8-2024

Predicting Individual Hiking Trail Intensity Using Statistical Learning

Kelci Hannan

Utah State University

PREDICTING INDIVIDUAL HIKING TRAIL INTENSITY USING STATISTICAL LEARNING

by

Kelci Hannan

A dissertation submitted in partial fulfillment
of the requirement for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

Christopher J. Dakin, Ph.D. Major Professor	David Bolton, Ph.D. Committee Member
Eadric Bressel, Ph.D. Committee Member	Breanna Studenka, Ph.D. Committee Member
Tyson Barrett, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

ABSTRACT

Predicting Individual Hiking Trail Intensity Using Statistical Learning

by

Kelci Hannan

Utah State University, 2024

Major Professor: Dr. Christopher J. Dakin
Program: Disability Disciplines

Hiking was the most popular outdoor recreation activity in 2023 and has been reported as the most common activity requiring Search and Rescue assistance. The low barriers to entry, paired with individuals' tendency to overestimate their abilities, may pose problems when individuals need to subjectively assess a trail's difficulty relative to their physical fitness capacity. Compounding this problem, each trail is assigned a difficulty from a difficulty rating system, to which all users must match the potentially inaccurate subjective assessment of their hiking abilities. With advancements in technology, other options exist to personalize trail difficulty ratings and reduce injury or overexertion risk while hiking. In this project, we evaluated the efficacy of generating a personalized prediction of trail difficulty. We first assessed the feasibility of using statistical learning models to estimate individual ratings of perceived exertion during a hike from biometric data, individual characteristics, and environmental characteristics. We then identified the variables with the greatest capacity to predict trail difficulty, compared accuracy of models using trail segment versus trail summary data, and assessed how changing the outcome variable impacted model accuracy.

Participants hiked the Wind Caves Trail in Logan, Utah while wearing an activity tracker and documenting their ratings of perceived exertion. Individual characteristics, including fitness level, hiking experience, and pain catastrophizing, were gathered using questionnaires before the hike. Based on the results, a large-scale study of this nature appears feasible, and a subset of predictor variables may be adequate to predict individual hiking trail intensity with sufficient sample size. Ultimately, using elastic net regression to predict ratings of perceived exertion resulted in the highest accuracy when pre-hike and intra-hike predictor variables were included. Classifying individual hiking intensity using a traditional difficulty scale resulted in every hiker being assigned the same difficulty and suggests that 46.67% of individuals might be misinformed of the trail difficulty and find it to be either harder or easier than anticipated. These projects lay a foundation for future research, within a relatively under-developed subject area, by illustrating study feasibility, exploring the data analytics of hiking difficulty prediction, and identifying variables important to these predictions.

(140 pages)

PUBLIC ABSTRACT

Predicting Individual Hiking Trail Intensity Using Statistical Learning

Kelci Hannan

Hiking is an increasingly popular choice for people wanting to engage in recreation and physical activity. It not only offers the many benefits of exercise but also the opportunity to explore nature and to socialize. To ensure that new hikers are prepared to hike and remain safe while on the trail, knowledge of a hiking trail's difficulty is important. Trail difficulty information is sometimes available at the trailhead if the trail is well-trafficked and maintained but is nearly always available online or through a hiking-specific mobile application, such as AllTrails. Both trailhead and online sources provide only one difficulty rating for all users regardless of personal characteristics, such as age and fitness level, which can lead to a misalignment of expectations and actual hiking experience. This project's overall goal was to attempt to generate personalized predictions of hiking trail difficulty and assess the effectiveness and utility of this approach. During a series of investigations, individuals hiked the Wind Caves trail in Logan, Utah while wearing a fitness watch that recorded heart rate, hiking speed, and location. Participants recorded how hard they felt they were exercising while hiking using a metric called Rating of Perceived Exertion. Before hiking, each participant answered questions about their fitness level, response to painful situations, and hiking experience. This information combined with statistical models was used to predict each individual's hiking intensity on the Wind Caves trail. These models were then evaluated for accuracy to assess how they might be used to improve estimates of hiking trail intensity prior to beginning a hike. We also explored which variables most improved hiking intensity prediction and how different perceived exertion scales impacted model performance. The results of this project suggest that personalized predictions of hiking difficulty may be feasible with refinement of the approaches used here, such as imposing one hiking speed or focusing on a sub-group of hikers, and lay a foundation for future research into applying statistical models to this question. Finally, we offer areas for improvement in future studies examining personalized trail intensity predictions.

DEDICATION

To Oleg and Edna, my constant companions during this project.

ACKNOWLEDGMENTS

To my advisor, Dr. Chris Dakin, thank you for your unwavering support, patience, and the endless lessons on how to provide high quality mentorship. I will be forever grateful for your encouragement to pursue questions that interested me and your reminder that anything can be learned with enough interest and time.

To my entire committee, I am exceedingly grateful for your flexibility with my change in dissertation topic due to the COVID-19 pandemic. Thank you for your excitement for and willingness to support this new direction.

To my mom and dad, Karla and James Besand, thank you for encouraging me to pursue as much education as I desired, even if that meant me moving further away from home with each degree.

To my partner, Joe Hannan, thank you for being in the messiness of every day with me, believing in me from the beginning, and gently reminding me that maybe I should write today.

CONTENTS

Page

Abstract iii

Public Abstract v

Dedication vii

Acknowledgments viii

Table of Contents ix

List of Tables xiii

List of Figures xiv

Chapter 1 1

Existing Hiking Trail Grading Systems 2

Performance and Intensity Prediction in Physical Activity 6

Prediction of Perceived Exertion as a Measure of Trail Difficulty 9

Specific Aims 12

Chapter 2 13

Introduction 13

Research Questions 16

Methods 16

Trail 16

Participants 17

Pre-Testing Procedures 17

Testing Procedures 18

Ratings of Perceived Exertion. 19

Data Processing 19

Results 22

Discussion 32

Efficacy of a larger sample size 32

VRPE versus Borg 6-20 33

Predictive independent variables 34

Limitations ... 35
Conclusion ... 35
Chapter 3 ... 36
Introduction ... 36
Research Questions ... 39
Methods ... 40
Trail ... 40
Participants ... 40
Pre-Testing Procedures ... 41
Testing Procedures ... 41
Ratings of Perceived Exertion. ... 42
Data Processing ... 42
Weather Data. ... 43
Trail Segment Data ... 43
Statistical Analysis ... 44
Results ... 47
Discussion ... 54
What is an accurate prediction of trail difficulty? ... 55
What is a meaningful change in VRPE? ... 57
Which Variables Improved Model Predictive Capacity? ... 58
Limitations ... 59
Conclusion ... 60
Chapter 4 ... 61
Introduction ... 61
Research Questions ... 63
Methods ... 63
Trail ... 63
Participants ... 64
Pre-Testing Procedures ... 64
Testing Procedures ... 65
Ratings of Perceived Exertion. ... 66
Data Processing ... 66

Statistical Analysis ... 66
Results ... 69
Discussion ... 73
Non-linear mapping of difficulty levels ... 73
Model prediction accuracy ... 74
Limitations ... 76
Conclusion ... 77
Chapter 5 ... 78
Why didn't predictive models perform better? Closer examination of trends in the data .. 78
Previous Trail Experience ... 79
Group Hiking ... 81
Chapter 6 ... 84
General Discussion and Conclusions ... 84
Methodology Considerations ... 84
Social Impact on Perceived Hiking Difficulty ... 85
Conclusions ... 87
References ... 89
APPENDICES ... 105
Appendix A. Description of Statistical Models ... 106
Appendix B. Borg RPE 6-20 Scale ... 107
Appendix C. Rating of Perceived Exertion Visual Analog Scale ... 108
Appendix D. Wind Caves Hiking Trail Information ... 109
Appendix E. Physical Activity Readiness Questionnaire (PAR-Q) ... 110
Appendix F. International Physical Activity Questionnaire (15-69 years) ... 111
Appendix G. Baecke Fitness Inventory ... 114
Appendix H. Pain Catastrophizing Scale ... 117
Appendix I. Demographic Questionnaire ... 119

Appendix J. Pre-Hike Survey 120

Appendix K. Traditional Hiking Difficulty Rating Scale 121

CURRICULUM VITAE 122

List of Tables

Table 2-1 *Results from Models Predicting VRPE* 24
Table 2-2 *Results from Models Predicting Borg 6-20* 24
Table 2-3 *Selected Questionnaire Items Used for Analysis* 31
Table 3-1 *Included Weather Variables* 43
Table 3-2 *List of Feature Sets.* 46
Table 3-3 *Results of Elastic Net Regression* 48
Table 3-4 *Features Identified with Variable Importance Analysis* 51

List of Figures

Figure 2-1 *Comparison of Borg 6-20 (left column) and VRPE (right column) Distributions* 23
Figure 2-2 *Correlations Between RPE of the Hike and PCS Items*. 26
Figure 2-3 *Pain Catastrophizing Scale (PCS) Variable Importance* 27
Figure 2-4 *Variable Importance of IPAQ and BFI Items* 28
Figure 2-5 *Correlations Between RPE of the Hike and IPAQ Items* 29
Figure 2-6 *Correlations Between RPE of the Hike and BFI Items* 30
Figure 3-1 *Hiking Project Difficulty Rating Scale* 38
Figure 3-2 *RMSE Values Resulting from Varying Regularization Parameter (lambda) and Mixing Percentage (alpha) Parameters in Elastic Net Regression* 49
Figure 3-3 *Top 10 Important Variables from Established Using Elastic Net Regression* 53
Figure 4-1 *Distribution of Categorical Response Variables* 67
Figure 4-2 *Classification cross-validation for the training set* 70
Figure 4-3 *RPE and Difficulty Confusion Matrices for the Test Data* 72
Figure 5-1 *Relationship Between Predictor Variables and VRPE Grouped by Previous Experience on the Wind Caves Trail* 80
Figure 5-2 *Relationships Between Predictor Variables and VRPE Grouped by Solo and Group Hikers* 82

Chapter 1

In 2023, participation in outdoor recreation reached a record high of 175.8 million people, with a little over half of Americans participating (Outdoor Industry Association & The Outdoor Foundation, 2024). Among outdoor recreation activities, hiking was the most popular (Outdoor Industry Association & The Outdoor Foundation, 2024). Hiking offers numerous physical health benefits, including lowering body weight, total cholesterol, and resting blood pressure (Greie et al., 2006) and improving insulin resistance (Schobersberger et al., 2010). Additionally, exposure to natural environments while hiking benefits overall well-being (Kaplan, 2001; R. Ulrich, 1984; R. S. Ulrich, 1979). However, hiking is not without risk. Among National Park Service units in 2005, hiking was the most common activity requiring Search and Rescue assistance (47% of incidents), and it was associated with 22.8% of all outdoor recreation-related fatalities (Heggie & Amundson, 2009). Under preparedness may be one source of this risk. For example, hiking contributed to 38% of all emergency medical services-involved injuries in Yellowstone National Park from 2003 to 2004 (Johnson et al., 2007) and was the leading contributor to injury (55%) at Mount Rainier and Olympic National Parks between 1997 and 2001, where 14.2% of injuries across all activities were due to overuse or exertion (Stephens et al., 2005). The majority of hiking-related injuries tend to be musculoskeletal in nature (Chrusch & Kavin, 2021) and mild or moderate in severity (Twombly & Schussman, 1995). Despite the well-documented injury risk associated with outdoor activity, hikers still underprepare, especially those that are younger, inexperienced, less fit, or plan shorter hikes and perceive this as less risky (Mason et al., 2013). Enhanced hiker preparation may alleviate some of these risks, and one means of doing this is to provide a better match between the trail's intensity rating and an individual's pre-hike expectations and fitness level.

Existing Hiking Trail Grading Systems

Generally, hiking trail grading systems offer five to seven difficulty categories, which are assigned based on the, more or less, subjective assessment of trail difficulty by an evaluator. An individual then must subjectively match themselves to the correct difficulty category during the process of selecting a trail to hike. These choices about trail difficulty are likely a contributing factor to the aforementioned risks associated with hiking, as they can be quite inaccurate (Heggie & Heggie, 2012). Outdoor recreation, in general, is plagued by overestimation of one's abilities, which manifests as an individual's inability to match their skill level to the difficulty of the task. This is a common enough occurrence in wilderness activities that one author coined the term "acute bad judgement syndrome" to refer to the poor decisions that result from a combination of overestimation of one's abilities and a lack of situational awareness (Trayers, 2004). Comparison of hiking to other outdoor activities, like mountaineering, reinforces the perception that hiking is easy, which may contribute to a general underestimation of the physical abilities required for safe participation (Heggie & Heggie, 2012). Therefore, expansion and subsequent standardization of trail rating systems could remove some of the subjectivity involved in hiking trail selection and prevent unnecessary injury due to the mismatching of individuals' fitness levels to trail difficulty.

Of the available hiking trail rating systems, one of the more objective systems uses average energy expenditure across treadmill gradients (Hugo et al., 1998) to categorize trail difficulty. This rating system has been applied to a variety of hiking trails using each trail's topographical map (Hugo, 1999b) and breaks trail difficulty into seven numerical levels with matching categorical designations that are 'Very Easy', 'Easy', 'Fair', 'Moderate', 'Difficult', 'Severe', and 'Extreme' (Hugo, 1999a, 1999b). An extension of this system is the Green Flags Trails Unique Trail Difficulty Rating System, which uses a 10-point effort scale divided into 4 difficulties (easy, moderate, difficult, extreme) that align with total energy expenditure for a given trail (McIntosh & Hugo, n.d.). While this system appears to be the most rigorously tested of those commonly used, it seems to be applied most frequently in South Africa and other select locations without wide acceptance in other countries (*International Hiking Trails*, n.d.). One potential reason for the limited use of this system is that individuals must know their energy expenditure (kJ) on a trail that has already been evaluated using the Green Flags system in order to 'calibrate' themselves for other trails (Hugo, 1999b). This presents a barrier to use because energy expenditure is neither available for all trails nor easily accessible at a trailhead. Even if energy expenditure information is available, most individuals do not have access to or know how to use this value without instruction. In reference to such energy expenditure based rating systems, Hugo (1999) stresses that they are independent of the individual hiker, individuals across fitness levels will find a difficult trail proportionally more tiring than an easy trail, and it is the trail, not the person, being evaluated (for its energy requirements). While important, this exact principle may be an inherent point of confusion and distress for a number of leisurely hikers. Although individuals familiar with hiking trail difficulty ratings may understand that the difficulty levels are meant to be 'easy' or 'difficult' relative to their fitness level, individuals with less knowledge about hiking may interpret a 'moderate' rating to mean it will be a leisurely experience, when, in fact, this trail may be difficult for a beginner. This, in turn, may lead to a negative experience and discourage an individual from pursuing hiking as a future activity.

Various other, less rigorous and thoroughly tested, hiking trail rating systems exist and are commonly developed and implemented by local parks services. Among these rating systems, the variables used to quantify a trail's difficulty are applied inconsistently, often differing between country, city, or park. Within the United States, for example, the National Park Service oversees 423 separate parks (National Park Service, 2021), of which some implement unique trail grading systems that differ from each other. Willamette National Forest grades trails as 'easy', 'moderate', or 'difficult' based on a trail's grade, width, and surface (Willamette National Forest, n.d.), whereas Shenandoah National Park in Virginia implements a formula that considers a trail's elevation gain and distance to then provide a numerical value that corresponds to five rating levels ('easiest', 'moderate', 'moderately strenuous', 'strenuous', 'very strenuous') (Shenandoah National Park Virginia, 2017). A mobile application, called Hiking Project, also has its own six-point rating system that provides trail ratings based on the average subjective rating of app users (Hiking Project, n.d.). These categories include 'easy', 'easy/intermediate', 'intermediate', 'intermediate/difficult', 'difficult', and 'very difficult' and are predominantly based on the steepness of the trail and the type of terrain (Hiking Project, n.d.).

A common theme across all trail rating systems is the requirement that individuals accurately estimate their fitness level to select an appropriate hike. However, evidence suggests that only slight to fair agreement exists between self-rated fitness and objectively measured fitness (Jensen et al., 2018; Obling et al., 2015). Further, an individual's estimate of their fitness level becomes less accurate with age (Germain & Hausenblas, 2006), and men, as well as less fit individuals, tend to overestimate fitness level (Jensen et al., 2018; Shook et al., 2016). The sometimes-poor accuracy of an individual's assessment of their fitness level may lead leisurely hikers to partake in hikes that are more difficult than they intend. An increase in inexperienced hikers seeking outdoor opportunities seems likely as outdoor recreation is encouraged by a growing number of initiatives, such as nature prescriptions (BC Parks Foundation, n.d.; Kondo et al., 2020; Mitten et al., 2018) and #OptOutside (REI, n.d.). This presents a unique challenge that calls for adaptation of classic hiking trail rating systems in order to accommodate an increasingly diverse group of individuals who have varying degrees of hiking experience.

With many hiking trail rating systems relying heavily on physical trail characteristics to determine the difficulty level, a new system that considers individual physical characteristics and fitness level could provide an alternative, perhaps more accurate, rating approach. Combined with the increasing role of technology in everyday life and the prevalence of cell phone use to collect movement data, a more advanced approach to the rating of hiking trails through statistical learning could allow the opportunity for increasingly accurate, personalized hiking trail difficulty classification. Providing guidance to hikers through prediction of individualized hiking trail difficulty, in turn, could remove the burden from medical providers to recommend specific trails as patients progress in physical fitness and help to increase the safety of outdoor activity as exercise prescription (Mitten et al., 2018). In fact, activity and intensity prediction algorithms used in tourism and sports domains could inspire such a rating system.

Performance and Intensity Prediction in Physical Activity

A personalized rating system using individual characteristics, biometric data, and trail data, that predicts the trail difficulty for a given individual, could improve the general population's experience of hiking and further encourage participation by sedentary individuals. With rapid improvements in technology, statistical approaches, and accessibility to technology, trail grading systems have the potential to be personalized to each hiker. Current use of such predictive statistical models is prevalent in tourist and leisure activity but often focuses on providing activity recommendations based on an individual's preferences, interests, and location. Currently, there are none that allow a user to select a specific hiking trail and then provide the user with an estimate of the difficulty level based on the user's own capability.

The tourism industry provides suggested outdoor recreation activities through recommender systems. These systems identify products or services a specific user would be interested in (Schumacher & Rey, 2011) and can provide targeted activity suggestions based on location, personal interests, and other preferences (Borràs et al., 2014). With a focus on hiking, Calbimonte et al. (2018) prototyped SanTour, a recommender system, that would provide users with hiking trail options by combining individual health profiles and trail profiles. This approach tailors the basic recommender system to consider the physical condition of each person and individual factors that may be limiting, such as tolerance of heights, when suggesting hiking trails. More recently, Calbimonte et al. (2021) presented Syris, a platform developed to provide more precise information regarding the difficulty and risk of hiking trails while still acting as a recommender system for tourism trails. Syris is based in semantic data models and includes a difficulty assessment model that defines the difficulty of a trail with three criteria: Effort, technique, and risk (Calbimonte et al., 2020). While Syris focuses on providing trail recommendations suitable for a participant's questionnaire-measured fitness level, it does not provide individualized hiking trail difficulty ratings for a given trail. It, instead, provides a selection of trails that fit the user's profile and allows for filtering of trails based on pre-assigned trail difficulties. Even with the consideration of individual characteristics when suggesting hiking trails and tourist attractions, at their core these tools remain recommender systems and do not provide direct information about a specific hiking trail in relation to the individual.

Among the recommender system literature are a few instances where statistical modeling approaches have been used to predict event performance times (Fogliato et al., 2020; Pitman et al., 2012; Sándor, 2018) and to classify average mountain bike trail difficulty based on bikers' subjective ratings (Langer et al., 2020). Statistical learning is a broad term for a variety of methods used to understand and learn from data (Hastie et al., 2018; James et al., 2017). These methods can be used to (1) predict an outcome based on certain features (i.e. inputs, predictors, independent variables) and their known responses (i.e. outputs, dependent variables) (supervised learning), (2) determine the importance of certain variables to the predictive capacity of the model and, (3) infer the nature of the relationship between variables when no outcome variable is present (unsupervised learning) (Hastie et al., 2018). Investigations implementing statistical learning approaches tend to be divided into two types of problems: regression and classification (James et al., 2017). Generally, we can think of these problems in terms of the type of data the model is trying to predict. Regression problems deal with predicting quantitative data, and classification problems deal with predicting qualitative data, but, as with most things, this is not a hard and fast rule and some flexibility exists (James et al., 2017). In either of these problems, the goal is to approximate the mathematical function that represents the relationship between the features and responses (Hastie et al., 2018). One recent study used statistical modeling to predict mountain bike trail difficulty from users' subjective ratings of trail difficulties (Langer et al., 2020). While valuable for potential standardization of mountain bike trail ratings, their grading system was limited to only three categories, which simplifies ratings but also does not allow for differentiation of widely varying trail segments. In predicting performance times, statistical learning models tend to outperform rule-of-thumb based methods used to estimate hiking time (Pitman et al., 2012; Sándor, 2018) and intercept-only models predicting running time (Fogliato et al., 2020). A variety of explanatory variables were used in these models including: Hike length, trail gradient, current progress on the trail (Pitman et al., 2012), steepness-velocity relationship, and previously recorded velocity values (Sándor, 2018). While these investigations resulted in final models with high performance, these approaches do not address the problem of matching an individual's fitness level to trail difficulty.

