 24.2 | 65.1 ± 38.3 | 53.9 ± 36.4 | 52.7 ± 26.5 |
| **BMIz-Score** | 0.602 ± 1.7 | -0.205 ± 1.3 | 1.27 ± 0.6 | 0.456 ± 1.1 | -0.0639 ± 1.0 | 1.07 ± 1.1 | .468 ± 1.0 | 0.282 ± 1.5 | 0.00871 ± 1.3 | 0.107 ± 0.8 |
| **PTC (steps/min)** | 151 ± 6.1 | 156 ± 5.5 | 146 ± 1.0 | 147 ± 4.0 | 146 ± 4.0 | 144 ± 2.7 | 143 ± 2.8 | 144 ± 5.5 | 140 ± 3.3 | 140 ± 1.3 |

Note: Data presented as mean ± SD. F = female; M = male. PTC = preferred transition cadence.

The list of independent variables considered for model development are presented in Table 2. Three sets of highly correlated (greater than 90%) independent variables were identified: waist circumference strongly correlated with weight; BMI percentile correlated with BMI z-score; and height correlated with leg length. Waist circumference, BMI percentile, and leg length were removed from consideration because weight and height are easier and more practical measures for an individual to attain and because BMI z-score is a more rigorous representation of a person’s BMI by accounting for age and sex (Must & Anderson, 2006).

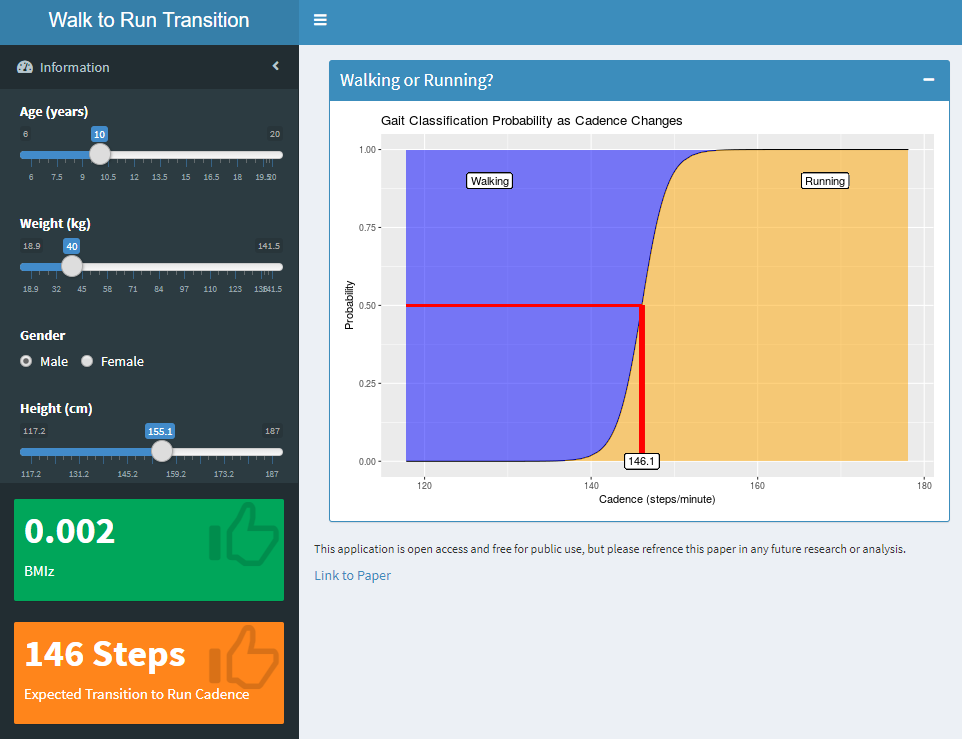
**Mathematical Models**

Each study participant represents two observations in the training data, one with their last walking cadence and another with their first running cadence, resulting in 138 total observations. Using the 138 training observations and the “purposeful selection” model-building technique, the final logistic regression model is the following:

These coefficients represent a change in the log odds of being in a running state. As age, height, BMIz, and cadence increase, the probability of that person being in a running state increases. The opposite is true for weight (note the sign of the coefficient). Cross validation was performed with a prediction accuracy of 0.974, indicating that this model accurately predicts gait classification as either walking or running. This is not a surprising result because the average gap from last walk cadence to first run cadence in the participants was 24 steps per minute, making classification much easier. To determine PTC, we determined the cadence at which the model is most uncertain as to the gait classification; i.e., where the probability of running or walking was 0.5.

