**METHODS**

**Study Design**

This study was designed to determine what variables predict the cadence at which a walk transitions to a run in children. To address this question we used data collected at the University of Massachusetts, Amherst’s Physical Activity and Health Laboratory to develop a model that predict the cadence at the walk to run transition from an easily measurable set of demographic and metabolic variables.

**Participants**

Some questions

1. When was the data collected and what was the primary outcome?
2. Has the data already been described elsewhere?
3. How were the participant’s recruited?
4. How many minutes were the children asked to walk at a constant pace?
5. How was “run” versus “walk” defined?
6. What is run cadence?

The dataset contained 122 participants who met transitioned from walk to run and maintained their running pace through the duration of their final stage. (50% female) between the ages of 7 and 20. Participants were guided to walk at a constant pace for \_\_\_ minutes. At the end of this interval, the speed was increased by 0.5 mph. The increase in speed was continued until the individual transitioned from walk to run. The experiment concluded at the walk to run transition interval. Of the 122 participants, 69 were able to attain the walk to run interval and only these data were used for model development. A number of the originally available variables on these 69 participants had linear dependencies. The list of independent variables considered for model development after accounting for these relationships are in Table 1 below. The dependent variable for all models was the step cadence of the participant in their final stage of the study (the stage in which they first began running); this is called the walk-to-run transition cadence.

Table 1 - List of considered independent variables and their explanations.

|  |  |
| --- | --- |
| Independent Variable | Explanation |
| Sex | Male or Female |
| Age | Age of participant in years |
| Height | Height of participant in cm |
| Weight | Weight of participant in kg |
| Waist | Waist circumference of participant in cm |
| BMI | Body Mass Index of participant |
| BMI percentile | The percentile of the participant’s BMI |
| BMI z-score | The standardized BMI score for the participant based on age and gender[[1]](#footnote-1) |
| Obesity Classification | 85th percentile BMI classified overweight, 95th percentile BMI classified obese |
| Tanita | Tanita Body Impedence Analysis Measure |
| Walk VO2 |  |
| Run VO2 |  |
| Run METS Youth1 |  |
| Run METS Youth 3 |  |
| Walk METS Youth 3 |  |

The 69 participants are a mix of males and females which are treated identically in model development. Physical attributes differ minimally in youth vice adulthood, and the mean critical characteristics to this study are within a standard deviation across genders (see Table 2).

Table 2 - Subject characteristic table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | N | Age  (mean ± sd) | BMI  (mean ± sd) | Walk-to-Run Transition Cadence (mean ± sd) |
| Males | 37 | 15.0 ± 3.88 | 22.6 ± 6.01 | 156 ± 10.2 |
| Females | 32 | 14.9 ± 3.70 | 21.9 ± 4.66 | 158 ± 8.92 |

**Mathematical Models**

Before constructing any regression models, two sets of highly correlated (greater than 90%) independent variables were discovered; waist circumference strongly correlates with weight while BMI percentile correlates with BMI z-score. Waist circumference and BMI percentile were removed from consideration because weight is an easier and more practical measure 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 gender.

We then used the leaps[[2]](#footnote-2) package in R[[3]](#footnote-3) to compare regression models for every possible subset of factors, selecting the model with the lowest BIC. All factors in the selected model were significant (have a p-value less than .01) while meeting the necessary linear regression assumptions. The final regression model was:

**Other Considered Models**

Linear regression was not the only type of model considered for this research. A k-means clustering approach feeding a Gaussian mixture model was also considered but did not provide the predictive power of the regression model based on BIC. Regularization methods were also considered but did not outperform multiple linear regression.

**App Development**

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 (age, weight, and height) are easily attained by any individual. The fourth variable, BMI z-score, can be calculated from the other three variables and the AGD[[4]](#footnote-4) package in R. With this possibility for application to the general population, a shiny application was created in R to provide users with the expected pace at which the minor will transition from walking to running. The user inputs the age, gender, height, and weight of the minor with the application returning the expected walk-to-run transition cadence with a 95% confidence interval. Based off the user’s inputs, the application calculates an individual’s BMI and subsequent BMI z-score; these data are fit to the final model to achieve the necessary outputs. The application is accessible at the following webpage: <https://dustyturner.shinyapps.io/KidsStep/> and a screen shot of the application is below (Figure 1).



Figure 1 - This screen shot of the Shiny App shows user inputs at the upper left. The bottom left shows the BMI (green) and BMI z-score (blue) of the user with the applicable results of the final model in orange. The graphs on the right depict how the walk-to-run transition cadence confidence interval changes with alterations to the various user inputs.

1. The AGD package in R was used for determining the BMI Z-score (cite). [↑](#footnote-ref-1)
2. Thomas Lumley based on Fortran code by Alan Miller (2017). leaps: Regression Subset Selection. R package version 3.0. https://CRAN.R-project.org/package=leaps [↑](#footnote-ref-2)
3. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. [↑](#footnote-ref-3)
4. Van Buuren S (2018). AGD: Analysis of Growth Data. R package version 0.39, <URL: https://CRAN.R-project.org/package=AGD>

   https://CRAN.R-project.org/package=AGD>. [↑](#footnote-ref-4)