Perhaps the body of literature most aligned with the current project is that concerning prediction of fatigue during sport using statistical learning approaches. The general goal of these studies was to identify the onset of fatigue in order to evaluate training loads and individualize training plans. The success of statistical models at predicting fatigue in Australian football (Bartlett et al., 2017; Carey et al., 2016), soccer (Geurkink et al., 2019; Vandewiele et al., 2017), and running (Davidson et al., 2020; Op De Beéck et al., 2018) suggests that statistical learning may serve as a useful starting point for predicting hiking difficulty. These studies employed a variety of statistical approaches to predict rating of perceived exertion (RPE), ranging from simple linear regression to support vector machines

(Appendix A), each with varying, but overall high, predictive accuracy. Specifically, Carey and colleagues (2016) provide a direct comparison of regression and classification approaches, which performed comparably in terms of predictive ability, although regression models resulted in slightly better predictions. Additionally, models that could account for non-linear relationships tended to perform better in both regression and classification across studies (Bartlett et al., 2017; Carey et al., 2016). The effectiveness of these studies' use of statistical learning to predict rating of perceived exertion across multiple sports domains with biometric information and/or training data suggests these predictor variables may also be effective for hiking trail difficulty prediction during hiking. In addition to these predictor variables, several other factors (e.g., weather) also likely affect trail intensity, and their importance can be assessed as part of the statistical learning process. Equally important is selecting the best outcome variable by taking into consideration the research question and the methods available to quantify the behavior being predicted.

Prediction of Perceived Exertion as a Measure of Trail Difficulty

Prior to the application of statistical learning models (further discussion in the methods), it is imperative to identify an appropriate outcome measure that provides an objective measure of trail difficulty and that does not prohibitively burden the hiker during its collection. Several options are available including a subjective assessment of trail difficulty, subjective ratings of perceived exertion, and heart rate.

A subjective assessment of overall trail difficulty can be obtained using a traditional rating scale of ‘easy’, ‘moderate’, and ‘difficult’. While this provides the simplest measure of trail difficulty, it does so with limited resolution of the individual’s experience during the hike, as discussed in the previous paragraphs. By forcing all trails into one of only three categories, we exclude much of a trail's inherent variation. A better measure may be a rating of perceived exertion (RPE), as used extensively in the exercise physiology literature and, more importantly to this study, in other activities for statistical learning approaches (Bartlett et al., 2017; Carey et al., 2016; Davidson et al., 2020; Geurkink et al., 2019; Op De Beéck et al., 2018; Vandewiele et al., 2017). RPE, recorded at several points during the hike, can provide finer detail about the user experience and will allow for better identification and separation of individual hiking levels during model application.

Perceived exertion is a subjective indicator of overall physical strain (Borg, 1982), which can be used to measure a hiker's subjective experience of exercise intensity while hiking, and perhaps a trail's difficulty. The Borg 6-20 scale (Appendix B) is a simple, effective tool for gauging an individual's level of perceived exertion quickly and accurately (Borg, 1970, 1998). Initially, the Borg 6-20 scale was validated against heart rate (Borg, 1962b, 1962a), and a strong correlation between these variables has been reported (Borg, 1982) but is not always observed (Chen et al., 2002). RPE, it seems, may be dependent on other variables (Robertson et al., 1998), which might affect the relationship between it and heart rate (Chen et al., 2002). Considering that measures of external training load have been useful for predicting RPE (Carey et al., 2016), obtaining accurate predictions of individualized trail difficulty using RPE likely requires some accommodation for variables known to influence measures of RPE and heart rate.

Ambient temperature, body weight, and use of hiking equipment can differentially affect individual RPE ratings and, therefore, they may be useful variables to include to improve the capacity of statistical models to predict hiking trail difficulty. Ambient temperature can influence RPE, which increases in hot environments compared to cool ones. This effect has been observed during cycling (Maw et al., 1993), treadmill walking (von Heimburg et al., 2019) and during thermo-neutral yoga (Boyd et al., 2018). There also appears to be an additive effect of heat and hypoxic environments (at higher altitudes) resulting in higher RPE ratings when combined, compared to hot or hypoxic environments alone (Levine & Buono, 2019). Similarly, individuals who are overweight consistently rate perceived exertion higher than individuals of an ideal weight during self-selected (Hulens et al., 2003) and imposed walking intensities (Ekkekakis & Lind, 2006; Marinov et al., 2002). The use of hiking equipment, including trekking poles or a pack, may also impact perceived exertion and/or heart rate during hiking. Treadmill walking at 0% grade (Rodgers et al., 1995) and hiking (Saunders et al., 2008) with trekking poles has been shown to result in significantly higher heart rate compared to not using poles, but RPE was not significantly altered by trekking poles regardless of trail grade (Saunders et al., 2008). Others report that treadmill walking at small grades with a backpack led to higher average heart rates and lower RPE when using poles compared to not using poles (Knight & Caldwell, 2000). Still others observed that trekking poles had no significant effect on heart rate or RPE (Perrey & Fabre, 2008). At steeper grades (25%) no change in physiological variables, including heart rate, was observed with and without poles, but RPE was lower when using trekking poles (Jacobson et al., 2000). The impact of trekking poles on physiological measures appears dependent on other variables, including steepness of the trail (Saunders et al., 2008). Given these previous findings, ambient temperature, body weight, and hiking equipment have the potential to improve the ability of statistical models to predict trail intensity as measured by RPE.

Ultimately, the ability to accurately estimate an individual's perceived difficulty on a given trail could enhance existing trail rating systems, serve to improve hiking safety and enjoyment (through appropriate trail selection), and improve the safety of medical recommendations for hiking as a prescription for outdoor activity. Therefore, the purpose of the current project was to develop and test a hierarchy of statistical models to determine the efficacy of developing a personalized prediction of hiking trail difficulty.

Specific Aims

Here we aimed to evaluate the efficacy of generating a personalized prediction of trail difficulty by addressing the following specific aims:

1. Assess the feasibility of using statistical learning models to estimate individual ratings of perceived exertion during a hike from biometric data measured by an activity tracker, individual characteristics, trail characteristics, and environmental characteristics (Chapter 2).
2. Identify the variables with the greatest capacity to predict trail difficulty (Chapters 2 and 3).
3. Compare the predictive accuracy of statistical learning models using trail segment data versus trail summary data (Chapter 3).
4. Contrast the accuracy of models predicting overall trail difficulty on a traditional categorical rating scale, to those predicting perceived exertion on a continuous scale (Chapter 4).

Chapter 2

Introduction

In light of the scant research related to assessment of hiking trail difficulty (Coetzee, 2018), the first step to making improvements in hiking trail rating systems is to identify variables that could be indicative of an individual's subjective trail difficulty and to explore options for the ideal study design. The overarching theme of this investigation is to address how we can obtain a full picture of a hiker's experience on a given trail without over-burdening them with questionnaires, equipment, physical assessments, and excessive documentation while hiking.

Before addressing this question, we need to identify and select a proxy measure for trail intensity with which to estimate individual hiking difficulty. As the goal is to use statistics to model individual hikers' perceived physical exertion on a trail, a subjective measure that mirrors this perception may be an ideal choice, as it will reflect each individual's perceived hiking experience. Hiking difficulty, as rated on a traditional difficulty scale, is a clear option as it measures an individual's perception of trail intensity using a familiar format. The primary problem with traditional difficulty scales is that they tend to have only a few rating levels, often five to seven difficulties from which to choose. Such few levels can reduce discriminability between hikers, potentially limiting an algorithm's ability to classify hikers into different tiers of physical fitness. Rating of perceived exertion (RPE) offers other benefits, such as a numeric scale, which offers rating flexibility and is important for some types of statistical models. Alternatively, common measures of RPE provide greater difficulty resolution (i.e., 15 levels), which could better separate hikers with differing fitness levels and provide more variation in individual trail intensity ratings (which is valuable when attempting to predict using regression-based models).

The Borg 6-20 scale, a commonly-used measure of RPE, has been well-established in a laboratory setting (Borg, 1962b, 1962a, 1970), but the categorical nature of this variable poses potential problems with regression-based modeling (Hastie et al., 2009). While it is possible to treat this variable as numeric, a direct measure of RPE as a continuous variable could be achieved by modifying the Borg 6-20 into a visual analog RPE scale (VRPE). The VRPE (Appendix C) has a moderate, positive relationship with the Borg scale ($r = 0.71$) and shows high reliability in test-retest analyses (ICC = 0.97) during a variety of standardized exercise tests (e.g., 1-mile run/walk, 1-minute push-up, 3-minute step test) (Casey et al., 2015) and submaximal exercise (Grant et al., 1999). While these provide general support for use of VRPE in place of the Borg 6-20, there appear to be no reports of VRPE and Borg 6-20 comparisons during hiking, which has the unique quality of being self-paced with no limit on number or length of breaks. Additionally, it is necessary to confirm that the RPE values measured using the VRPE provide as much (or more) information as the Borg 6-20.

While a host of variables from human physiology, biomechanics, and physical assessments could inform trail difficulty research, narrowing these down to the most pertinent, while considering their accessibility and collection time burden, will benefit the practicality of any applicable findings and is essential to motivate broad use from the general population. Some insight into useful variables may arise from investigation into the prediction of fatigue in Australian football (Bartlett et al., 2017; Carey et al., 2016), soccer (Geurkink et al., 2019; Vandewiele et al., 2017), and running (Davidson et al., 2020; Op De Beéck et al., 2018). In these studies, the authors used a variety of tests to provide an indication of fitness level (e.g., Step-Up Test, VO_2). These fitness tests yield reliable and valid information, and are likely predictive of individuals' trail difficulty ratings, but many may not be easily accessible to the public (in that they require effort to research and complete) and, therefore, will reduce the use of a trail rating system based upon them. For example, the Step-Up Test (De Villiers & Thiart, 1988) is a strong predictor of exertion on hiking trails (Coetzee, 2018) but not everyone is inclined to take a physical fitness assessment prior to undertaking leisure activity. Moreover, a primary barrier to the identification of predictive variables is that the majority of previous publications focus on elite athletes (Geurkink et al., 2019; Vandewiele et al., 2017) with access to more assessment tools (e.g., VO_2 max, muscle fiber type, anaerobic and aerobic thresholds) compared to leisure hikers. If a personalized hiking difficulty prediction system is to be broadly adopted by the public, barriers to its use, such as excessive data collection and pre-hike assessments, likely need to be minimized.

Therefore, the purpose of this study was not to predict hiking difficulty with any and all measures possible but to do so in a way that would make this approach broadly accessible to the general public. Using traditional laboratory measures could limit the accessibility of this approach by increasing the burden on participants and thereby creating a negative incentive to use predictive models dependent on these measures. Consequently, the focus of this feasibility study was to identify variables predictive of trail intensity whose collection present the least time burden to participants, in order to maximize the applicability and breadth of use of the final statistical model.

Research Questions

With this first study, we sought to determine the feasibility of collecting a range of variables potentially predictive of perceived exertion while hiking in order to identify and refine those variables most predictive of perceived exertion (Overarching Specific Aims 1 and 2). This study also served as a proof of concept through preliminary implementation of the data collection protocol in which we investigated the following research questions:

1) Is a larger study based on similar methods feasible, considering the recruitment of participants, hiker interest, and the broad range of measures potentially necessary to make accurate predictions?
2) Which version of the RPE scale (Borg 6-20 or VRPE) might best differentiate between participants of varying fitness levels?
3) Which independent variables are the strongest predictors of RPE during the entire hike?

Methods

Trail

The hiking trail used to address these questions was the Wind Caves Trail in the Uinta-Wasatch-Cache National Forest in Cache County, Utah (Appendix D). This 1.8 mile out-and-back trail contains a variety of grades and varying trail features which allow for discriminating between participants of differing fitness levels and demand a range of physical abilities across different trail sections. Additionally, this trail is heavily trafficked making it the best regional option for recruiting participants.

Participants

Individuals 18 years and older with an interest in hiking the Wind Caves trail and willingness to wear an activity tracker and chest strap heart rate monitor were recruited for this study through physical flyers placed around the university and surrounding community, in-class announcements, social media posts, and by word of mouth. All individuals completed the Physical Activity Readiness Questionnaire (PAR-Q) (Appendix E) prior to enrollment in the study to ensure there were no underlying conditions that contraindicated the level of physical activity required by this investigation. Individuals who met the pre-screening health criteria were enrolled in this study. Of the 112 individuals that completed the pre-screening survey, 51 (F: 33; M: 18; 30.8 ± 14.96 yrs; 71.11 ± 13.83 kg; 171.52 ± 9.86 cm) participated in the hike. All procedures were approved by Utah State University's Institutional Review Board (#12021).

Pre-Testing Procedures

Prior to arriving at the trailhead, interested individuals completed the pre-screening survey (Appendix E). Eligible individuals then completed the informed consent form, as well as a series of forms meant to determine their fitness history and level (International Physical Activity Questionnaire [Appendix F], Baecke Fitness Inventory [Appendix G]) and measure other factors that may influence their exertion and difficulty ratings (Pain Catastrophizing Scale [Appendix H], demographic questionnaire [Appendix I]), and finally scheduled a time to hike. All forms were hosted on REDCap (Research Electronic Data Capture) (Harris et al., 2009, 2019). If the participant had not hiked the Wind Caves Trail before, they were sent a fact sheet about the trail to inform them of the trail's characteristics, and they were encouraged to hike with another individual.

Testing Procedures

Upon arrival at the trailhead, participants completed a pre-hike survey (Appendix J), which included items referencing circumstances that could have changed before arrival at the trailhead (i.e., use of trekking poles, size of hiking group, etc.). Each participant's dominant hand (preferred writing hand) grip strength was measured using a hand grip dynamometer (T.K.K. 5001 GRIP-A Analog Grip Dynamometer, Takei Scientific Instruments Co., Ltd, Niigata City, Japan) as others have reported this measure to be a significant predictor of muscle endurance (Trosclair et al., 2011) and aerobic capacity (Dag et al., 2021). Three maximal efforts were recorded with the participant standing, elbow held at 90 degrees, and wrist in a neutral position. If the participant planned to carry a pack while hiking, the weight of the pack was measured using a handheld digital force gauge (FGE-HXY Digital Force Gauge, Nidec-Shimpo Corporation, Kyoto, Japan). Next, participants were instructed to fit the commercially available fitness tracker (Garmin Forerunner 235, Garmin Ltd., Olathe, Kansas, USA) on their non-dominant wrist. A researcher then checked it for proper fit, ensured data recording was enabled, and gave instructions for how to rate perceived exertion and document their exertion and difficulty ratings during the hike (see below). Participants hiked the Wind Caves Trail (without a researcher) at a self-selected pace and provided ratings of perceived exertion a predetermined number of times, as described below. Participants were free to take breaks when desired, and their time spent at the top of the trail was not restricted. The activity tracker worn on the wrist collected biometric data throughout the hike, including heart rate, distance, date, time, latitude, and longitude, at a sampling rate of 1 Hz. Only data between the GPS coordinates for the start and end of the hike were analyzed. When finished hiking, participants returned all equipment to the research station at the trailhead and rated the entire trail difficulty on a seven-point scale (Appendix K).

Ratings of Perceived Exertion.

Participants recorded their ratings of perceived exertion on paper using a clipboard and pencil. All participants completed a pre-hike, post-ascent, post-descent, and post-hike rating using a traditional seven-point difficulty scale and using their pre-assigned rating of perceived exertion scale. Participants completed 1 to 15 RPE ratings during both the ascent and descent of the hike using the assigned RPE scale, for a total of 2 to 30 ratings over the entire hike. A participant assigned 1 RPE rating completed the scale twice during their hike, once during the ascent and again during the descent, while a participant assigned 15 ratings completed the scale a total of 30 times, 15 during the ascent and 15 during the descent. The number of ratings and the RPE scale were quasi-randomly assigned. If participants hiked in pairs or a group, all hikers received the same RPE scale and number of ratings to complete during the ascent and descent. Participants were instructed to spread out their RPE ratings during the ascent and descent to be distributed over the entire hike as best they could. The general guideline used was to complete a rating every 2-5 minutes depending on the assigned number of ratings and the individual's expected hiking speed, which would allow for RPE observations over the entirety of the hike as opposed to all ratings being clustered over a certain trail section. Participants were also instructed to complete a rating prior to stopping if they wanted or needed to take a break during the hike.

Data Processing

RPE was obtained from visual analog RPE scales by measuring, with a ruler, in millimeters the distance from the zero anchor to the intersecting line drawn by the participant. All RPE and difficulty responses were manually input into REDCap, and fitness tracker data (i.e., .fit file) were transferred to a secure university computer for storage and data analysis. Using the Python programming language (Python Core Team, 2023), all .fit files from the fitness tracker (the GIS data file used by Garmin equipment and software) were downloaded and saved as a text file. All subsequent data wrangling and analysis of the fitness tracker data (e.g., altitude, heart rate, latitude, longitude, speed, date, and time) were performed in R (R Core Team, 2021). To accommodate differing data lengths across participants due to hiking speed, all measured and derived variables were summarized over the ascent, descent, and entire hike, resulting in metrics with a data length of one for all individuals (e.g., maximum heart rate during the ascent, average speed during the descent, etc.). All demographic and fitness questionnaire data collected in REDCap were exported and merged with fitness tracker data.

Histograms of individual RPE during the hike's ascent and descent were used to identify which RPE scale (Borg 6-20 vs. VRPE) to use in the next stages of this project. This was done by qualitatively assessing the uniformity of and detail provided by the data gathered with each scale over the course of the hike. A sub-purpose of generating these plots was to check that participants' responses were relatively evenly distributed over the entire hike. Two sets of elastic net regression models were built to explore the two outcome variables of interest (VRPE and Borg 6-20 over the entire hike). This aided in determining which outcome variable resulted in models with higher predictive accuracy. Four models were built for each outcome variable using feature sets based on the included questionnaires and wearable-derived variables. The four feature sets included (1) all pre-hike variables, (2) only Baecke Fitness Inventory (BFI) items, (3) only Pain Catastrophizing Scale (PCS) items, and (4) only International Physical Activity Questionnaire (IPAQ) items. Predictive accuracy of all models using different outcome variables was assessed using RMSE estimated from five-fold cross-validation. All models were built with the caret package (Kuhn, 2008; Kuhn et al., 2023) in R.

A multi-step approach was used to identify the strongest predictors of RPE from the measured independent variables. First, Pearson correlations between all continuous independent and dependent variables were used to identify the strongest relationships between variables. Those independent variables showing strong correlations (relative to other predictor variables) with RPE were marked as potentially important. Independent variables that were strongly correlated with other independent variables were noted as potentially less important or problematic because a high correlation between predictor variables can lead to multi-collinearity in regression-based statistical models and less reliable statistical inferences (James et al., 2017). Visual assessment of correlation matrices, and subsequent confirmation of the correlation coefficients, was used to select those independent variables exhibiting the strongest relationships (relative to the strength of other predictor variables) with RPE. During this process, we focused predominantly on questionnaire items, as these presented the largest pre-hiking time barrier for participants and the most potential for variables with overlapping information (i.e., IPAQ and BFI) that could be removed to reduce burden to participants.

Next, we examined variable importance within each feature set to identify the predictor variables that most strongly contribute to the models' predictive accuracy. The independent variables identified in the step above as having the strongest correlations to the outcome variable were used in combination with the highest-ranking features from the variable importance analysis to determine which variables to include in a larger study. The final determination on which variables were kept was made using both the statistical results and by considering the ease of collection and potential time burden to participants.

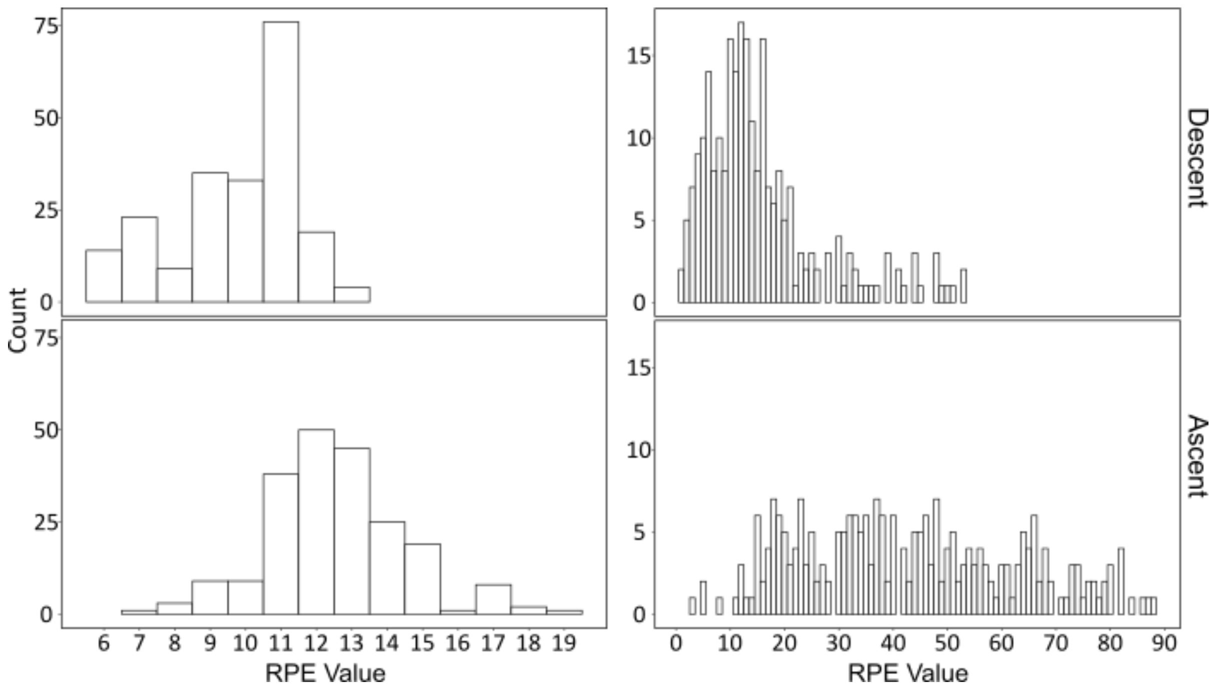
Results

Performance of the Borg 6-20 and the VRPE was evaluated by examining the distribution of RPE ratings during the hike (Figure 2-1). The Borg 6-20 shows a relatively normal distribution during the
ascent (Figure 2-1, bottom left) and a left-skewed distribution (lower ratings) during the descent (Figure 2-1, top left). The VRPE exhibits a more uniform distribution of ratings during the ascent (Figure 2-1, bottom right) and a right-skewed distribution during the descent of the hike (Figure 2-1, top right). Of the 211 Borg 6-20 ratings provided during the ascent, 45% of them were a 12 or 13. Of the 213 ratings during the descent, 35% were an 11 and no ratings above 13 were used. The total number of ratings during the ascent and descent differ due to participant error (participants missing assigned ratings). Models predicting VRPE consistently have lower RMSE than Borg 6-20 across feature sets (Table 2-1 & Table 2-2), but this is likely due to the discrete nature of the Borg 6-20 scale. As the VRPE has a

more uniform distribution of scores, and more flexibility to record participant's responses, from this point forward VRPE is used as the outcome measure.