**App Development**

Figure 1 depicts a screen shot of the user interface for the developed R Shiny app, which is available at <https://westpointmath.shinyapps.io/KidsStep/>. After the user inputs age, sex, height, and weight, the application returns the expected PTC with a graph visualizing the probability of being in either gait behavior.



**Figure 1:** Screen shot of the R Shiny App. After user input (age, weight, height, sex), BMIz (green) is displayed with expected preferred transition cadence (orange). The graph on the right displays the probabilities associated with being in either gait classification.

# Discussion

The purpose of this study was to develop a model to predict age- and anthropometry-specific PTCs in individuals 6-20 years old. We hypothesized that the logistic regression modeling approach using cadence and anthropometric parameters would classify gait behavior with reasonable accuracy. The results support this hypothesis, as the model we report herein displayed a prediction accuracy value of 0.974. Moreover, the Shiny App herein provides researchers and clinicians with an easy tool to estimate the PTC for physical activity behavior assessment.

The independent variables selected for the final model were fortuitous with regards to application of this model to the general population. Three of the four variables used in the model (age, weight, and height) are easily attained by any individual. The fourth variable, BMI z-score, can be calculated from the other three variables combined with the individual’s sex, and the AGD package (Van Buuren, 2018) in R. This potential for application to the general population was the impetus behind creation of the Shiny app. In the development of this model, sex was removed as not a significant factor in the presence of other variables such as height and BMI z-score which account for much of the sex-related differences in cadence in children. The Shiny App does require the user to input sex to determine their BMI z-score.

Previous efforts to establish the walk-to-run transition based on walking speed i.e., PTS, have provided ample evidence that this threshold is between 2.0-2.2 m/s (e.g., 2.03 m/s (Shih, Chen, Lee, Chan, & Shiang, 2016), 2.06 m/s (Hreljac, 1995), 2.08 m/s (Ganley et al., 2011), 2.09 m/s (Diedrich & Warren, 1995), 2.10 m/s (Prilutsky & Gregor, 2001), or 2.21 m/s (Ranisavljev, Ilic, Soldatovic, & Stefanovic, 2014)). Moreover, PTS’s have been reported in children (Tseh, Bennett, Caputo, & Morgan, 2002). However, with respect to cadence, while previous research has indicated that the transition from walking to running occurs at a PTC ~ 140 steps/minute in adults, the study by Tseh and colleagues (2002) did not report on the PTCs for walking to running. To our knowledge, the study presented herein is the first to use cadence to predict PTC in children, adolescents, and young adults. This information has potential to enhance the measurement of physical activity behavior. For example, the ActiLife software that processes laboratory or free-living ActiGraph data reports various outcome measures, including minute-by-minute step data (i.e., cadence), but not gait speed. Using the PTC values reported herein, a researcher or clinician could easily estimate minutes per day that a participant performed running behavior.

**Limitations**

One limitation for this study was that the observed PTC was based on trials that always began with walking and progressed to running. Considering there may be a hysteresis effect (i.e., running to walking may yield a different PTC), future research should incorporate protocols that include running to walking transitions as well. Also, the logistic regression model reports the value at which there is a 50% chance that the individual is walking, and 50% chance the individual is running. For researchers wishing to perform analyses of running behavior from a more conservative approach (i.e., maximize true positives and risk increased false negatives), they may choose to shift the cadence value towards higher probability of running. This can be easily assessed qualitatively using the R Shiny app.

**Conclusion**

Using standard anthropometric information (i.e., age, height, sex, and weight mass), the cadence (~144 ± 5.2 steps/minute) corresponding with the transition from walking to running can be accurately predicted in children, adolescents, and young adults. This information may be beneficial for researchers, practitioners, wearable device manufacturers, and the general public who are attempting to characterize locomotor behavior in the free-living setting. Moreover, herein we provide a free, user-friendly app that can be used by researchers or clinicians to determine an individual’s threshold without the need to program the equation.

# Acknowledgements

The authors would like to thank Dr. William Johnson, X, Y and Z for the original conceptualization and development of this project, data collection, and analyses at Pennington Biomedical Center. CADENCE-Kids was prospectively registered at ClinicalTrials.gov (NCT01989104). This work was supported by an award NIH NICHD 1R21HD073807 and in part by 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. These funding bodies had no role in design, in the collection, analysis, or interpretation of data, or in the writing or decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

# Conflicts of Interest

The authors declare no conflicts of interest

# References

Alexander, R. M. (1989). Optimization and gaits in the locomotion of vertebrates. *Physiological Reviews, 69*(4), 1199-1227. doi:10.1152/physrev.1989.69.4.1199

Alexander, R. M. (2002). Energetics and optimization of human walking and running: the 2000 Raymond Pearl memorial lecture. *American Journal of Human Biology, 14*(5), 641-648. doi:10.1002/ajhb.10067

Diedrich, F. J., & Warren, W. H. (1995). Why change gaits? Dynamics of the walk-run transition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(1), 183-202.