Figure 2-1

Comparison of Borg 6-20 (left column) and VRPE (right column) Distributions



Note. The Borg 6-20 shows a normal distribution during the ascent (bottom left), while the VRPE exhibits a more uniform distribution during the ascent (bottom right). During the descent, the Borg 6-20 exhibits a left-skewed distribution (top left), while VRPE shows a right-skewed distribution (top right).

Table 2-1

Results from Models Predicting VRPE

Feature Set	RMSE
Pre-Hike Variables	5.49 e-15
BFI Items	9.69 e-16
IPAQ Items	1.29 e-15
PCS Items	6.14 e-15

Note. All RMSE values here result from the elastic net models. Pre-Hike Variables include those measured prior to the hike. BFI: Baecke Fitness Inventory; IPAQ: International Physical Activity Questionnaire; PCS: Pain Catastrophizing Scale; VRPE: visual analog rating of perceived exertion scale; RMSE: root-mean squared error

Table 2-2

Results from Models Predicting Borg 6-20

Feature Set	RMSE
Pre-Hike Variables	0.048
BFI Items	0.346
IPAQ Items	1.355
PCS Items	0.502

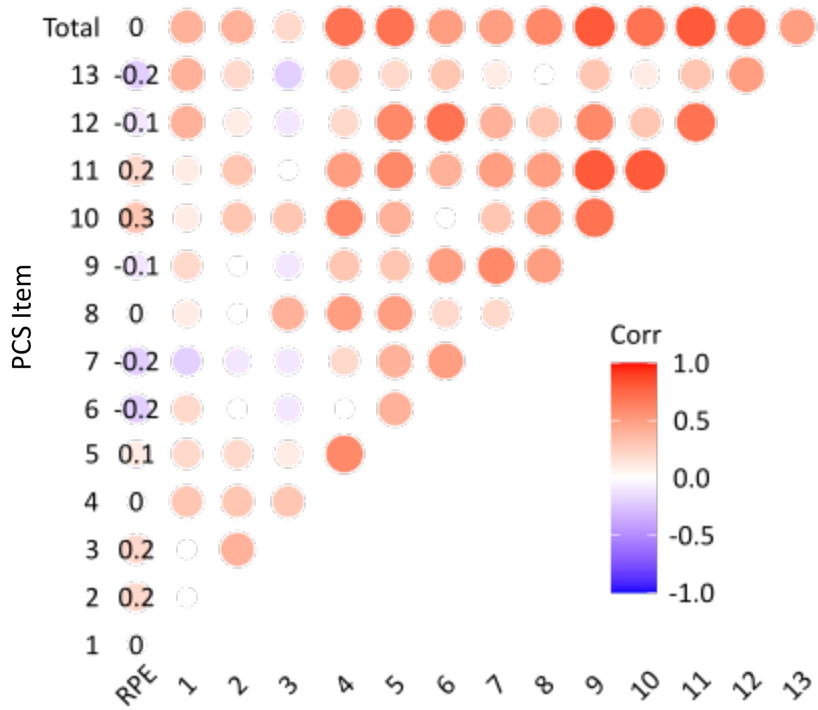
Note. All RMSE values here result from the elastic net models. This outcome variable was treated as a numeric variable. Pre-Hike Variables include those measured prior to the hike. BFI: Baecke Fitness Inventory; IPAQ: International Physical Activity Questionnaire; PCS: Pain Catastrophizing Scale; RMSE: root-mean squared error

Two approaches were used to identify important predictor variables: (1) the strength of correlation between predictor and outcome measures and (2) variable importance weighting arising from L1 regularization performed by the elastic net regression on the feature sets. Predictor measures and VRPE with high correlations were marked as potentially informative (Figures 2-2, 2-5, 2-6). From the PCS, items 7, 10, and 13 were identified as important both by the elastic net regression model's variable selection (Figure 2-3) and from their correlation with RPE (Figure 2-2). These variables had the strongest relationships with VRPE compared to the other PCS items. Additionally, PCS questions 10 and 13 (Figure 2-3) tended to occur often when assessing variable importance in the constructed models. PCS 2 and 3 were also among the variables with the strongest relationship to RPE. However, because these relationships were in the same direction as PCS 10 but weaker, they were excluded from future data collection. We also excluded PCS 6 and 11 because of their strong correlations with PCS 7 and 10, respectively.

From the IPAQ, days of vigorous exercise (IPAQ 1) and total time sitting (IPAQ 7) were identified as important (see Table 2-2 for item details). Days of vigorous exercise (IPAQ 1) had the strongest relationship with RPE relative to other items of the IPAQ (Figure 2-5) and also contributed most to the elastic net model, based on variable importance (Figure 2-4 A). Total time sitting (IPAQ 7) had a relatively strong correlation with RPE (Figure 2-5). Although IPAQ 3 (days of moderate exercise) contributed to the elastic net model (Figure 2-4 A), this item was correlated with IPAQ 1, which increases the likelihood of multicollinearity and therefore it was not used in the following investigations. IPAQ 2 (time spent performing vigorous activity on one day) was excluded as it showed no correlation

Figure 2-2

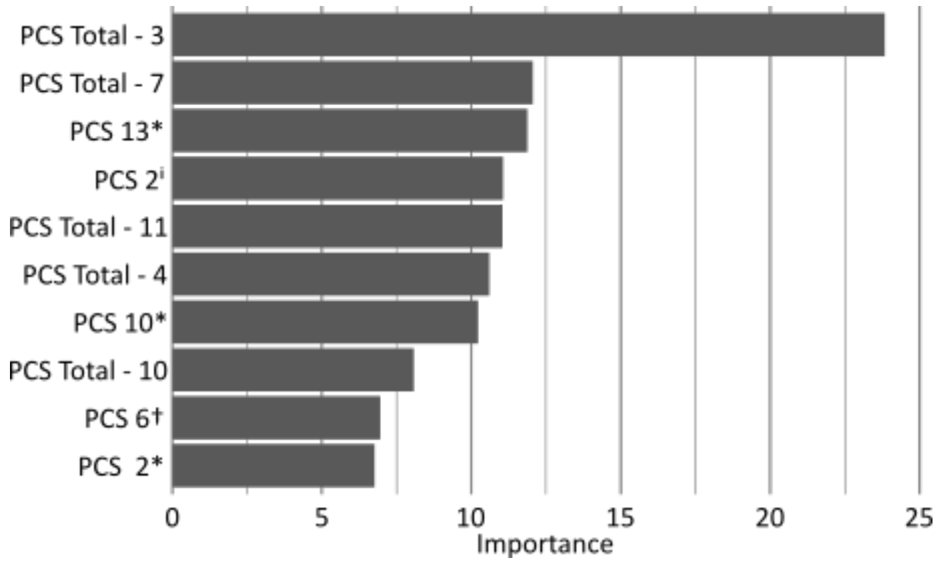
Correlations Between RPE of the Hike and PCS Items.



Note. Circle size maps to correlation coefficient magnitude, while circle color maps to correlation coefficient magnitude and direction. Coefficients are shown for the relationship between PCS items and RPE of the hike. The vertical and horizontal axes indicate the item number on the PCS.

Figure 2-3

Pain Catastrophizing Scale (PCS) Variable Importance

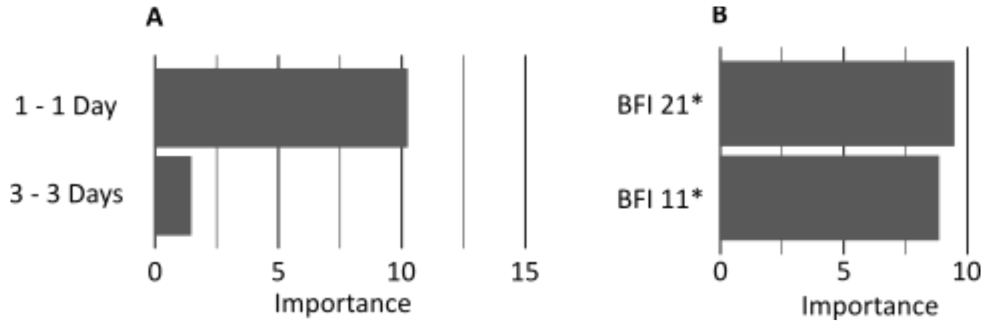


Note. The specific item or the total of all items is indicated on the vertical axis. ‘Total’ represents the sum of all PCS item ratings for a given participant, where lower totals indicate responses that suggest lower levels of catastrophizing and higher totals align with higher levels of catastrophizing. The length of the bars are the regression coefficients, and they represent a variable’s importance to the model.

The following superscripts indicate the level of response at which a questionnaire item was important: * to a slight degree, † to a moderate degree, ⁱ to a great degree. For example, PCS Item 13 was an important variable when the participant responded with ‘to a slight degree’.

Figure 2-4

Variable Importance of IPAQ and BFI Items

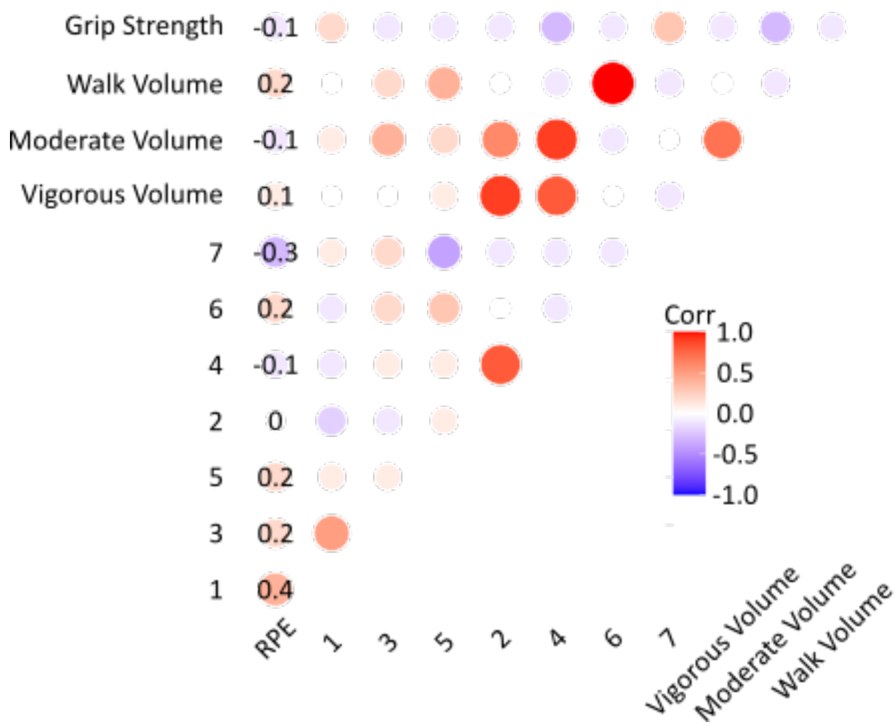


Note. A) Only two items from the IPAQ were identified as important using elastic net regression models. The specific item is indicated on the vertical axis. Item 1: During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling? Item 3: During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or double tennis? Do not include walking. B) Only two items from the BFI were identified as important using elastic net regression models. The specific item is indicated on the vertical axis.
* indicates that this item with the response 'sometimes' was important in the regression model.

with RPE and was not a top variable in the importance analysis. Similarly, IPAQ 5 (days spent walking 10 minutes or more) had a relatively weak relationship with RPE compared to other IPAQ items, and therefore it was also excluded moving forward. It is important to note that questions from the IPAQ and BFI (below) tended to overlap in their content and so questions from the IPAQ were preferentially pruned over the BFI due to the potential utility of the BFI's summary scores, which are described further below.

Figure 2-5

Correlations Between RPE of the Hike and IPAQ Items

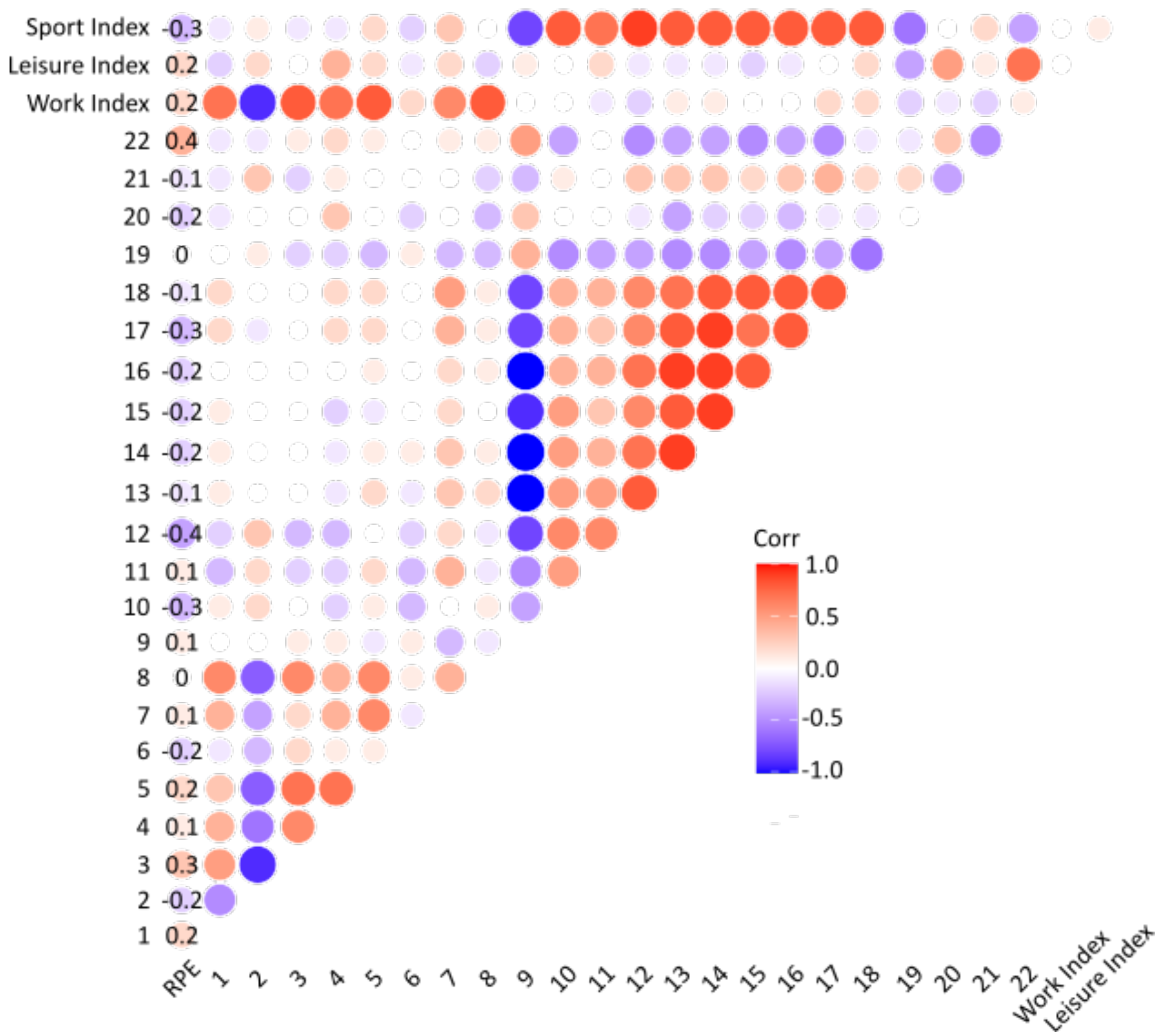


Note. Circle size maps to correlation coefficient magnitude, while circle color maps to correlation coefficient magnitude and direction. Coefficients are shown for the relationship between IPAQ items and RPE of the hike. Numbers on the vertical and horizontal axes indicate the item number on the IPAQ.

The BFI is unique among the selected questionnaires in that it uses three summary indices that are derived from the questions presented to participants. Surprisingly, the indices were not among the top variables in the elastic net model (Figure 2-4 B) but they were weakly correlated with RPE (Figure 2-6). Given that the BFI items were meant to work together and the correlation coefficients for the majority of BFI items were between 0.1-0.4, we chose to include this questionnaire in its entirety, rather than prune questions like in the IPAQ above, in the studies that follow. This allowed for flexibility in the use of the three

Figure 2-6

Correlations Between RPE of the Hike and BFI Items



Note. Circle size maps to correlation coefficient magnitude, while circle color maps to correlation coefficient magnitude and direction. Coefficients are shown for the relationship between BFI items and RPE of the hike. The vertical and horizontal axes indicate the item number on the BFI.

indices or individual BFI items in future models but at the risk of collinearity as a result of the correlations between some of the questions.

Grip strength showed a weak correlation with the hike's RPE compared to the questionnaire items above, and considering that individuals would generally not have access to a hand dynamometer, it was excluded from future data collection.

Table 2-3

Selected Questionnaire Items Used for Analysis

Item	Item Text
IPAQ 1	During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?
IPAQ 7	During the **last 7 days**, how much time did you spend **sitting** on a **week day**?
PCS 7	When I'm in pain, I keep thinking of other painful events.
PCS 10	When I'm in pain, I keep thinking about how much it hurts.
PCS 13	When I'm in pain, I wonder whether something serious may happen.
BFI – All Items	All items from the BFI were selected for inclusion in future studies.

Note. IPAQ: International Physical Activity Questionnaire; PCS: Pain Catastrophizing Scale; BFI: Baecke Fitness Inventory.

Discussion

Our overarching aims, here, were to determine the feasibility of collecting a range of variables potentially predictive of perceived exertion during a hike in order to identify and refine those variables most predictive of perceived exertion (Overarching Specific Aims 1 and 2). The specific questions we sought to answer in this investigation were:

1. Is a larger study based on similar methods feasible, considering the recruitment of participants, hiker interest, and the barrier presented by the required completion of a broad range of measures potentially necessary to make accurate predictions?

2. Which version of the RPE scale (Borg 6-20 or VRPE) might best differentiate between participants of varying fitness levels?

3. Which independent variables are the strongest predictors of RPE over the entire hike.

Overall, the rate at which participants were recruited, and the ultimate recruitment of the proposed sample over a reasonable period, suggests that a future study, using a larger sample and only a subset of the independent variables collected here, is indeed feasible. Going forward, the VRPE is the favored dependent measure because, as a continuous measure, it reduces the clustering of hikers into a few RPE ratings that we observed with the Borg 6-20 and consequentially may provide better discrimination between individuals. Finally, these results suggest that several independent (predictor) variables could be removed to reduce the load on future participants by selecting only those questions that contributed most to the model's predictive accuracy. These results are encouraging and suggest that a statistical model using a larger sample and a subset of the independent variables collected here may predict hikers' perceived exertion to a practical accuracy.

Efficacy of a larger sample size

Despite the time commitment required by individuals participating in this study, over 100 individuals expressed their interest in participating via the pre-screening questionnaire over a 4-month period. Nearly half of these individuals completed the entire study. The results of this study suggest that the number of questionnaires and activities completed before participation could be reduced significantly, decreasing the time burden placed on participants and potentially reducing the attrition rate in a future study. Considering these results, it appears that a larger sample could be recruited for a larger scale study and, importantly, that the majority of participants would follow researcher-provided instructions sufficiently, while hiking the trail without a researcher, to provide the data necessary for training statistical models to predict perceived exertion.

VRPE versus Borg 6-20

Notably, not only did the VRPE capture RPE with increased resolution than the Borg 6-20 but also exhibited a more uniform distribution over its range (versus a Gaussian-like distribution). The difference in distributions of the VRPE and Borg 6-20 suggest that these two scales may capture participant exertion information in differential ways. Specifically, the VRPE offers more rating levels with which participants can align their experience, while the Borg 6-20 uses fewer categories, many of which are anchored by text descriptors that might influence participant ratings. As a continuous variable, the VRPE offers hikers more options when reporting perceived exertion, but the same attribute could result in a noisier distribution. Its advantages are potentially more obvious when it is compared to the Borg 6-20's tendency to cluster participants into one or two moderate-level RPE categories. This clustering could potentially be problematic for models using this measure of individual performance as their dependent variable due to the information loss inherent to the reduced variability. Similarly, during hiking descent, the tight range around which the RPE values cluster when using either measure suggests that the descent may have little predictive utility on its own and, if combined with the ascent, may reduce the predictive capacity of some models. While further analysis is needed to fully address and understand this concern, consideration of this effect in future research may help to avoid problems with prediction of individual trail difficulty when using the VRPE.

Predictive independent variables

Narrowing down possible predictor variables was an important purpose of this investigation. To do this, we trained several models to evaluate the large number of predictor variables, focusing primarily on those items adding to participant time burden prior to hiking in order to lower the attrition rate. A model containing only wearable-derived and demographic variables as predictors would be ideal for a simple, streamlined data collection but would not be useful for predicting a hike's expected difficulty since the wearable measures are collected during the hike. On the other hand, a model containing all questionnaire items, demographic information, and wearable variables as predictors would place increased burden on participants to complete all questionnaires and tests at the trailhead, potentially hindering the model's usage. Since only a few variables from these questionnaires appeared to contribute to the model's predictive accuracy and some had strong correlations with each other, it appears it may be unnecessary to collect all of these items in order to effectively predict individual hiking intensity. Therefore, the use of only a subset of the variables collected in this study could decrease the time burden on participants in future studies while still producing an effective predictive model for hiking trail intensity. A direct comparison of predictions using simple survey data and wearable-derived variables could be a useful direction for future research investigating hiking trail difficulty. Researchers may also consider investigating the relationships between the predictor variables used here to better understand variable interactions and their impact on RPE.

Limitations

Due to the limited sample size, separate datasets were not used to train and test the statistical models, which could lead to inflated predictive accuracy values compared to what might be observed if assessed on a test data set. Consequently, little focus was placed on comparisons of the predictive capacity of the models.

Conclusion

The results of this investigation suggest that it is likely feasible to conduct a large-scale study investigating whether simple questionnaire data could be used to provide individual predictions of a hiking trail's difficulty. However, when implementing this design on a larger scale, it appears that, compared to the Borg 6-20, VRPE could allow for better separation of individuals of differing fitness levels based on the wider spread of ratings across the scale and the increased number of rating levels with which hikers can align their experience. In addition, we were able to identify a subset of items from the questionnaires that provided the best predictive capacity of the variables collected, which may allow for fewer questions asked before the hike and, consequently, a reduced time burden on the participants. Overall, the results of this study suggest that a larger-scale project could be feasible and have the potential to provide foundational knowledge for improving on current hiking trail rating systems.

Chapter 3

Introduction

Hiking is an increasingly popular recreational activity (Outdoor Industry Association & The Outdoor Foundation, 2024) that has relatively low barriers to entry and offers both the benefits of physical activity and exposure to nature (Kaplan, 2001; R. Ulrich, 1984; R. S. Ulrich, 1979). Despite the popularity, and obvious benefits, there is not a universal hiking trail rating system to consistently inform hikers of the difficulty of the trail on which they are about to embark. In fact, a variety of difficulty rating systems are used within the United States (Shenandoah National Park Virginia, 2017; Willamette National Forest, n.d.) and even more so globally (*International Hiking Trails*, n.d.). Few of these systems are seemingly underpinned by peer-reviewed research (Hugo et al., 1998; McIntosh & Hugo, n.d.), and no system offers individualized predictions of hiking trail difficulty. Considering the growing number of individuals engaging in hiking (Outdoor Industry Association & The Outdoor Foundation, 2024), there is a need for updated trail rating systems that can provide a personalized estimate of a trail’s difficulty. Such a system could improve individual hiking experiences by better aligning the hiker’s expectation of a trail’s difficulty with the actual trail difficulty, thereby improving their overall experience.

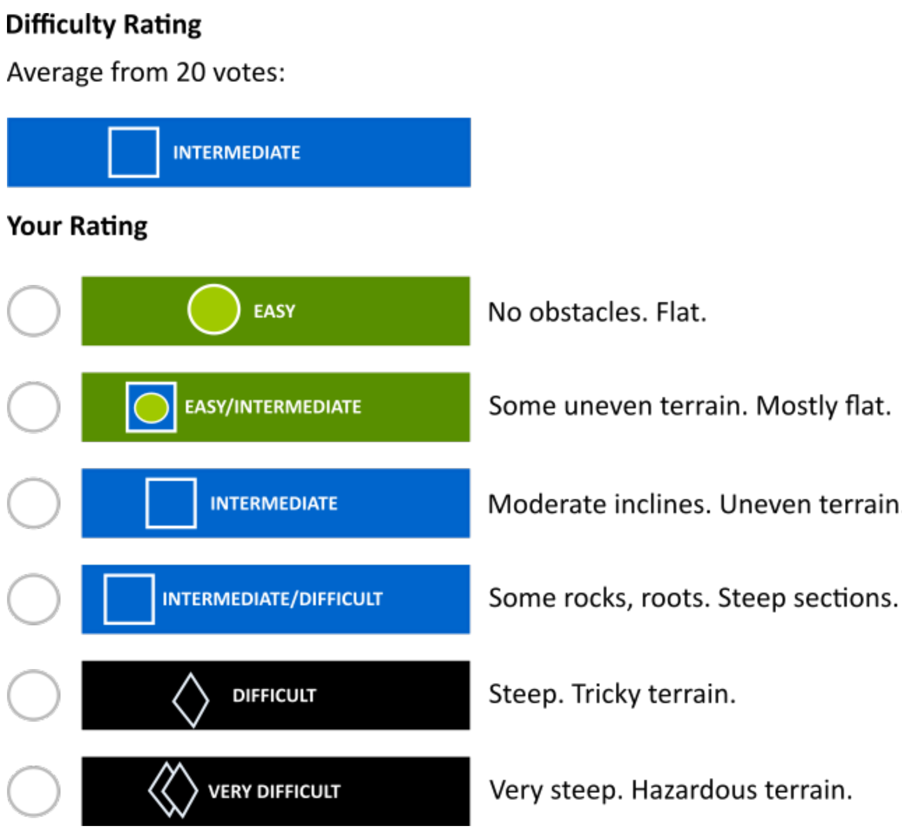
Currently, some hiking trail difficulty ratings systems provide only a single difficulty estimate for all hikers based on the characteristics of the trail and of the average user (Hugo et al., 1998; McIntosh & Hugo, n.d.; Shenandoah National Park Virginia, 2017; Willamette National Forest, n.d.). These systems seek to inform hikers about the trail intensity prior to their hike but fall short of considering individual characteristics that might also influence hiking trail intensity, such as age and fitness level. Two popular mobile applications, AllTrails and Hiking Project, aim to help hikers identify nearby trails and their current conditions. AllTrails provides one trail intensity rating for all hikers, much like one might see on an informational sign at the trailhead. While users of this application are encouraged to write a review of each trail, the review prompts do not ask for their evaluation of the perceived difficulty of the trail after they have completed the hike. Instead, these prompts include a rating (out of five stars), a free response section to provide up-to-date information about trail conditions, and a report of the activity performed on the trail (e.g., hiking, running, etc.). As of January 2023, AllTrails rates the Wind Caves Trail as moderate. With 45 million AllTrails users globally (as of January 2023), a more personalized evaluation of trail difficulty could benefit a significant number of people, especially entry-level hikers.

Hiking Project comes closer to providing a dynamic average difficulty rating by crowdsourcing hiking trail information, including trail difficulty, from individual users (Hiking Project, n.d.). An average difficulty rating is then assigned to the trail based upon the user-submitted ratings. While this approach appears to be drawing closer to a more accurate rating of trail difficulty, the difficulty scale used by this application focuses on trail features instead of hiker exertion (Figure 3-1). This mismatch between a hiker's perceived difficulty (which is often tied to exertion) and the difficulty defined by different trail terrain (and not physical exertion) may lead to confusion by users. An additional consideration is the low number of users of this app who choose to provide ratings for each trail. Only 20 user submissions contribute to the intermediate rating for the Wind Caves Trail on February 6, 2023 (Figure 3-1). Considering that this is a heavily-trafficked trail, 20 ratings may be too small of a sample to obtain a precise measure of the trail average and will likely be skewed towards more active individuals who use the application often. Ignoring the potential issues with the rating scale used in this app, its average trail difficulty rating is as close as we get to a predicted trail difficulty, and even then, this is not an individualized prediction.

The rating systems and mobile applications currently in use may benefit from the incorporation of an individual hiker's assessment of a hiking trail's difficulty. This relationship between trail and hiker is arguably one of the more important factors contributing to new hikers' engagement with the trail and, more generally, their hiking experience. When considering how to better integrate individual characteristics with hiking trail difficulty predictions, we should also evaluate the tools required for individuals to use a

Figure 3-1

Hiking Project Difficulty Rating Scale



Note. Reproduced from Hiking Project (*FAQ: Overview of Hiking Project Features*, n.d.).

newly developed system and whether these tools would be readily accessible for beginning hikers. For example, wrist-based wearables often track heart rate during exercise, which seems like a simple variable to use for individualized hiking trail difficulty predictions, but this approach would require the individual to own this tool and also to hike the trail prior to obtaining a difficulty prediction. Alternatively, fitness and demographic questionnaires that are free online and available prior to a hike could be a viable alternative to using heart rate data to predict hiking difficulty if the prediction accuracy is comparable to models using biometric data collected during a hike. To better match hikers' experience with trail difficulty, new means of estimating the overall difficulty of a trail relative to the individual should be explored, and simplification of tools required for an accurate trail difficulty estimate is needed.

Research Questions

This study sought to assess the accuracy with which statistical learning models could predict an individual's exertion during a hike and to evaluate the contribution each variable made to these predictions through the following research questions (Overarching Specific Aims 2 and 3):

1) Contingent on the available sample size, what is the accuracy with which statistical models can predict an individual's hiking trail difficulty, as measured on a visual analog RPE scale, during a hike on the Wind Caves trail?
2) Which predictor variables contribute most to the predictive capacity of these statistical models?

3) Does the incorporation of intra-hike information, such as trail segment and aggregate trail information, improve the models' predictive accuracy over that observed when using only variables available prior to the hike?

Methods

Trail

The hiking trail used to address these aims was the Wind Caves Trail in the Uinta-Wasatch-Cache National Forest in Cache County, Utah. This 1.8 mile out-and-back trail contains a variety of grades and varying trail features that will be potentially sufficient to discriminate between participants of differing fitness levels by demanding a range of physical abilities across different sections. Additionally, this trail is heavily trafficked, making it ideal for recruiting a large number of participants.

Participants

Individuals 18 years and older with an interest in hiking the Wind Caves trail and willingness to wear an activity tracker were recruited for this study through physical flyers posted in the surrounding community, in-class announcements, social media posts, digital signage, and by word of mouth. All individuals completed the Physical Activity Readiness Questionnaire (PAR-Q) prior to enrollment in the study to ensure there were no underlying conditions that contraindicated the level of physical activity required by this investigation. Individuals who met the pre-screening health criteria were enrolled in the study. Of the 353 individuals that completed the pre-screening survey, 108 (F: 58; M: 46 [n = 104]; 30.57 ± 12.12 yrs [n = 104]; 72.36 ± 14.56 kg [n = 97]; 171.25 ± 9.75 cm [n = 104]) participated in the hike. Four participants recruited at the trailhead did not respond to requests to complete the physical activity and demographic questionnaires and were excluded from analyses. All procedures were approved by Utah State University's Institutional Review Board (#11825).

Pre-Testing Procedures

Prior to arriving at the trailhead, interested individuals completed the pre-screening survey (Appendix E). Eligible individuals then completed the informed consent form, the Baecke Fitness Inventory (Appendix G), selected questions from the International Physical Activity Questionnaire (Appendix F) and Pain Catastrophizing Scale (Appendix H), and the demographic questionnaire (Appendix I), and scheduled a time to hike. This is the reduced pre-hiking packet based on the results of the feasibility study presented in Chapter 2 and applied to this new sample of participants. All surveys were hosted on REDCap (Research Electronic Data Capture) (Harris et al., 2009, 2019).

Testing Procedures

Upon arrival at the trailhead, participants completed a pre-hiking survey (Appendix J), which included items referencing circumstances subject to change before arrival at the trailhead (i.e., use of trekking poles, size of hiking group, etc.). If the participant planned to carry a pack while hiking, the weight of the pack was measured using a handheld digital force gauge (FGE-HXY Digital Force Gauge, Nidec-Shimpo Corporation, Kyoto, Japan). Next, participants were instructed to fit the commercially available fitness tracker (Garmin Forerunner 235, Garmin Ltd., Olathe, Kansas, USA) on their non-dominant wrist. A researcher checked the fitness tracker for proper fit, ensured data recording was enabled, and provided instructions to the participant on how to rate perceived exertion and document ratings during the hike (see below). Participants hiked the Wind Caves Trail (without a researcher) at a self-selected pace while providing perceived exertion ratings at predetermined times, as described below. Participants were free to take breaks when desired, and their time spent at the top of the trail was unrestricted. The activity tracker worn on the wrist collected biometric data throughout the hike, including heart rate, distance, date, time, latitude, and longitude, at a sampling rate of 1 Hz. When the hike was complete, participants returned all equipment to the research station at the trailhead.

Ratings of Perceived Exertion.

Participants recorded perceived exertion ratings on a provided clipboard using a pencil and paper packet. All participants completed a pre-hike, post-ascent, post-descent, and post-hike rating on a traditional seven-point hiking trail difficulty scale (Appendix K) and on a visual analog perceived exertion scale (VRPE) (Appendix C). Individuals hiking in a group were told not to discuss or share their ratings with other hikers in the group to improve the independence of each individual's perceived exertion ratings.

Data Processing

RPE was obtained from the visual analog RPE scales by measuring with a ruler, in millimeters, the distance from the zero anchor to the intersecting line drawn by the participant. All questionnaire and RPE responses were manually input into REDCap, and fitness tracker data (i.e., .fit file) were transferred to a secure university computer for storage and data analysis. Using the Python programming language (Python Core Team, 2023), all .fit files (the GIS data file used by Garmin equipment and software) were downloaded and saved as a text file. All subsequent data wrangling and analysis of the fitness tracker data (e.g., altitude, heart rate, latitude, longitude, speed, date, and time) were performed in the programming language R (R Core Team, 2021). Then, in R, all relevant data from the text files were combined (e.g., altitude, heart rate, latitude, longitude, speed, date, and time). To accommodate differing data lengths across participants, all measured and derived variables were summarized across the entire hike, then exported and merged. All demographic and fitness questionnaire data collected in REDCap were then exported and merged with fitness tracker data.

Weather Data.

Weather data were downloaded from climate.usu.edu using Logan Golf Station from the AGWX (Utah AgWeather) network as the recording station. Hourly recordings were used, and the time closest to a hiker's start time was used to identify the weather variables for that given hike. See Table 3-1 for a complete list of the weather variables used in this investigation.

Table 3-1

Included Weather Variables

Variable (units)	Measures
Relative humidity (%)	Maximum, minimum
Dew point temperature (°F)	Average, maximum, minimum
Air temperature from thermistor (°F)	Average
Air temperature from RH sensor PRT (°F)	Average, maximum, minimum
Wind speed (mph)	Average, maximum

Note. RH: relative humidity; PRT: platinum resistance thermometer

Trail Segment Data

To determine whether segment-based information improves model predictive accuracy, two trail segments were defined: the segment with the steepest grade and the segment following. As most of the Wind Caves trail is composed of a low to moderate incline, the steepest segment poses the greatest challenge to individual fitness and could serve as a way to separate individuals based on fitness level. The segment immediately after the steepest segment could also provide information about heart rate recovery, which differs based on training status (Darr et al., 1988). First, the trail segment (Segment 1) with the steepest grade was identified (~30% grade), and the start and end coordinates obtained using Google Maps were used to identify the fitness tracker data recorded during this segment. The same approach was used to gather information for the segment (~16% grade) immediately following (Segment 2) the steepest trail segment. Variables were then aggregated over each segment to provide additional predictor variables for use in predictive modeling of RPE over the entire trail.

Statistical Analysis

All statistical analyses were completed using R (R Core Team, 2021). Due to missing GPS data, two participants were removed prior to the fitting of statistical models, resulting in 106 subjects with the potential to be included in the final analysis. Of the 106 total observations, only 73 observations were complete, and therefore, only these observations were included in models fit with all predictor variables. For the remaining measures, participants were excluded if they were missing any of the variables necessary for a particular statistical model.

After limiting the dataset to only complete observations (n = 73), the dataset was partitioned into training (n = 63) and test (n = 14) datasets using RPE over the entire hike as the outcome variable and an 80/20 split of the whole dataset into the training dataset (80%) and the test dataset (20%). The training dataset was used to train all models built to address the aforementioned research questions, and the test dataset was held out to assess the final model's predictive accuracy on unseen data. All regression models were trained using five-fold cross validation. Elastic net regression was chosen as the main approach as it has the ability to handle over-parameterized models through regularization. Models were built with five different feature sets. Set 1 includes only information available prior to the hike that was provided by the hiker, including demographic data and questionnaire responses. This is the ideal feature set from a participant burden perspective as well as from a predictive capacity lens; it takes only five to ten minutes of a hiker's time, requires no additional equipment be used while hiking, and would allow a hiking difficulty prediction to be given prior to ever hiking a given trail. Set 2 is a small step up from Set 1 as it includes weather features obtained while the participant is hiking. This introduces intra-hike variables into the model while requiring no additional time or equipment from the hiker but does limit model application to a trail hiked only after another hike where weather data are available (unless weather prior to the hike is predictive of performance, which was not tested here). Sets 3, 4, and 5 introduce participant-level intra-hike performance variables. These three feature sets have the highest level of participant burden, as they require equipment use during the hike, which could also present a barrier to access if such a model was available for public use. While heart rate could be used as an outcome variable here, it was selected for predictor set inclusion as validation to address whether or not it greatly improves prediction over using variables available pre-hike (Sets 1 and 2 versus Sets 3-5). Each feature set with the included predictor groups is described in Table 3-2.

Table 3-2

List of Feature Sets.

Feature Set	Included Feature Groups	Description
Set 1	Pre-Hike	Variables collected before the hike, including demographic and questionnaires
Set 2	Pre-Hike + Weather	Set 1 + weather information
Set 3	Pre-Hike + Weather + Segment	Set 2 + variables from two trail segments
Set 4	Pre-Hike + Weather + Aggregate Trail	Set 2 + variables derived from the entire hike
Set 5	Pre-Hike + Weather + Aggregate Trail + Segment	Set 4 + variables from two trail segments

Note. Sets 1 and 2 present the lowest burden to hikers both prior to and during the hike. Sets 3, 4, and 5 contain participant-level intra-hike features and require equipment be used during the hike.

Once trained, all models were fit using the test data to assess predictive accuracy of each feature set. All elastic net regression models were built with the train function from the glmnet package (Friedman et al., 2010, 2023), and all data cleaning and analyses were conducted using R Studio. Predictive accuracy of all models was assessed using RMSE. Additionally, a model predicting the sample average RPE for each participant was used to provide a performance benchmark for the five feature set models. While the five feature set models predicted each individual's RPE for the hike as the dependent variable, the benchmark model predicted the average RPE of the sample for every hiker regardless of predictor variable values.

We also explored simple linear regression and multiple linear regression but found that these models overfit to the training data and, in turn, provided inaccurate predictions for the test data. In addition, we used principal component analysis as a feature reduction strategy to address the large number of predictor variables present in the dataset, but this did not improve model predictions above what was observed using elastic net regression.

Results

The regression model predicting the sample average RPE for each participant produced a training RMSE of 16.44 mm and a test RMSE of 17.79 mm. Elastic net regression models using the feature sets listed above (Table 3-2), resulted in comparable training RMSE values across feature sets (Table 3-3). Training these models on limited pre-hike variables (Set 1, Table 3-2) resulted in an RMSE of 15.09 mm on the training data and a prediction RMSE of 23.82 mm on the test data. Adding weather variables (Set 2, Table 3-2) resulted in the same training RMSE (15.09 mm) but an improvement in test RMSE (19.23 mm) – the top performing model based on test RMSE. An elastic net model using all available features (Set 5, Table 3-2) resulted in only a slight improvement in test data prediction accuracy (RMSE: 22.41 mm) compared to the pre-hike variable model (Set 1) but was outperformed by the pre-hike and weather model (Set 2). The range of training RMSE values obtained across varying parameters for each of the feature sets is shown in Figure 3-2.

Table 3-3

Results of Elastic Net Regression

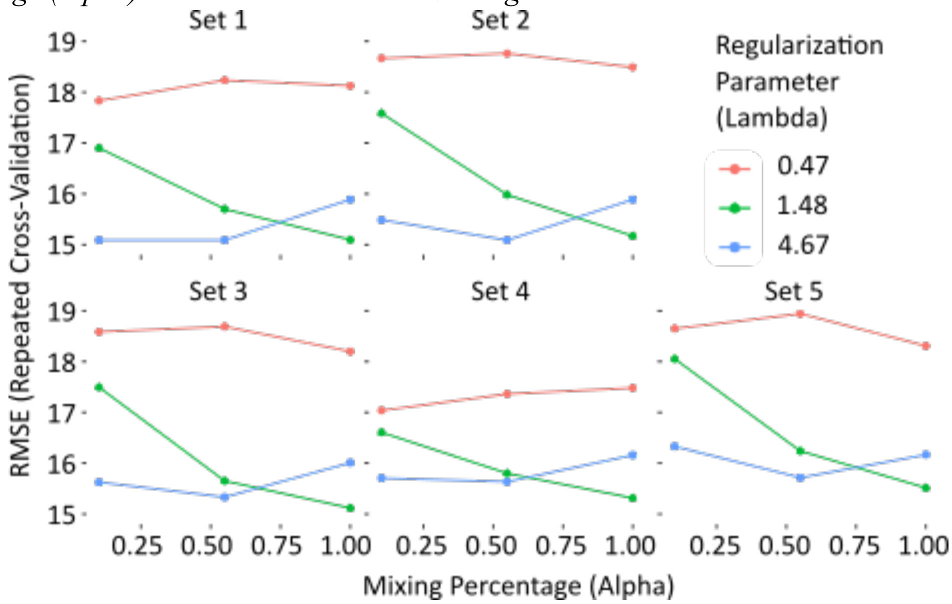
Feature Set	Included Feature Groups	Training RMSE ± SD	Test RMSE
Set 1	Pre-Hike	15.09 ± 2.12	23.82
Set 2	Pre-Hike + Weather	15.09 ± 2.85	19.23
Set 3	Pre-Hike + Weather + Segment	15.12 ± 3.59	21.35
Set 4	Pre-Hike + Weather + Aggregate Trail	15.31 ± 3.03	22.57
Set 5	Pre-Hike + Weather + Aggregate Trail + Segment	15.52 ± 3.19	22.41

Note. Set 2 resulted in the lowest test RMSE of the five feature set models, outperforming even Sets 3, 4, and 5 which included participant-level intra-hike variables. Additional feature set details are in Table 3-2.