Ganley, K. J., Stock, A., Herman, R. M., Santello, M., & Willis, W. T. (2011). Fuel oxidation at the walk-to-run-transition in humans. *Metabolism, 60*(5), 609-616. doi:10.1016/j.metabol.2010.06.007

Hansen, E. A., Kristensen, L. A. R., Nielsen, A. M., Voigt, M., & Madeleine, P. (2017). The role of stride frequency for walk-to-run transition in humans. *Sci Rep, 7*(1), 2010. doi:10.1038/s41598-017-01972-1

Hansen, E. A., Nielsen, A. M., Kristensen, L. A. R., Madeleine, P., & Voigt, M. (2017). Prediction of walk-to-run transition using stride frequency: A test-retest reliability study. *Gait & Posture, 60*, 71-75. doi:10.1016/j.gaitpost.2017.11.008

Hosmer, J., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Hreljac, A. (1993). Preferred and energetically optimal gait transition speeds in human locomotion. *Med Sci Sports Exerc, 25*(10), 1158-1162.

Hreljac, A. (1995). Determinants of the gait transition speed during human locomotion: kinematic factors. *Journal of Biomechanics, 28*(6), 669-677.

Kuczmarski, R. J., Ogden, C. L., Grummer-Strawn, L. M., Flegal, K. M., Guo, S. S., Wei, R., . . . Johnson, C. L. (2000). CDC growth charts: United States. *Adv Data, 314*, 1-27.

Li, L. (2000). Stability landscapes of walking and running near gait transition speed. *Journal of Applied Biomechanics, 16*(4), 428-435. doi:10.1123/jab.16.4.428

Minetti, A. E., Ardigo, L. P., & Saibene, F. (1994). The transition between walking and running in humans: metabolic and mechanical aspects at different gradients. *Acta Physiologica Scandinavica, 150*, 315-323.

Must, A., & Anderson, S. E. (2006). Body mass index in children and adolescents: considerations for population-based applications. *International Journal of Obesity, 30*(4), 590-594. doi:10.1038/sj.ijo.0803300

Noble, B. J., Metz, K. F., Pandolf, K. B., Bell, C. W., Cafarelli, E., & Sime, W. E. (1973). Perceived exertion during walking and running. II. *Med Sci Sports, 5*(2), 116-120.

Prilutsky, B. I., & Gregor, R. J. (2001). Swing- and support-related muscle actions differentially trigger human walk-run and run-walk transitions. *Journal of Experimental Biology, 204*(Pt 13), 2277-2287.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>.

Ranisavljev, I., Ilic, V., Soldatovic, I., & Stefanovic, D. (2014). The relationship between allometry and preferred transition speed in human locomotion. *Human Movement Science, 34*, 196-204. doi:10.1016/j.humov.2014.03.002

Shih, Y., Chen, Y. C., Lee, Y. S., Chan, M. S., & Shiang, T. Y. (2016). Walking beyond preferred transition speed increases muscle activations with a shift from inverted pendulum to spring mass model in lower extremity. *Gait & Posture, 46*, 5-10. doi:10.1016/j.gaitpost.2016.01.003

Tseh, W., Bennett, J., Caputo, J. L., & Morgan, D. W. (2002). Comparison between preferred and energetically optimal transition speeds in adolescents. *European Journal of Applied Physiology, 88*(1-2), 117-121. doi:10.1007/s00421-002-0698-x

Tudor-Locke, C., Schuna, J. M., Jr., Han, H., Aguiar, E. J., Larrivee, S., Hsia, D. S., . . . Johnson, W. D. (2018). Cadence (steps/min) and intensity during ambulation in 6-20 year olds: The CADENCE-Kids study. *International Journal of Behavioural Nutrition and Physical Activity, 15*(20).

Usherwood, J. R. (2005). Why not walk faster? *Biology Letters, 1*(3), 338-341. doi:10.1098/rsbl.2005.0312

Van Buuren, S. (2018). \_AGD: Analysis of growth data\_ (Version 0.38). CRAN: The Comprehensive R Archive Network. Retrieved from <https://CRAN.R-project.org/package=AGD>