L1 normalization, inherent to elastic net regression, reduces the impact of variables that contribute little to prediction, in effect selecting the variables of the greatest predictive importance. Elastic net regression revealed that multiple variables were consistently in the top ten most important variables across feature sets (Figure 3-3 and Table 3-4). All feature sets showed question ten of the Pain Catastrophizing Scale (PCS), specifically for those responding with ‘all the time’, as the most important variable. Sets 4 and 5 showed average heart rate normalized to estimated maximum heart rate as the second most important variable. The individual Baecke Fitness Inventory (BFI) questions were also consistently important, as was average speed normalized to individual maximum speed during the

Figure 3-2

RMSE Values Resulting from Varying Regularization Parameter (lambda) and Mixing Percentage (alpha) Parameters in Elastic Net Regression



Note. The regularization parameter adjusts the strength of regularization: lower lambda values create a more flexible model but with increased risk of overfitting to the data, whereas higher lambda values reduce risk of over fitting but can increase model bias. The mixing percentage weighs the amount of each of the two forms of regularization contributing to the model's fit. An alpha value of 0 reduces the model to ridge regression (a form of regression that helps control for multicollinearity), an alpha value of 1 reduces the model to lasso regression (a form of regression that performs variable selection), and values in between are a weighted combination of the two. During model training, a range of lambda (color) and alpha (x-axis) values are used to adjust the regression model and assess which combination of parameters results in the lowest RMSE. In this case, a low lambda (red lines) results in high RMSE regardless of the alpha value, while there appears to be a trade-off between mid and high lambda (green and blue) and alpha values across all feature sets. This information is then used to set the parameters of the final regression model. Each subplot shows the training RMSE for the feature set indicated by the title. Feature set details are provided in Table 3-2.

steepest trail segment, in models where segment info was included (Sets 3 & 5). Whether or not the individual hiked with others was also an important feature in all five models.

Table 3-4

Features Identified with Variable Importance Analysis

Feature Name	Feature Description
Average Speed Normalized – Segment 1	Average speed during steepest trail segment normalized by the individual's maximum speed during this segment.
Average Heart Rate Normalized	Average heart rate from the entire hike as a percentage of the hiker's estimated heart rate maximum.
Maximum Heart Rate Normalized – Segment 2	Maximum heart rate during the trail segment immediately following the steepest segment as a percentage of the hiker's estimated heart rate maximum.
PCS Total	Sum of ratings on three PCS items included in the questionnaires.
PCS 10*	When I'm in pain, I keep thinking about how much it hurts.
PCS 13*	When I'm in pain, I wonder whether something serious may happen.
BFI 1	What is your main occupation? Low activity (1), moderate activity (3), high activity (5)
BFI 3[a]	At work I stand:
BFI 5[a]	At work I lift heavy loads:
BFI 6†	After working I am tired:
BFI 7†	At work I sweat:
BFI 10	In comparison with others of my own age I think my physical activity during leisure time is: much more (5), more (4), the same (3), less (2), much less (1)
BFI 12†	During leisure time I play sport:
BFI 20†	During leisure time I walk:
Leisure Index	A composite measure of all responses to BFI leisure section items
Group	Binary variable indicating whether an individual hiked alone or with at least one other person.

Note. Features with the top 10 variable importance following elastic net regression. Estimated heart rate maximum was calculated using the equation: 220 - age. PCS: Pain Catastrophizing Scale; BFI: Baecke Fitness Inventory. Superscripts indicate response options for the noted questions.

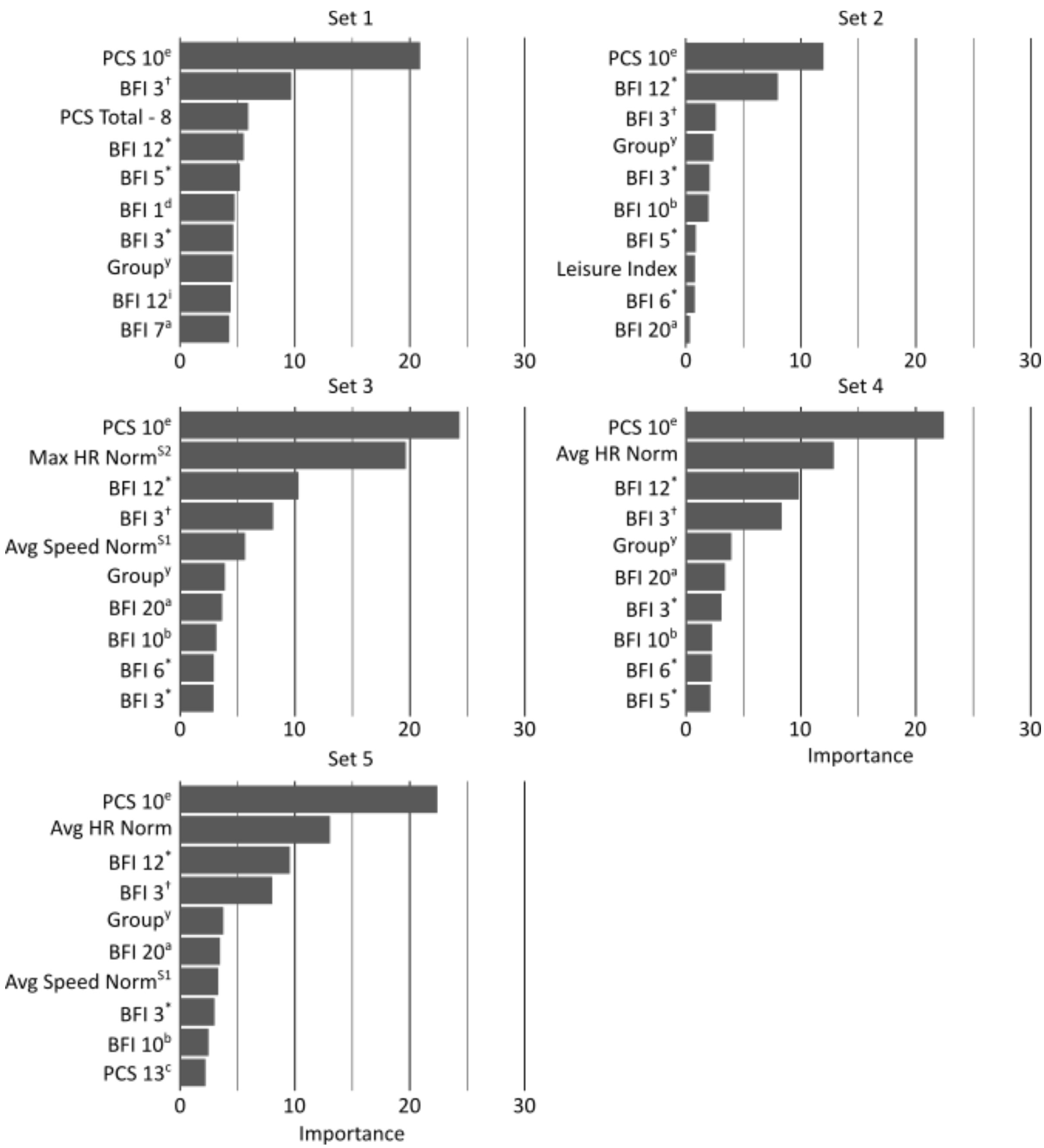
* Not at all (0), to a slight degree (1), to a moderate degree (2), to a great degree (3), all the time (4)

[a] never (1), seldom (2), sometimes (3), often (4), always (5)

† Never (1), seldom (2), sometimes (3), often (4), very often (5)

Figure 3-3

Top 10 Important Variables from Established Using Elastic Net Regression



Note. All sets showed item ten of the Pain Catastrophizing Scale as the top most important variable. Sets 4 and 5 showed average heart rate normalized to estimated maximum heart rate in the top two variables. Individual Baecke Fitness Inventory questions were consistently important, as were average speed normalized to individual maximum speed during the steepest trail segment and hiking in a group. The variables on the y-axis correspond to the variables, descriptions, and levels presented in Table 3-4.

*seldom, [i]sometimes, [†]always/all the time, [a]very often, [b]more, [y]yes, [c]to a slight degree, [d]high activity, [e]all the time, [S1]segment 1, [S2]segment 2

norm: normalized; PCS: Pain Catastrophizing Scale; BFI: Baecke Fitness Inventory

Discussion

The purpose of this investigation was to (1) assess the accuracy of statistical models predicting individual hiking trail difficulty rated on a visual analog RPE scale, (2) identify predictor variables important to the models' predictive capacity, and (3) determine whether aggregate trail features and segment-derived features improve model accuracy more than pre-hike variables. The results suggest that with this sample, pre-hike variables can predict VRPE to an accuracy of approximately 19-23 mm and that the addition of aggregate or segment information does not greatly improve accuracy (~22 mm RMSE), in this case. One variable identified here as most predictive, PCS item 10, was also an important predictor in the previous chapter, while no other predictive variables were consistently important between this chapter and the last. These findings suggest that basic questionnaire data (Set 1) and weather data (Set 2) may help improve the accuracy with which trail intensity, measured via perceived exertion, can be estimated. Importantly, the included questionnaires present relatively low barriers for individuals to complete if they are interested in using these algorithms to aid their hiking trail choices.

What is an accurate prediction of trail difficulty?

When assessing predictive ability, there must be a comparative standard to use as a frame of reference for the predictions being made. At this point in time, the standard likely includes either an estimate of trail intensity based on trail elevation and length, the trail intensity provided by the managing park system (NPS or local trail service), or trail intensity provided by an app (such as *All Trails* or *Hiking Project*). Standardized rating systems and general trail ratings provide only one rating meant for all individuals. As a result, these ratings do not consider the individual experience of each hiker and require that each hiker evaluate their fitness level relative to the standard, typically without knowledge of how the standard was developed. In contrast, approaches like the *Hiking Project* crowdsource trail ratings via their users, which may be an improvement from the ratings of static systems. However, it is important to consider the sample from which *Hiking Project* draws these ratings and whether this sample of users is representative of the general population. For example, these users are likely to be individuals who regularly engage in hiking (at least enough to warrant retaining a hiking-specific app) and thus, may be able to compare experiences across different trails. With this in mind, crowdsourced trail app predictions are likely to skew toward a more physically active population and therefore, might not accurately represent the general population or those new to hiking. While these estimates of the average trail intensity are far from perfect, an approach such as this provides an ideal tradeoff between estimation efficiency (easy to estimate) and prediction accuracy.

Comparable accuracy results were obtained by predicting the sample average RPE for all participants (test RMSE = 17.79 mm) and from a model including pre-hike and weather features (Set 2 test RMSE = 19.23 mm). This suggests that the information provided by these data does not improve upon an average intensity prediction for our sample. While the change in predictive capacity between an average and an individualized model may not be very different here, the observation that individual subject variables are important for the predictive capacity of the model suggests that with greater and more diverse data there is potential to improve individual engagement in hiking through better trail intensity predictions. While expanding to a larger, more heterogeneous training set could improve inferences drawn from these models, caution must be exercised to ensure that the models are not overfitting to the training data.

What is a meaningful change in VRPE?

Since there are no standardized norms to use as a comparison to different outdoor activity-related VRPE measures, it is challenging to evaluate the practical significance of the elastic net model's accuracy (RMSE). As a starting point, to determine what difference in RMSE is meaningful we might consider using minimum clinically significant differences reported in Visual Analog Scale (VAS) scores for other biological constructs. The minimum clinically significant difference in VAS scores for patient satisfaction is 7-11 mm (group SD: 10-15 mm) (Singer & Thode, 1998), 15 mm for nausea (95% CI 11.0 to 19.8 mm) (Hendey et al., 2005), 10-14 mm for pain regardless of severity (Kelly, 2001), and 10 mm for sleep quality (95% CI 8.0 to 12.6 mm) (Zisapel & Nir, 2003). While these differences are specific to the construct being measured, the use of a 100-mm visual analog scale to obtain subjective measures directly from the individual can provide a general guide for the meaning of changes on the visual analog scale used for measuring perceived exertion. The range of minimum clinically significant differences from 7-15 mm (with even larger 95% CI in some cases), places our model's RMSE just above the range of what is considered clinically significant for other biological constructs. While far from perfect—as this would suggest that, on average, the model's predictions are likely statistically different from the true scores—these values can provide a baseline from which to improve upon. Moreover, a small improvement in predictive capacity, for example an RMSE lower than the minimum clinically significant difference, could suggest that the predicted perceived exertion is close enough to the actual perceived exertion that it would, on average, place an individual in the correct exertion 'category', depending on the resolution of the categorical scale used.

Which Variables Improved Model Predictive Capacity?

Beyond predictive capacity, the models presented here also provide insight into which predictor variables should be considered if trying to predict RPE during a hike. Out of 101 features, 16 regularly appeared in the top ten features across the five feature sets. Of these variables, one was heart rate, which should be expected, as heart rate is a strong predictor of exertion during physical activity. Age-normalized average heart rate (as a percent of maximum heart rate) also helps to control for age-related changes in mean heart rate. Whether or not individuals hiked in a group was also important to multiple models, suggesting that there is some social impact on RPE when hiking, which will be explored in Chapter 5.

Most of the top features were questionnaire-derived variables. These include multiple items from the Baecke Fitness Inventory, as well as items from and the total score of the Pain Catastrophizing Scale. The identified Baecke Fitness Inventory items captured information about physical activity during work and leisure, as well as an individual's perceived fitness level relative to their peers. This last variable is unique as it not only compares individuals of differing fitness levels but also asks for one's subjective perception of their fitness relative to their age group. The items from the Pain Catastrophizing Scale also provide unique information compared to other questionnaire items. Specifically, these items ask about the individual's response to painful circumstances or events (Sullivan et al., 2000). Item 10 of the Pain Catastrophizing Scale, in particular, asks individuals to what degree they keep thinking about how much it hurts when they are in pain. Surprisingly, this variable was the top most important variable in each feature set, particularly for individuals that responded with 'all the time'. Individuals that tend to fixate on pain or discomfort during the hike have a tendency to experience higher RPE over the hike, which may be related to the relationship between greater catastrophizing thoughts and greater cardiovascular response (Lentini et al., 2021). Finally, segment-specific information (Set 3) did slightly improve predictions of RPE in the training set compared to including only aggregate trail information (Set 4). Inclusion of the normalized average speed during the steepest segment improved training RMSE compared to models using only aggregate data, suggesting that there is unique information provided by this segment-specific feature but not as much as might be expected. Although the addition of trail information, whether segment-specific or aggregate, did not greatly improve model accuracy over using only pre-hike variables.

Elastic net regression with only pre-hike variables (Set 1, test RMSE = 23.82 mm) was comparably accurate to an elastic net regression model that included all predictor variables (Set 4, test RMSE = 22.41 mm), while a model including pre-hike variables and weather during the hike outperformed both (Set 2, test RMSE = 19.23 mm). This suggests that the addition of participant-specific variables measured during the hike may not meaningfully improve RPE predictions and thus, with further refinement, pre-hike variables and publicly available weather data may be all that are necessary to provide a functional personalized prediction of RPE prior to a hike. However, further work is clearly necessary to refine these methods.

Limitations

A larger and more diverse sample would likely have improved the predictive capacity of the model as well as its generalizability, making the final model more accurate and applicable to a wider range of individuals. Collection of a larger sample would also have permitted use of larger statistical models that may better capture the variability in the sample.

Here, only 31% of individuals that submitted the prescreening survey completed the hike. Examination of the relatively high attrition rate could provide clues to improve retention of those individuals interested in participating, which could increase the efficiency of data collection in the future.

Conclusion

This study highlighted the accuracy with which elastic net regression can predict perceived exertion during hiking, given this specific trail and sample characteristics. It also provides direction for future investigations by identifying and confirming which variables may be important to prediction, such as whether the hiker is carrying a pack and the hiker's response to painful situations, that may improve predictions of VRPE while hiking. Additionally, information about hiking performance broken down by segment may improve VRPE prediction accuracy during hiking. The highest prediction accuracy was observed when a combination of participant-level pre-hike data and intra-hike weather data were included in the model. However, these results could likely be improved with a larger sample size and correspondingly larger predictive models, and ultimately the goal of providing individual predictions of trail intensity may be realized.

Chapter 4

Introduction

Hiking trail intensity is traditionally rated on a difficulty scale with five to seven levels (Hugo, 1999b) where each trail is assigned one difficulty level that is used for all hikers. This approach is widely used and accepted, but scientific support for it is limited (Hugo, 1999b; Hugo et al., 1998). Despite this, the difficulty scale has several benefits, including distinct levels rooted in the estimated bioenergetic demands of the trail (Hugo, 1999a) and labelled using common terms (e.g., fair, moderate, extreme) that align relatively closely with hikers' conceptualization of trail difficulty. It is important, however, to recognize that the intended use of this scale is to evaluate the bioenergetic cost of a trail and not an individual's perception of its difficulty (Hugo, 1999b). Because the trail rating is not a measure of experienced intensity, inconsistent ratings may arise if the mapping between the scale's levels, and hiker's perception of the intensity associated with each level, varies between participants. While this 'rating the trail' approach might be suitable for generalization, if non-linearities are introduced by individual differences in mapping category names to their experiences, it may pose challenges to accurately predicting an individual's perception of trail difficulty. In the previous projects, I have used this difficulty scale (Hugo, 1999b) to provide categorical descriptions of different intensity ratings (Chapter 2). However, it is important to better understand how this categorical rating scale compares to the numerical visual analog rating of perceived exertion (VRPE) scale, used in previous chapters, for predicting individual hikers' perceptions.

In contrast to the aforementioned categorical measure of difficulty, perceived exertion can be measured on a linear visual analog scale from 0 to 100 mm, using 1 mm increments. A benefit of the VRPE is that a user's rating of perceived exertion is not dependent upon their interpretation of the descriptive term associated with different intensity levels (e.g., 'moderate'). Instead, users indicate where on the range of exertions possibly experienced this current exertion is without use of descriptive terms that have the potential for misinterpretation, other than the descriptive terms used for the scale anchors. This approach allows for a more precise measure of perceived exertion, as it permits a gradation of hiking exertions free from the constraints of language. One drawback is that a VRPE value might not have meaning for many people (e.g., Is moderate exertion at 60 mm or 50 mm?). Thus, it might be difficult for hikers to score their current exertion using this measure within the broader context of their possible exertion without repeated exposure to the scale. As the visual analog scale has no positional, only extreme, anchors, its scores are likely to have higher intra-rater and inter-measure variability, which could make prediction more difficult.

While both RPE measurement scales have their limitations, it remains unclear whether the difference in resolution (numeric vs. categorical) between these scales impacts the accuracy of models predicting individual perceived trail difficulty. Considering the limited predictive accuracy observed using VRPE (numeric) in Chapter 3, this Chapter aims to evaluate whether predictive accuracy might be improved by transforming the continuous measure into a discrete measure and descriptively comparing its performance as a categorical measure to that of the categorical difficulty scale. By recoding VRPE as a categorical variable, we aim to achieve two objectives:

1. Reduce variability in participant's VRPE ratings through pooling: By grouping similar intensities into discrete categories, the analysis becomes less susceptible to outliers and variation of ratings using the VRPE scale.
2. Minimize the impact of individual differences in how participants map category names (e.g., 'moderate') to their perceived experience: Discretizing the numerical VRPE data may help counter this source of variability and thus provide a more accurate representation of participant's perceptions.

This approach allows for a comparison between two commonly used methods for measuring individual hiking trail difficulty and may offer insights into more effective ways to collect and analyze RPE data while improving a model's capacity to predict individualized trail difficulty.

Research Questions

This study aims to descriptively investigate the benefits of discretizing VRPE data on a statistical model's predictive capacity by comparing the accuracy of a model predicting trail perceived exertion (a proxy for difficulty) using the discretized VRPE to a model predicting ratings of trail difficulty using a traditional categorical rating scale (Overarching Specific Aim 4).

Methods

Trail

The hiking trail used to address this aim was the Wind Caves Trail in the Uinta-Wasatch-Cache National Forest in Cache County, Utah (Appendix D). This 1.8 mile out-and-back trail contains a variety of grades and varying trail features to potentially allow for better discrimination between participants of differing fitness levels. Additionally, this trail is heavily trafficked making it the best local option for recruiting participants.

Participants

Individuals 18 years and older with an interest in hiking the Wind Caves Trail and willingness to wear an activity tracker were recruited for this study through flyers posted around the community, in-class announcements, social media posts, digital signage, and by word of mouth. All individuals completed the Physical Activity Readiness Questionnaire (PAR-Q) prior to study enrollment to ensure there were no underlying conditions that contraindicated the level of physical activity required by this investigation. Individuals who met the pre-screening criteria were enrolled in the study. Of the 353 individuals that completed the pre-screening survey, 108 participated in the hike (F: 58; M: 46 [n = 104]; 30.57 ± 12.12 yrs [n = 104]; 72.36 ± 14.56 kg [n = 97]; 171.25 ± 9.75 cm [n = 104]). Four participants recruited at the trailhead did not respond to requests to complete the set of physical activity and demographic questionnaires and, therefore, were excluded from the analysis. All procedures were approved by Utah State University's Institutional Review Board (#11825).

Pre-Testing Procedures

Prior to arriving at the trailhead, interested individuals completed the pre-screening survey (Appendix E). Eligible individuals then completed the informed consent form, a demographic questionnaire (Appendix I), the Baecke Fitness Inventory (Appendix G), selected questions from the International Physical Activity Questionnaire and the Pain

Catastrophizing Scale (Appendix H), and scheduled a time to hike. All surveys were hosted on REDCap (Research Electronic Data Capture) (Harris et al., 2009, 2019).

Testing Procedures

Upon arrival at the trailhead, participants completed a pre-hiking survey (Appendix J), which included items referencing circumstances subject to change prior to the hike (i.e., use of trekking poles, size of hiking group, etc.). If the participant planned to carry a pack while hiking, a researcher measured the weight of the pack using a handheld digital force gauge (FGE-HXY Digital Force Gauge, Nidec-Shimpo Corporation, Kyoto, Japan). Next, participants were instructed to properly fit the commercially available fitness tracker (Garmin Forerunner 235, Garmin Ltd., Olathe, Kansas, USA) on their non-dominant wrist. A researcher then checked it for proper fit, ensured data recording enabled, defined rating of perceived exertion, and provided instructions for documentation of RPE and difficulty during the hike (see below). Participants hiked the Wind Caves Trail (without a researcher) at a self-selected pace and provided ratings of perceived exertion and difficulty at predetermined locations, as described below. Participants were free to take breaks when desired, and their time spent at the top of the trail was not restricted. The activity tracker worn on the wrist collected biometric data throughout the hike, including heart rate, distance, date, time, latitude, and longitude, at a sampling rate of 1 Hz. When the hike was complete, participants returned all equipment to the research station at the trailhead and rated the entire trail difficulty on a seven-point scale (Appendix K).

Individuals recruited at the trailhead completed the prescreening survey to assess their eligibility for participation. If eligible, participants read and signed the informed consent document, completed the pre-hiking survey, and then followed the same pre-hiking preparation and data collection steps described above. These participants were then emailed a request to complete the Baecke Fitness Inventory (Appendix G), selected questions from the International Physical Activity Questionnaire (Appendix F) and the Pain Catastrophizing Scale (Appendix H), and demographic questionnaire.

Ratings of Perceived Exertion.

Participants recorded perceived exertion ratings with a clipboard, pencil, and paper packet. All participants completed a pre-hike, post-ascent, post-descent, and post-hike rating on a traditional hiking trail difficulty seven-point scale (Appendix K) and on the VRPE (Appendix C). If hiking in a group, participants were instructed to avoid sharing or discussing ratings.

Data Processing

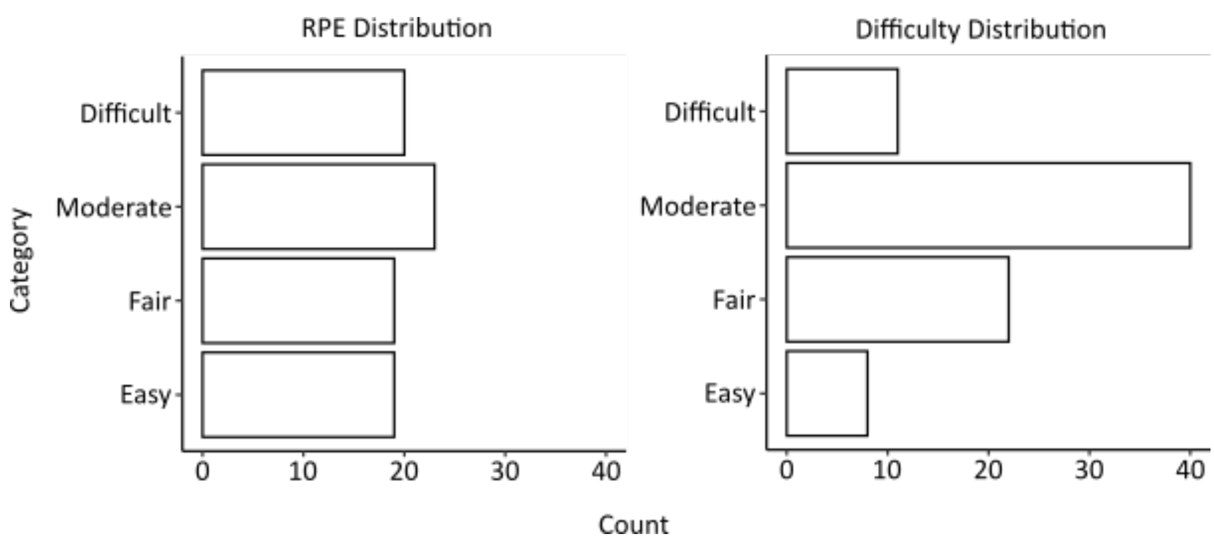
RPE was obtained from VRPE scales by measuring with a ruler in millimeters the distance from the zero anchor to the intersecting line drawn by the participant. All VRPE and difficulty responses were manually input into REDCap, and fitness tracker data (i.e., .fit file) were transferred to a secure university computer for storage and data analysis. Using the Python programming language (Python Core Team, 2023), all .fit files from the fitness tracker (the GIS data file used by Garmin equipment and software) were downloaded, converted to, and saved as a text file. All subsequent data wrangling and analysis of the fitness tracker data (e.g., altitude, heart rate, latitude, longitude, speed, date, and time) were performed in R (R Core Team, 2021). All demographic and fitness questionnaire data collected in REDCap were then exported and merged with fitness tracker data.

Statistical Analysis

To make comparisons between models using VRPE and those using difficulty as the response variable, VRPE was recoded to a categorical variable (Figure 4-1), and classification was applied to both outcome variables. Hiking difficulty was measured on a seven-point categorical scale, but participants used only five of these levels. VRPE was measured as a numeric variable from 0 to 100 and was recoded by dividing the 100-point scale into 7 equal levels to match the levels of the difficulty scale (under the assumption that the difficulty scale levels increase linearly like the VRPE scale). Very few values were in the two extreme categories, therefore the bottom two levels were combined as well as the top two levels. This resulted in a categorical RPE scale with five levels. Both categorical response variables (difficulty and RPE) resulted in level sparsity during classification due to very few observations in the 'Severe' category. Therefore, these observations were recoded

Figure 4-1

Distribution of Categorical Response Variables



Note. Each histogram represents the distribution of RPE and difficulty on a categorical scale and displays the levels used in the classification models.

for inclusion into the next highest level, ‘Difficult’, which resulted in four levels (‘Easy’, ‘Fair’, ‘Moderate’, ‘Difficult’) of each response variable used for classification.

Due to missing GPS data, two participants were removed prior to the fitting of statistical models, resulting in 106 subjects with the potential to be included in the final analysis. The missing variables for these two subjects meant that the models would automatically exclude them for predictions whenever a predictor variable was missing from the dataset. Similarly, since there were only 81 ‘data complete’ observations, models including all predictor variables were limited to this sample size. The samples used for predictive modeling, including those with only a subset of all predictor variables, were restricted to observations that were ‘data complete’ for that respective model (i.e., all the data required for that particular model was present).

After limiting the dataset to only data complete observations (n = 81), the data were partitioned into training and test datasets using an 80/20 split for each of the two response variables (categorical RPE and difficulty). Models were trained using the training dataset to address the aforementioned specific aim, and the test datasets were withheld to assess the final model’s predictive accuracy on unseen data for a given response variable. All models were built with a feature set that included only variables available before the hike using five-fold cross-validation to assess accuracy on the training set. This feature set included pre-hike questionnaire responses and demographic and hiking equipment information.

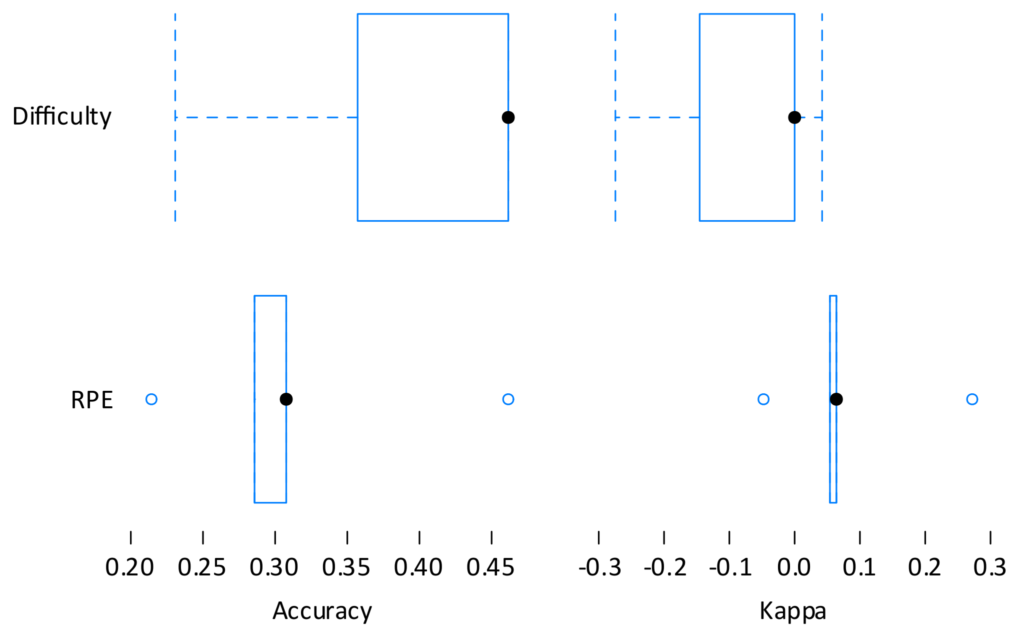
Elastic net classification was used as the main statistical approach as it has the ability to avoid overfitting when over-parameterized. Model performance was assessed using classification accuracy, with higher accuracy indicating better performance. All models were built using the ‘train’ function from the caret package (Kuhn, 2008; Kuhn et al., 2023), and all statistical analyses were completed using R (R Core Team, 2021).

Results

Categorical VRPE resulted in a more uniform distribution of scores across the possible ratings compared to the difficulty ratings. After recoding VRPE responses, elastic net classification resulted in a test accuracy 28.57% for RPE and 53.33% for difficulty. The training accuracy ranges obtained via cross-validation for each response variable are shown in Figure 4-2. The accuracy variability of the difficulty classifier was greater during training but its accuracy was on average higher than that of the discrete RPE classifier (Figure 4-2 Left). Compared to a model predicting at random chance levels, the RPE classifier was a small improvement, as indicated by a Cohen’s Kappa value consistently greater than zero. In contrast, the difficulty classifier’s performance was comparable to random chance, with a Kappa value closer to zero, although with greater variability. The confusion matrices from the models’ test predictions, presented in Figure 4-3, provide further detail about model performance on the test data. The RPE classifier demonstrated a more accurate categorization at low intensity RPE predictions (‘easy’, ‘fair’) compared to high intensity RPE predictions (‘moderate’, ‘difficult’), whereas the difficulty classifier always predicted ‘moderate’. Due to clear differences in covariation between the measured and predicted scores in the discretized RPE versus difficulty, we performed a Spearman rank correlation on each variable to determine the strength of the relationship between the measured and predicted values. There was a moderate, positive (rho = 0.34) relationship between the measured and predicted categorical VRPE scores but an undefined relationship between the measured and predicted difficulty scores (due to the lack of variability in the predicted difficulty scores). This result suggests that while the discretized RPE performed worse on average, it better captured participant outcome variability across changes in RPE compared to the difficulty classifier.

Figure 4-2

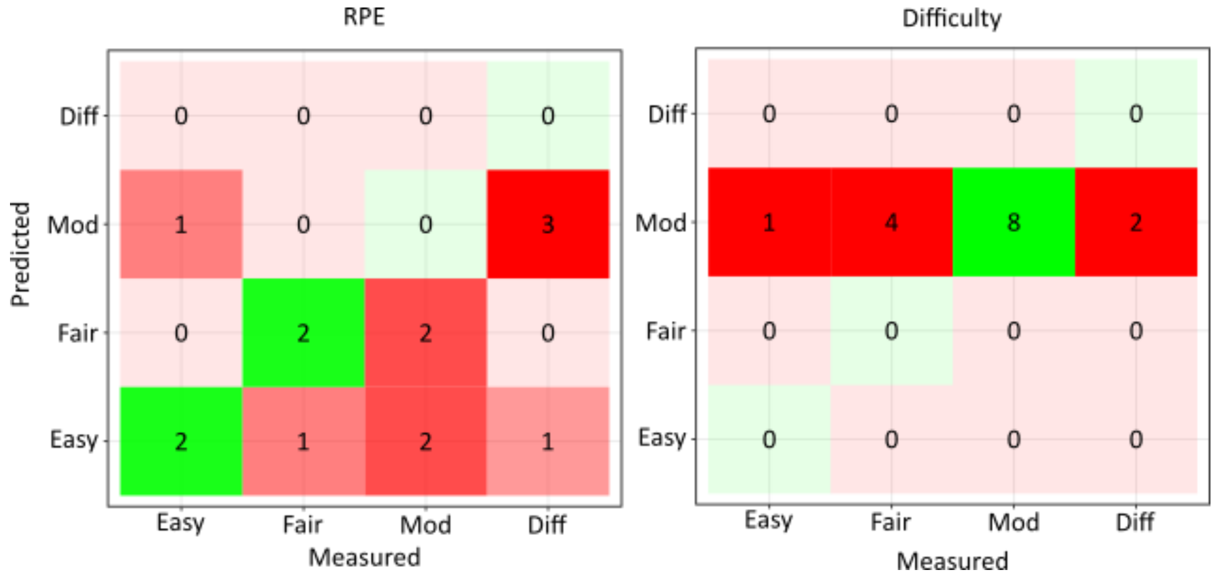
Classification cross-validation for the training set



Note. Model accuracy (left) and kappa (right) for difficulty and categorical RPE during the training of the classification model. Left: Accuracy indicates the number of observations classified correctly for each round of cross-validation. Contrasting the classification of difficulty and RPE, difficulty nearly always out-performs RPE based on accuracy. This measure is reported as a percentage. Right: Kappa compares the observed accuracy to the expected accuracy (random chance) and presents the classification accuracy of the model being tested relative to a model classifying based only on random chance. Kappa ranges from -1 to 1, where 0 indicates the classification accuracy is equal to random chance, -1 is less than random chance, and 1 indicates perfect agreement between the observations and model outcomes. In terms of Kappa, the RPE classifier performs slightly better than the difficulty classifier, which performs similar to a random chance classifier. In all subplots above, the box, tails, and dots represent the spread of values observed during cross-validation for accuracy and kappa.

Figure 4-3

RPE and Difficulty Confusion Matrices for the Test Data



Note. RPE (left) was classified correctly in the 'easy' and 'fair' categories but was misclassified in the 'moderate' and 'difficult' categories. Difficulty (right) was always classified as 'moderate'. The measured category is represented on the x-axis, and the category in which the model classified the observation is shown on the y-axis. If a model were to classify correctly all values in every category, the squares along the diagonal would be green with no other squares highlighted in red. A square's color indicates whether the model's classification category was correct (green) or incorrect (red). The opacity of each square reflects the proportion of all measured observations in a given level that were classified correctly or incorrectly. The more opaque, the greater proportion of measured values covered by that specific classification (predicted) category. The more transparent, the lower the proportion of measured values represented by the associated square.

Discussion

This investigation sought to examine whether recoding of the numerical VRPE measure as a categorical variable would improve prediction accuracy. Ultimately, difficulty classification resulted in better test accuracy (53.33%) than for categorical VRPE classification (28.57%). However, this improved accuracy for difficulty came as a consequence of predicting a moderate intensity for every hiker. The results of this study suggest that model accuracy was superior for difficulty compared to categorical VRPE, although the difficulty model sacrificed individual variability for accuracy by predicting 'moderate' for every hiker. In contrast, the model classifying VRPE appeared to capture the subject-to-subject variability more effectively, especially at lower intensity levels.

Non-linear mapping of difficulty levels

One concern of using the difficulty scale to indicate how challenging participants found the trail is the possibility of there being variation in how participants interpreted the relationship between the scale's category names (i.e., 'easy', 'fair', 'moderate', etc.) and the exertion that they experienced. The VRPE likely bypasses this particular effect by using descriptive anchors only at the scale extremes and avoiding categorical associations. After the VRPE was recoded to four categorical levels, its distribution was more uniform than the difficulty distribution, suggesting that indeed the mapping between degree of exertion and the ratings provided between the two measures differed at least to some degree. Participants were more likely to associate their experience with a moderate difficulty when presented with the categorical difficulty scale than to select an exertion level on the VRPE approximately equivalent to this level. This finding raises questions as to whether the categorical labels acted to bias the scoring of participants (e.g., participants selected moderate more often for psychological or social reasons rather than mark their true experience) or whether participants had difficulty associating their exertion with a spot on the VRPE line resulting in more variable scores and the appearance of a uniform distribution as a result of our categorical recoding. Future research is necessary to disentangle these possible influences on these dependent measures.

Model prediction accuracy

Accuracy is a simple way to communicate results and compare classification models, however, it provides only surface-level information about the total number of observations classified correctly and neglects level-specific information about where the model performs poorly. For example, while classification accuracy of VRPE was lower than that of difficulty (as seen in the confusion matrices above), VRPE classification performed well at lower intensity levels ('easy', 'fair'), and the predicted RPE appeared to covary with the measured RPE across all levels. However, because only a moderate relationship (rho = 0.34) exists between the two variables, a larger sample size is necessary to determine if this relationship is significantly different from zero. A simulated power analysis using this correlation magnitude suggests that approximately 64 subjects would be necessary to establish whether this relationship is significantly different from zero with greater than 80% power.

Although the classification accuracy was superior for difficulty compared to VRPE, this outcome came at the cost of individual variability. Every difficulty observation was classified as 'moderate' and thus this model did not consider individual variability in making its predictions. While this outcome counters the goal of this thesis (to use information supplied by participants to improve and provide personalized predictions), given the relatively 'normal' distribution of the data in the test and training set, this may be the contextually optimal outcome. Specifically, for normally distributed continuous numerical data, the estimate that minimizes mean absolute and squared errors are the median and mean of the distribution (measures of central tendency), assuming there is no knowledge available regarding the mechanism of the generative model. Since the data here are ordinal categorical data, the median or mode of the data may represent a similar compromise. Given the size of this sample, and the relative weakness of the relationship between the predictor variables and the outcome variables, the model may not have been able to perform better than predicting a single central, or most probable, difficulty because no benefit was gleaned from information provided by the individual participants. Notably, assigning a single difficulty to every hiker is essentially how difficulty scales are currently applied, and these data indicate that this approach may only be accurate approximately 53% of the time. If we extend this finding to our current use of difficulty scales, and assume a similar distribution of data to that observed here, it suggests that 46.67% of individuals would potentially be misinformed of the trail difficulty and find it to be either harder or easier than anticipated (moderate). Considering that we want to decrease barriers to hiking engagement, especially for beginner or leisurely hikers, and these individuals are likely to have lower physical fitness and familiarity with hiking trails, an accuracy rate of 53.33% is less than desirable, although it is comparable to previous findings (Geurkink et al., 2019). In contrast, since the categorical VRPE scores were fairly uniformly distributed, the advantage to predicting the center of the distribution, or most probable category, was not as large and therefore the model's accuracy seemingly benefitted from incorporating participant information into the predictions over simply predicting a single value (category) for everyone. Compared to the difficulty model, the categorical VRPE model resulted in a lower average classification accuracy but did classify observations in the ‘fair’ and ‘easy’ categories better compared to the ‘moderate’ and ‘difficult’ categories (though the consequences of categorizing a hard trail as easy are greater than categorizing an easy trail as hard). Interestingly, while the VRPE model captured some of individual variability it also tended to underestimate three of the four categories. The reason for this is currently unclear.

While there are pros and cons to using either of these response variables, combining them may provide a better solution. A scale that combines the flexibility of the VRPE with a traditional difficulty scale, where VRPE ranges are stratified by difficulty level using terminology that is easily interpretable by all hikers, could circumvent the absent meaning, to those unfamiliar, of a 0 to 100 scale but would also allow for regression-like resolution. Additionally, a hybrid scale could help limit the impact of variation in participants’ interpretation of the relationship between the category names and the difficulty they experience. Such a scale could better standardize participant responses, potentially increasing response consistency within and between individuals.

Limitations

Ultimately, while there was an increase in accuracy when classifying difficulty compared to VRPE, the low overall accuracy of both provides an opportunity for improvement. As stated earlier, it is likely that training these models on a larger, more diverse training sample and testing them on a larger test set could result in greater predictive accuracy for both measures. Additionally, changes to the experimental design, such as using a hybrid scale or further refining the included features, may improve the predictive accuracy of such models.

Conclusion

This study aimed to determine whether converting the numerical VRPE measure to categorical would enhance predictive accuracy. Our findings suggest that classifying trail difficulty results in better test accuracy than discretized RPE. However, this improved accuracy was due to the model predicting a moderate intensity for every hiker, rather than capturing the subject-to-subject variability more effectively. These findings suggest that while treating VRPE as a categorical measure may not necessarily improve prediction accuracy, it could potentially lead to improved outcomes if the correlation between these two measures can be strengthened. Future research should explore ways to address this correlation and optimize prediction accuracy, perhaps by combining scales and refining the predictor variables to allow for improved response accuracy and reliability.

Chapter 5

Why didn't predictive models perform better? Closer examination of trends in the data

When interpreting the results of previous chapters at face value, it appears that predicting individualized hiking difficulty using statistical learning methods does not greatly improve upon current hiking trail rating systems and, therefore, may not be overly useful. One factor contributing to the relatively poor predictive accuracy may be that individuals, whether active (Rose & Parfitt, 2007; Spurway, 1992; Zamparo et al., 2001) or sedentary (Parfitt et al., 2006), tend to self-select an exercise intensity around their lactate threshold, which often falls within the ACSM guidelines of 50-85% VO_2max for moderate exercise (Glass & Chvala, 2001; Parfitt et al., 2000). Importantly, when self-selecting exercise intensity, individuals report higher RPE (as measured with Borg 6-20) than they report when a below-lactate intensity is imposed (Parfitt et al., 2006). The ability to self-regulate hiking speed and take breaks during the study likely led participants to hike at a moderate exercise intensity, on average, which may have led to the moderate-centered RPE distribution. This self-selection of moderate intensity exercise may partially explain the tendency of models presented in previous chapters to also predict near the average VRPE for the entire group or classify every observation as moderate difficulty. Due to this, future investigations might consider a study design that implements experimental controls for hiking speed, break number, and break lengths during the hike.

Regardless of potential limitations resulting from individuals self-selecting hiking intensity, further investigation into the data revealed that interactions between sub-populations within the sample may be reducing the predictive capacity of the models above. Identification of these trends may be useful to improve future efforts toward predicting hiking trial difficulty. In particular, both an individual's prior trail experience and whether or not they hiked in a group appear to moderate the relationship between several variables, which could negatively impact the predictive capacity of some models if these variables are not controlled for, or there is insufficient sample size to account for them. Below, several variables are plotted by whether a participant had hiked the trail before (Black) or not (Gray) (Figure 5-1) and by whether or not they hiked in a group (Figure 5-2).

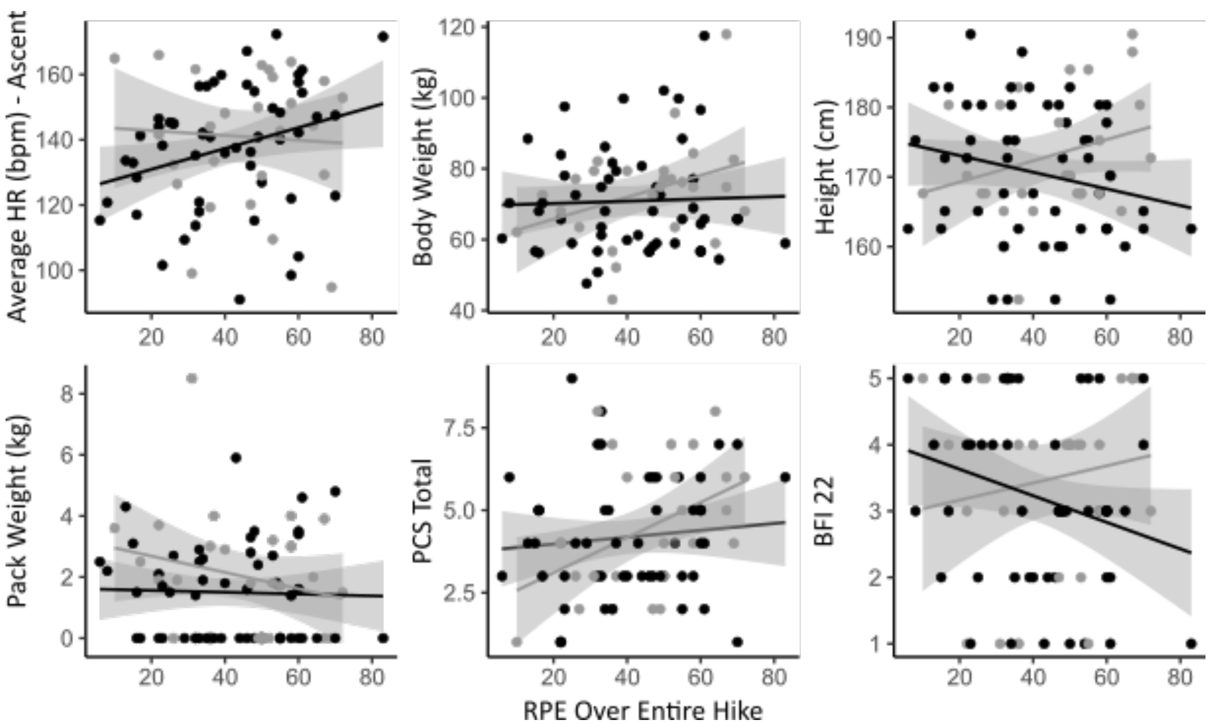
Previous Trail Experience

The first sub-populations we identified were repeat-hikers (hiked the Wind Caves trail previously (n = 52)) and first-time hikers (never hiked this trail before (n = 27)). This grouping variable appears to influence the relationship between VRPE over the entire hike across a variety of predictor variables, including those that typically have a consistent relationship with RPE, such as heart rate during the hike (Borg, 1982). Select scatter plots with linear model fits are shown in Figure 5-1.

Familiarity with the Wind Caves trail, and hiking as physical activity, may be playing a role here, as repeat- and first-time hikers report hiking for different reasons, including exercise and health (Wilcer et al., 2019). Familiarization with exercise may explain differences in motivational and affective responses to self-selected exercise intensity in sedentary women (Rose & Parfitt, 2012). Whether our model's predictive capacity was hindered by trail-specific or general physical activity familiarity is unclear. However, there does appear to be some influence of first-time hiking trail exposure on several predictor variables in this sample that were explored in previous chapters. For example, when

Figure 5-1

Relationship Between Predictor Variables and VRPE Grouped by Previous Experience on the Wind Caves Trail



Note. Different behaviors may be observed depending on whether or not the participant had previous experience on the trail. Gray = first-time hiker. Black = repeat-hiker. The shaded region indicates the 95% confidence interval for predictions from a linear model.

considering average heart rate's utility in predicting the rating of perceived exertion, we see that trail-unfamiliar hikers tend to rate perceived exertion higher for lower heart rates compared to trail-familiar hikers. Heart rate was not the only predictor variable to exhibit what appears to be different behavior for trail-familiar and unfamiliar hikers. Height and BFI 22 also appeared to exhibit differing trends across RPE depending on whether the hiker had experience with the trail (Figure 5-1). Importantly, we included these variables here to try to account for their influence. However, there may ultimately have been insufficient power and/or model complexity to use the information provided by these relationships, potentially leading to poorer predictive capacity of the models.

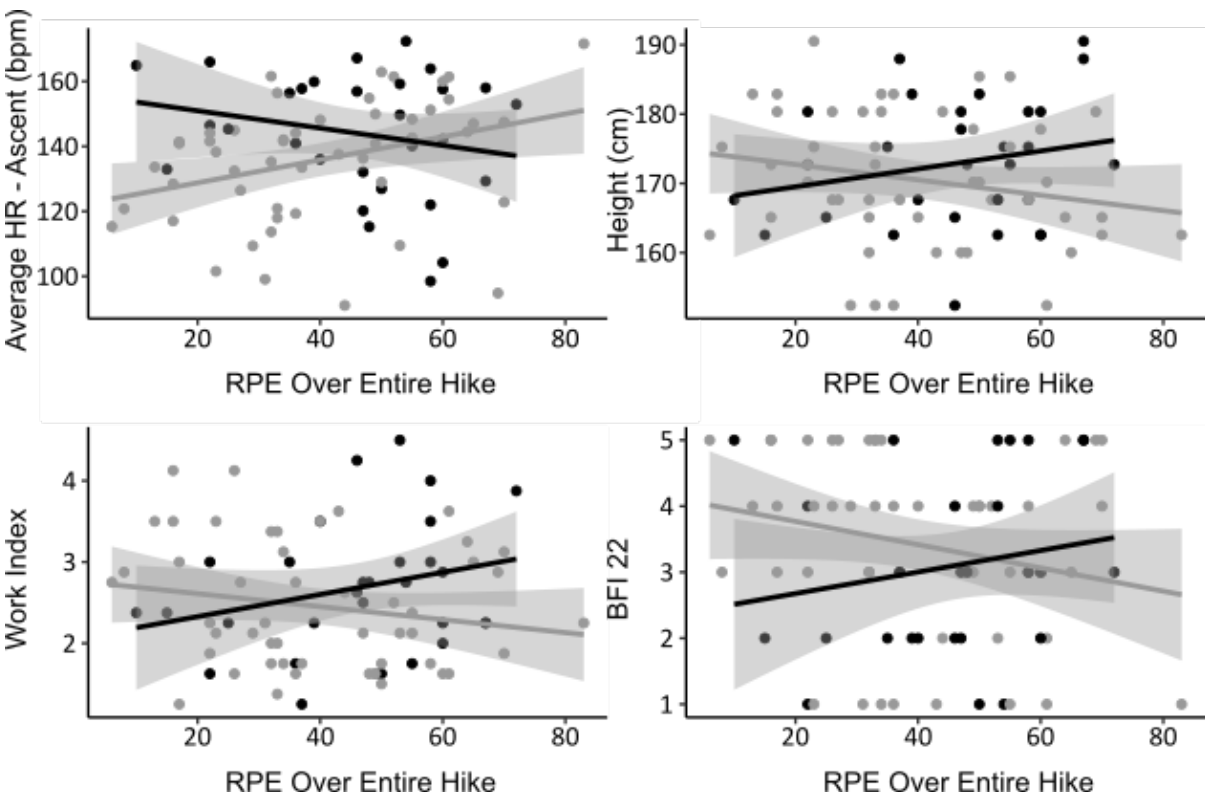
Group Hiking

The second moderating variable we identified as potentially influencing the relationship between select predictor variables and VRPE is whether individuals hiked with at least one other person (Figure 5-2). In our sample, 50 people hiked alone and 29 hiked in a group. Hiking in a group introduces a social component into the hike and likely alters a participant's hiking velocity and when and/or how many breaks are taken during the hike so that the group can stay together.

Underlying motivations, such as autonomy (Coble et al., 2003) and social interaction (Cronan et al., 2008; Kim et al., 2015; Whiting et al., 2017), may drive individuals to hike solo or partake in group hiking. At the same time, fears associated with hiking alone (e.g., getting lost, accidental injury, injury by another person (Coble et al., 2003)) may also contribute to an individual's solo hiking experience or decision to forego solo hiking altogether. Individual motivations and fears related to hiking contribute to a complex experience involving not only physiological measures considered here but also psychological

Figure 5-2

Relationships Between Predictor Variables and VRPE Grouped by Solo and Group Hikers



Note. Opposite behaviors may be observed in measures of RPE depending on whether or not participants hiked in a group. Gray = Solo-hiker. Black = Group-hiker. The shaded region indicates the 95% confidence interval for predictions from a linear model.

and social factors. These complex, personal decisions that guide an individual to hike solo or in a group may also influence the measures presented here, leading to identifiable differences in each type of hiking group but washing out the predictive accuracy of models not considering these relationships. Ultimately, it would seem that these studies would benefit from a larger sample size by better capturing the relationship between the independent variables and potentially opening the door for the use of more complex models.

Conclusion

These results suggest that heterogeneity within our sample may be impairing models' capacity to accurately predict individual hiking trail difficulty at this sample size. A larger overall sample size may potentially allow inclusion of these variables in the model if they continue to behave differently. However, if such a dataset is unavailable, future researchers may consider focusing only on one sub-population, such as individuals that have never hiked the trail before or those solo-hiking, to reduce the confounding impact of these relationships. Another option is to impose a particular hiking speed on all individuals in an effort to control for the effect of self-selected intensity tending toward a moderate level of exercise in all individuals. Either of these approaches may improve model training and the accuracy of their predictions.

Chapter 6

General Discussion and Conclusions

Improving hikers' self-awareness of physical fitness, trail conditions, and trail difficulty may decrease recreational injuries and incidents and positively influence overall hiking experience (Heggie & Heggie, 2012; Trayers, 2004). Others seeking to align individual fitness level and trail difficulty to assess hiking readiness report weak relationships between popular fitness questionnaires (International Physical Activity Questionnaire) and individual RPE on a hiking trail (Coetzee et al., 2021). The alternative suggestion to implement standardized fitness assessments (i.e., step test and one-mile walking test) presents its own barriers to informed hiking trail difficulty, namely required time, equipment, and additional physical exertion for prospective hikers (Coetzee, 2018; Coetzee et al., 2021). We proposed an alternative approach to determine individual hiking trail difficulty using statistical models to predict RPE during a hike using trail, individual, and weather feature sets. Our results are encouraging when considering the potential application to improve individual hiking experiences, but there are still many important areas to explore and many unanswered questions.

Methodology Considerations

A prevailing limitation across our aims is that the sample size is smaller than ideal and, consequently, likely limits the predictive capacity of the models presented, as well as additional statistical approaches that could be explored. While imputation could be used to replace missing data and increase the sample size, this approach can lead to biased estimates when applied under specific circumstances (Donders et al., 2006). Similar studies have used smaller samples and applied complex machine learning models after imputation of the original sample to create a larger dataset (Davidson et al., 2020; Geurkink et al., 2019). These publications report classification accuracy of upwards of 86% to predict RPE in a variety of sports contexts but only after simplifying the problem by reducing RPE to two larger categories ("Somewhat hard to hard" (RPE ≤ 15) vs. "Hard to very hard" (RPE > 15)) (Davidson et al., 2020) or using a "loose accuracy" approach where classification of predictions within ±1 of the observed value was considered correct (Geurkink et al., 2019). While imputation could be used to replace missing data and increase the sample size, this approach can lead to biased estimates when applied under specific circumstances (Donders et al., 2006). Another option is to explore time series analysis using the raw data extracted from the fitness tracker. This could be a promising approach as it has previously been used to predict fatigue and RPE (Hajifar et al., 2021) with relative success.

As seen in recent studies predicting fatigue via participant-reported RPE, an alternative approach to this problem could be to focus on a small sample of hikers that repeat the hiking task multiple times and consider each hike as separate training observations (Bartlett et al., 2017; Carey et al., 2016; Vandewiele et al., 2017). This approach would provide information from which the models could learn how RPE changes within a smaller group of individuals based on the time of year, environmental factors, and as the hiking season progresses. While a model trained on a smaller sample may still encounter issues of overfitting to these individuals' characteristics, it may allow for greater focus on hiking performance and fluctuating weather conditions.

Social Impact on Perceived Hiking Difficulty

Hiking with one or more other people may influence the level of exertion experienced during a given hike (Coetzee et al., 2021). Beyond hiking faster to keep pace with group members with higher physical fitness, there may be some level of distraction that makes the perceived experience less intense than objective measures (such as heart rate) suggest. This may lead to lower RPE levels despite heightened physiological measures and necessitate the inclusion of a group interaction in regression-based models. This change in the relationship between certain variables may, in turn, impact model performance when predicting RPE.

Motivations and hiking styles that characterize different hikers may also influence subjective perception of hiking difficulty. Long-distance hikers on the Pacific Crest Trail have been categorized into two groups, “purists” and “social hikers”, with each group having distinct approaches to and motivations for hiking (Lum et al., 2019). Purists primarily seek connection to the trail itself and are engaged solely for the act of hiking, while Social Hikers are characterized by a desire to partake in emotional connection to others (Lum et al., 2019). While there is currently no literature that directly connects hiking characterization and subjective trail difficulty, it is worth considering when predicting individual hiking experience based on a subjective measure, such as rating of perceived exertion.

Similar relationships between gender and RPE may also exist as women tend to have higher concerns about confidence and fear of getting hurt despite sharing similar interests in outdoor recreation as men (Ceurvorst et al., 2018). The subjective and multi-faceted nature of RPE may mean that personal attributes, specifically psychological factors, may impact an individual’s perception of trail exertion. As described here, an individual’s level of general pain catastrophizing appears to be correlated with their perceived hiking experience as measured with RPE.

Generally, this dataset would also benefit from more observations in older age groups, as well as broader variety of fitness levels. Anecdotally, when recruiting participants, it was difficult to recruit older individuals who had never hiked the trail before due to fears of getting hurt, lost, or needing help while on the trail. As this study required physical activity, specifically hiking, it likely suffered from participation bias towards those who enjoy hiking and physical activity. While this is anecdotal, a greater focus on non-hiking audiences during recruitment and implementation of a variety of participation opportunities may increase recruitment of individuals less likely to naturally volunteer for a study such as this. In light of this, future studies might consider organizing group hikes, not only for older hikers but for anyone who feels more comfortable hiking with others. This approach may help with diversifying the age range as well as the type of hikers who are willing to participate in the study, especially if a more challenging hike is used.

Conclusions

Overall, this series of projects lays a foundation for future research within a relatively under-developed subject area by illustrating the feasibility of a real-world study of this nature, exploring preliminary data analytics for prediction of individual hiking difficulty, and identifying future directions through the identification of variables that appear important to predicting individualized trail difficulty. While the results of this project are encouraging, we are limited by sample size, participant demographics, and self-selected hiking intensity. To make the approach presented here a truly viable approach to rating trail difficulty, it will need to be extended to other hiking trails, particularly those with more difficult terrain to further differentiate between hikers of differing fitness levels. A first step may be to expand it to include other trails within the Wasatch region to understand how models need to be built for different trails and how individual performance on different trails will impact predictions. Next, expansion to trails in different regions of the United States will also allow us to explore samples with different demographics and understand what these differences might indicate about perceived hiking difficulty based upon regional and geographic differences.

References

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: Evidence for an individualized approach. *International Journal of Sports Physiology & Performance*, *12*(2), 230–234.

BC Parks Foundation. (n.d.). *PaRx: A Prescription for Nature*. PaRx. Retrieved December 23, 2023, from https://www.parkprescriptions.ca/

Borg, G. (1962a). A simple rating scale for use in physical work test. *Kungliga Fysiogra Ska Sallskapets i Lund Forhandlingar, 2*, 7–15.

Borg, G. (1962b). *Physical performance and perceived exertion.* Gleerup.

Borg, G. (1970). Perceived exertion as an indicator of somatic stress. *Scandinavian Journal of Rehabilitation Medicine*, *2*(2), 92–98.

Borg, G. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, *14*(5), 377–381.

Borg, G. (1998). *Borg's Perceived exertion and pain scales*. Human Kinetics.

Borràs, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, *41*(16), 7370–7389. https://doi.org/10.1016/j.eswa.2014.06.007

Boyd, C. N., Lannan, S. M., Zuhl, M. N., Mora-Rodriguez, R., & Nelson, R. K. (2018). Objective and subjective measures of exercise intensity during thermo-neutral and hot yoga. *Applied Physiology, Nutrition, and Metabolism*, *43*(4), 397–402. https://doi.org/10.1139/apnm-2017-0495

Calbimonte, J. P., Martin, S., Calvaresi, D., & Cotting, A. (2021). A platform for difficulty assessment and recommendation of hiking trails. In W. Wörndl, C. Koo, & J. L. Stienmetz (Eds.), *Information and Communication Technologies in Tourism 2021* (pp. 109–122). Spring, Cham.

Calbimonte, J. P., Martin, S., Calvaresi, D., Zappelaz, N., & Cotting, A. (2020). Semantic data models for hiking trail difficulty assessment. In J. Neidhardt & W. Wörndl (Eds.), *Information and Communication Technologies in Tourism 2020* (pp. 295–306). Springer, Cham. https://doi.org/10.1007/978-3-030-36737-4_24

Calbimonte, J. P., Zappellaz, N., Hébert, E., Simon, M., Délétroz, N., Hilfiker, R., & Cotting, A. (2018). SanTour: Towards personalized recommendation of hiking trails to health profiles. In C. Pautasso, F. Sánchez-Figueroa, K. Systä, & J. M. Murillo Rodríguez (Eds.), *Current Trends in Web Engineering* (Vol. 11153, pp. 238–250). Springer International Publishing. https://doi.org/10.1007/978-3-030-03056-8_23

Carey, D. L., Ong, K., Morris, M. E., Crow, J., & Crossley, K. M. (2016). Predicting ratings of perceived exertion in Australian football players: Methods for live estimation. *International Journal of Computer Science in Sport*, *15*(2), 64–77. https://doi.org/10.1515/ijcss-2016-0005

Casey, J., Herron, R., Bishop, S., Ryan, G., & Bishop, P. (2015). *Validity of a visual analog scale for assessing RPE*. https://doi.org/10.13140/RG.2.1.3802.4802

Ceurvorst, R. L., Fuller, R., Childers, D., Dubois, M., & Steiger, S. (2018). University Outdoor Recreation Program Participation: Comparing Benefits, Constraints, and Willingness to Pay Between Males and Females. *Journal of Outdoor Recreation,* *Education, and Leadership*, *10*(3), Article 3. https://doi.org/10.18666/JOREL-2018-V10-I3-9047

Chen, M. J., Fan, X., & Moe, S. T. (2002). Criterion-related validity of the Borg ratings of perceived exertion scale in healthy individuals: A meta-analysis. *Journal of Sports Sciences*, *20*(11), 873–899. https://doi.org/10.1080/026404102320761787

Chrusch, A., & Kavin, M. (2021). Survey of Musculoskeletal Injuries, Prehike Conditioning, and On-Trail Injury Prevention Strategies Self-Reported by Long-Distance Hikers on the Appalachian Trail. *Wilderness & Environmental Medicine*, *32*(3), 322–331. https://doi.org/10.1016/j.wem.2021.04.004

Coble, T. G., Selin, S. W., & Erickson, B. B. (2003). Hiking Alone: Understanding Fear, Negotiation Strategies and Leisure Experience. *Journal of Leisure Research*, *35*(1), 1–22. https://doi.org/10.18666/jlr-2003-v35-i1-608

Coetzee, B. A. (2018). *Self-reported fitness levels, actual fitness levels and recorded energy expenditure on graded hiking trails* [Thesis, University of the Free State]. http://scholar.ufs.ac.za/xmlui/handle/11660/9041

Coetzee, B. A., Bloemhoff, H. J., Coetzee, F. F., & Schall, R. (2021). Relationship between hikers' self-reported physical activity responses and ratings of perceived exertion. *African Journal for Physical Activity and Health Sciences (AJPHES)*, *27*(4), 424–438. https://doi.org/10.37597/ajphes.2021.27.4.2

Cronan, M., Shinew, K., & Stodolska, M. (2008). Trail use among Latinos: Recognizing diverse uses among a specific population. *Journal of Park and Recreation Administration*, *26*, 62–86.

Dag, F., Tas, S., & Cimen, O. B. (2021). Hand-grip Strength is Correlated with Aerobic Capacity in Healthy Sedentary Young Females. *Montenegrin Journal of Sports Science and Medicine*, *10*(1), 55–60. https://doi.org/10.26773/mjssm.210308

Darr, K. C., Bassett, D. R., Morgan, B. J., & Thomas, D. P. (1988). Effects of age and training status on heart rate recovery after peak exercise. *American Journal of Physiology-Heart and Circulatory Physiology*, *254*(2), H340–H343. https://doi.org/10.1152/ajpheart.1988.254.2.H340

Davidson, P., Düking, P., Zinner, C., Sperlich, B., & Hotho, A. (2020). Smartwatch-derived data and machine learning algorithms estimate classes of ratings of perceived exertion in runners: A pilot study. *Sensors*, *20*(9), Article 9. https://doi.org/10.3390/s20092637

De Villiers, S. W., & Thiart, B. F. (1988). ‘n Moeilikheidsgradering van Voetslaanpaaie in die RSA met Riglyne vir Selfevaluasie deur Voetslaners. *South African Journal for Research in Sport, Physical Education and Recreation*, *11*(1), 37–46.

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014

Ekkekakis, P., & Lind, E. (2006). Exercise does not feel the same when you are overweight: The impact of self-selected and imposed intensity on affect and exertion. *International Journal of Obesity*, *30*, 652–660.

Fogliato, R., Oliveira, N. L., & Yurko, R. (2020). TRAP: A predictive framework for the assessment of performance in trail running. *Journal of Quantitative Analysis in Sports*, *ahead-of-print*. https://doi.org/10.1515/jqas-2020-0013

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., & Yang, J. (2023). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models* (4.1-7) [Computer software]. https://CRAN.R-project.org/package=glmnet

Germain, J. L., & Hausenblas, H. A. (2006). The relationship between perceived and actual physical fitness: A meta-analysis. *Journal of Applied Sport Psychology*, *18*(4), 283–296. https://doi.org/10.1080/10413200600944066

Geurkink, Y., Vandewiele, G., Lievens, M., de Turck, F., Ongenae, F., Matthys, S. P. J., Boone, J., & Bourgois, J. G. (2019). Modeling the prediction of the session rating of perceived exertion in soccer: Unraveling the puzzle of predictive indicators. *International Journal of Sports Physiology & Performance*, *14*(6), 841–846.

Glass, S. C., & Chvala, A. M. (2001). Preferred Exertion Across Three Common Modes of Exercise Training. *Journal of Strength and Conditioning Research*, *15*(4), 474–479.

Grant, S., Aitchison, T., Henderson, E., Christie, J., Zare, S., McMurray, J., & Dargie, H. (1999). A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects During Submaximal Exercise. *Chest*, *116*(5), 1208–1217. https://doi.org/10.1378/chest.116.5.1208

Greie, S., Humpeler, E., Gunga, H.-C., Koralewski, E., Klingler, A., Mittermayr, M., Fries, D., Lechleitner, M., Hoertnagl, H., Hoffmann, G., Blasche, G., & Schobersberger, W. (2006). Improvement of metabolic syndrome markers through altitude specific hiking vacations. *Journal of Endocrinological Investigation*, *29*, 497–504. https://doi.org/10.1007/BF03344138

Hajifar, S., Sun, H., Megahed, F. M., Jones-Farmer, L. A., Rashedi, E., & Cavuoto, L. A. (2021). A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Applied Ergonomics*, *90*, 103262. https://doi.org/10.1016/j.apergo.2020.103262

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hastie, T., Friedman, J., & Tisbshirani, R. (2018). *The Elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Stanford University.

Heggie, T. W., & Amundson, M. E. (2009). Dead men walking: Search and rescue in US national parks. *Wilderness & Environmental Medicine*, *20*(3), 244–249. https://doi.org/10.1580/08-WEME-OR-299R.1

Heggie, T. W., & Heggie, T. M. (2012). Dead Men Hiking: Case Studies from the American Wilderness. *Medicina Sportiva*, *16*(3), 118–121.

Hendey, G. W., Donner, N. F., & Fuller, K. (2005). Clinically significant changes in nausea as measured on a visual analog scale. *Annals of Emergency Medicine*, *45*(1), 77–81. https://doi.org/10.1016/j.annemergmed.2004.07.446

Hiking Project. (n.d.). *FAQ: Overview of Hiking Project Features—Difficulty Ratings*. Hiking Project. Retrieved March 25, 2021, from https://www.hikingproject.com/help/21/overview-of-site-name-features

Hugo, M. L. (1999a). A comprehensive approach towards the planning, grading and auditing of hiking trails as ecotourism products. *Current Issues in Tourism*, *2*(2–3), 138–173. https://doi.org/10.1080/13683509908667849

Hugo, M. L. (1999b). Energy equivalent as a measure of the difficulty rating of hiking trails. *Tourism Geographies*, *1*(3), 358–373. https://doi.org/10.1080/14616689908721327

Hugo, M. L., Kruger, P. E., Van Vuuren, B., & Hugo, M. L. (1998). Difficulty grading of hiking trails. *S.A. Journal for Research in Sport, Physical Education and Recreation*, *21*(1), 47–54.

Hulens, M., Vansant, G., Claessens, A. L., Lysens, R., & Muls, E. (2003). Predictors of 6-minute walk test results in lean, obese and morbidly obese women. *Scandinavian Journal of Medicine & Science in Sports*, *13*(2), 98–105. https://doi.org/10.1034/j.1600-0838.2003.10273.x

International Hiking Trails. (n.d.). Green Flag Trails. Retrieved March 25, 2021, from https://greenflagtrails.co.za/hiking-trails-international/

Jacobson, B. H., Wright, T., & Dugan, B. (2000). Load Carriage Energy Expenditure With and Without Hiking Poles During Inclined Walking. *International Journal of Sports Medicine*, *21*(5), 356–359. https://doi.org/10.1055/s-2000-3775

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning: With applications in R*.

Jensen, K., Rosthøj, S., Linneberg, A., & Aadahl, M. (2018). The association between self-rated fitness and cardiorespiratory fitness in adults. *International Journal of Sports Medicine*, *39*(06), 419–425. https://doi.org/10.1055/s-0044-102128

Johnson, R. M., Huettl, B., Kocsis, V., Chan, S. B., & Kordick, M. F. (2007). Injuries sustained at yellowstone national park requiring emergency medical system activation. *Wilderness & Environmental Medicine*, *18*(3), 186–189. https://doi.org/10.1580/06-WEME-OR-046R1.1

Kaplan, R. (2001). The nature of the view from home: Psychological benefits. *Environment and Behavior*, *33*(4), 507–542. https://doi.org/10.1177/00139160121973115

Kelly, A.-M. (2001). The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. *Emergency Medicine Journal*, *18*(3), 205–207. https://doi.org/10.1136/emj.18.3.205

Kim, H., Lee, S., Uysal, M., Kim, J., & Ahn, K. (2015). Nature-Based Tourism: Motivation and Subjective Well-Being. *Journal of Travel & Tourism Marketing*, *32*(sup1), S76–S96. https://doi.org/10.1080/10548408.2014.997958

Knight, C. A., & Caldwell, G. E. (2000). Muscular and metabolic costs of uphill backpacking: Are hiking poles beneficial? *Medicine & Science in Sports & Exercise*, *32*(12), 2093–2101.

Kondo, M. C., Oyekanmi, K. O., Gibson, A., South, E. C., Bocarro, J., & Hipp, J. A. (2020). Nature Prescriptions for Health: A Review of Evidence and Research Opportunities. *International Journal of Environmental Research and Public Health*, *17*(12), 4213. https://doi.org/10.3390/ijerph17124213

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2023). *caret: Classification and Regression Training* (6.0-94) [Computer software]. https://CRAN.R-project.org/package=caret

Langer, S., Müller, R., Schmid, K., & Linnhoff-Popien, C. (2020). Difficulty classification of mountainbike downhill trails utilizing deep neural networks. *Communications in Computer and Information Science*, 270–280. https://doi.org/10.1007/978-3-030-43887-6_21

Lentini, M., Scalia, J., Lebel, F. B., Touma, F., Jhajj, A., Darlington, P. J., & Dover, G. (2021). Association Between Pain Catastrophizing and Pain and Cardiovascular Changes During a Cold-Pressor Test in Athletes. *Journal of Athletic Training (Allen Press)*, *56*(5), 473–483. https://doi.org/10.4085/1062-6050-016-20

Levine, A., & Buono, M. J. (2019). Rating of perceived exertion increases synergistically during prolonged exercise in a combined heat and hypoxic environment. *Journal of Thermal Biology*, *84*, 99–102. https://doi.org/10.1016/j.jtherbio.2019.06.006

Lum, C. S., Keith, S. J., & Scott, D. (2019). The long-distance hiking social world along the Pacific Crest Trail. *Journal of Leisure Research*, *51*(2), 165–182. https://doi.org/10.1080/00222216.2019.1640095

Marinov, B., Kostianev, S., & Turnovska, T. (2002). Ventilatory efficiency and rate of perceived exertion in obese and non-obese children performing standardized exercise. *Clinical Physiology and Functional Imaging*, *22*(4), 254–260. https://doi.org/10.1046/j.1475-097X.2002.00427.x

Mason, R. C., Suner, S., & Williams, K. A. (2013). An Analysis of Hiker Preparedness: A Survey of Hiker Habits in New Hampshire. *Wilderness & Environmental Medicine*, *24*(3), 221–227. https://doi.org/10.1016/j.wem.2013.02.002

Maw, G. J., Boutcher, S. H., & Taylor, N. A. S. (1993). Ratings of perceived exertion and affect in hot and cool environments. *European Journal of Applied Physiology and Occupational Physiology*, *67*(2), 174–179. https://doi.org/10.1007/BF00376663

McIntosh, F., & Hugo, L. (n.d.). Difficulty Rating – Green Flag Trails. *Unique Difficulty Rating System*. Retrieved March 21, 2021, from https://greenflagtrails.org/difficulty-rating/

Mitten, D., Overholt, J. R., Haynes, F. I., D'Amore, C. C., & Ady, J. C. (2018). Hiking: A low-cost, accessible intervention to promote health benefits. *American Journal of Lifestyle Medicine*, *12*(4), 302–310. https://doi.org/10.1177/1559827616658229

National Park Service. (2021, February 16). *National Park System Units*. National Park Service. https://www.nps.gov/aboutus/national-park-system.htm

Obling, K. H., Hansen, A.-L. S., Overgaard, K., Normann, K., Sandbaek, A., & Maindal, H. T. (2015). Association between self-reported and objectively measured physical fitness level in a middle-aged population in primary care. *Preventive Medicine Reports*, *2*, 462–466. https://doi.org/10.1016/j.pmedr.2015.05.010

Op De Beéck, T., Meert, W., Schütte, K., Vanwanseele, B., & Davis, J. (2018). Fatigue prediction in outdoor runners via machine learning and sensor fusion. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 606–615. https://doi.org/10.1145/3219819.3219864

Outdoor Industry Association, & The Outdoor Foundation. (2024). *2024 outdoor participation trends report: Executive summary*. https://oia.outdoorindustry.org/research/2024-outdoor-participation-trends-report-executive-summary

Parfitt, G., Rose, E. A., & Burgess, W. M. (2006). The psychological and physiological responses of sedentary individuals to prescribed and preferred intensity exercise. *British Journal of Health Psychology*, *11*(1), 39–53. https://doi.org/10.1348/135910705X43606

Parfitt, G., Rose, E. A., & Markland, D. (2000). The Effect of Prescribed and Preferred Intensity Exercise on Psychological Affect and the Influence of Baseline Measures of Affect. *Journal of Health Psychology*, *5*(2), 231–240. https://doi.org/10.1177/135910530000500213

Perrey, S., & Fabre, N. (2008). Exertion During Uphill, Level and Downhill Walking With and Without Hiking Poles. *Journal of Sports Science & Medicine*, *7*(1), 32–38.

Pitman, A., Zanker, M., Gamper, J., & Andritsos, P. (2012). Individualized hiking time estimation. *2012 23rd International Workshop on Database and Expert Systems Applications*, 101–105. https://doi.org/10.1109/DEXA.2012.51

Python Core Team. (2023). *Python: A dynamic, open source programming language.* [Computer software]. Python Software Foundation. https://www.python.org/

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

REI. (n.d.). *#OptOutside on Black Friday and every day | REI Co-op*. REI. Retrieved February 17, 2024, from https://www.rei.com/opt-outside

Robertson, R. J., Goss, F. L., & Metz, K. F. (1998). Perception of Physical Exertion during Dynamic Exercise: A Tribute to Professor Gunnar A. V. Borg. *Perceptual and Motor Skills*, *86*(1), 183–191. https://doi.org/10.2466/pms.1998.86.1.183

Rodgers, C. D., Vanheest, J. L., & Schachter, C. L. (1995). Energy expenditure during submaximal walking with Exerstriders. *Medicine & Science in Sports & Exercise*, *27*(4), 607–611.

Rose, E., & Parfitt, G. (2007). A Quantitative Analysis and Qualitative Explanation of the Individual Differences in Affective Responses to Prescribed and Self-Selected Exercise Intensities. *Journal of Sport & Exercise Psychology*, *29*, 281–309. https://doi.org/10.1123/jsep.29.3.281

Rose, E., & Parfitt, G. (2012). Exercise experience influences affective and motivational outcomes of prescribed and self-selected intensity exercise. *Scandinavian Journal of Medicine & Science in Sports*, *22*(2), 265–277. https://doi.org/10.1111/j.1600-0838.2010.01161.x

Sándor, A. M. (2018). Personalised hiking time estimation. *Pure Mathematics and Applications*, *27*(2), 35–42. https://doi.org/10.1515/puma-2015-0017

Saunders, M. J., Hipp, G. R., Wenos, D. L., & Deaton, M. L. (2008). Trekking Poles Increase Physiological Responses to Hiking Without Increased Perceived Exertion. *The Journal of Strength & Conditioning Research*, *22*(5), 1468–1474. https://doi.org/10.1519/JSC.0b013e31817bd4e8

Schobersberger, W., Leichtfried, V., Mueck-Weymann, M., & Humpeler, E. (2010). Austrian Moderate Altitude Studies (AMAS): Benefits of exposure to moderate altitudes (1,500–2,500 m). *Sleep and Breathing*, *14*(3), 201–207. https://doi.org/10.1007/s11325-009-0286-y

Schumacher, M., & Rey, J.-P. (2011). Recommender systems for dynamic packaging of tourism services. In R. Law, M. Fuchs, & F. Ricci (Eds.), *Information and Communication Technologies in Tourism 2011* (p. 11). Springer, Vienna.

Shenandoah National Park Virginia. (2017, December 5). *How to Determine Hiking Difficulty*. National Park Service. https://www.nps.gov/shen/planyourvisit/how-to-determine-hiking-difficulty.htm

Shook, R., Gribben, N., Hand, G., Paluch, A., Welk, G., Jakicic, J., Hutto, B., Burgess, S., & Blair, S. (2016). Subjective estimation of physical activity using the IPAQ varies by fitness level. *Journal of Physical Activity & Health*, *13*, 79–86. https://doi.org/10.1123/jpah.2014-0543

Singer, A. J., & Thode, H. C. (1998). Determination of the Minimal Clinically Significant Difference on a Patient Visual Analog Satisfaction Scale. *Academic Emergency Medicine*, *5*(10), 1007–1011. https://doi.org/10.1111/j.1553-2712.1998.tb02781.x

Spurway, N. C. (1992). Aerobic exercise, anaerobic exercise and the lactate threshold. *British Medical Bulletin*, *48*(3), 569–591. https://doi.org/10.1093/oxfordjournals.bmb.a072564

Stephens, B. D., Diekema, D. S., & Klein, E. J. (2005). Recreational injuries in Washington state national parks. *Wilderness & Environmental Medicine*, *16*(4), 192–197. https://doi.org/10.1580/1080-6032(2005)16[192:RIIWSN]2.0.CO;2

Sullivan, M. J. L., Tripp, D. A., Rodgers, W. M., & Stanish, W. (2000). Catastrophizing and pain perception in sport participants. *Journal of Applied Sport Psychology*, *12*(2), 151–167. https://doi.org/10.1080/10413200008404220

Trayers, F. J. (2004). Wilderness Preventive Medicine. *Wilderness & Environmental Medicine*, *15*(1), 1–3. https://doi.org/10.1580/1080-6032(2004)015[0001:WPM]2.0.CO;2

Trosclair, D., Bellar, D., Judge, L. W., Smith, J., Mazerat, N., & Brignac, A. (2011). Hand-Grip Strength as a Predictor of Muscular Strength and Endurance. *The Journal of Strength & Conditioning Research*, *25*, S99. https://doi.org/10.1097/01.JSC.0000395736.42557.bc

Twombly, S. E., & Schussman, L. C. (1995). Gender differences in injury and illness rates on wilderness backpacking trips. *Wilderness & Environmental Medicine*, *6*(4), 363–376. https://doi.org/10.1580/1080-6032(1995)006[0363:GDIIAI]2.3.CO;2

Ulrich, R. (1984). View through a window may influence recovery from surgery. *Science*, *224*(4647), 420–421. https://doi.org/10.1126/science.6143402

Ulrich, R. S. (1979). Visual landscapes and psychological well-being. *Landscape Research*, *4*(1), 17–23. https://doi.org/10.1080/01426397908705892

Vandewiele, G., Geurkink, Y., Lievens, M., Ongenae, F., Turck, F. D., & Boone, J. (2017). *Enabling training personalization by predicting the session rate of perceived exertion (sRPE)*. 12.

von Heimburg, E., Sandsund, M., Rangul, T. P., & Reinertsen, R. E. (2019). Physiological and perceptual strain of firefighters during graded exercise to exhaustion at 40 and 10 °C. *International Journal of Occupational Safety and Ergonomics*, *25*(3), 412–422. https://doi.org/10.1080/10803548.2017.1381468

Whiting, Larson, Green, & Kralowec. (2017). Outdoor recreation motivation and site preferences across diverse racial/ethnic groups_ A case study of Georgia state parks | Elsevier Enhanced Reader. *Journal of Outdoor Recreation and Tourism*, *18*, 10–21. http://dx.doi.org/10.1016/j.jort.2017.02.001

Wilcer, S., Larson, L., Hallo, J., & Baldwin, E. (2019). Exploring the Diverse Motivations of Day Hikers: Implications for Hike Marketing and Management. *The Journal of Park and Recreation Administration*. https://doi.org/10.18666/JPRA-2019-9176

Willamette National Forest. (n.d.). *Hiking—Trail Difficulty Ratings*. United States Department of Agriculture - Forest Service. Retrieved March 25, 2021, from https://www.fs.usda.gov/detailfull/willamette/recreation/hiking/?cid=stelprdb5103006&width=full

Zamparo, P., Perini, R., Peano, C., & Di Prampero, P. E. (2001). The Self Selected Speed of Running in Recreational Long Distance Runners. *International Journal of Sports Medicine*, *22*(8), 598–604. https://doi.org/10.1055/s-2001-18559

Zisapel, N., & Nir, T. (2003). Determination of the minimal clinically significant difference on a patient visual analog sleep quality scale. *Journal of Sleep Research*, *12*(4), 291–298. https://doi.org/10.1046/j.0962-1105.2003.00365.x

APPENDICES

Appendix A. Description of Statistical Models

Support Vector Machines

The support vector machine (SVM) is an extension of the support vector classifier. Support vector classifiers address the problem of non-linearity by enlarging the feature space through the addition of polynomial versions of predictor variables (James et al., 2017). The SVM also does feature space expansion but uses kernels to make computations more efficient (James et al., 2017). Simply, a kernel is a mathematical function representing the relationship between two outcomes; this quantity is then used to create a decision rule for classification problems (James et al., 2017).

Appendix B. Borg RPE 6-20 Scale

6	No exertion at all
7	Extremely light
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard (heavy)
16	
17	Very hard
18	
19	Extremely hard
20	Maximal Exertion

(Borg, 1970, 1998)

Appendix C. Rating of Perceived Exertion Visual Analog Scale

Maximal Exertion

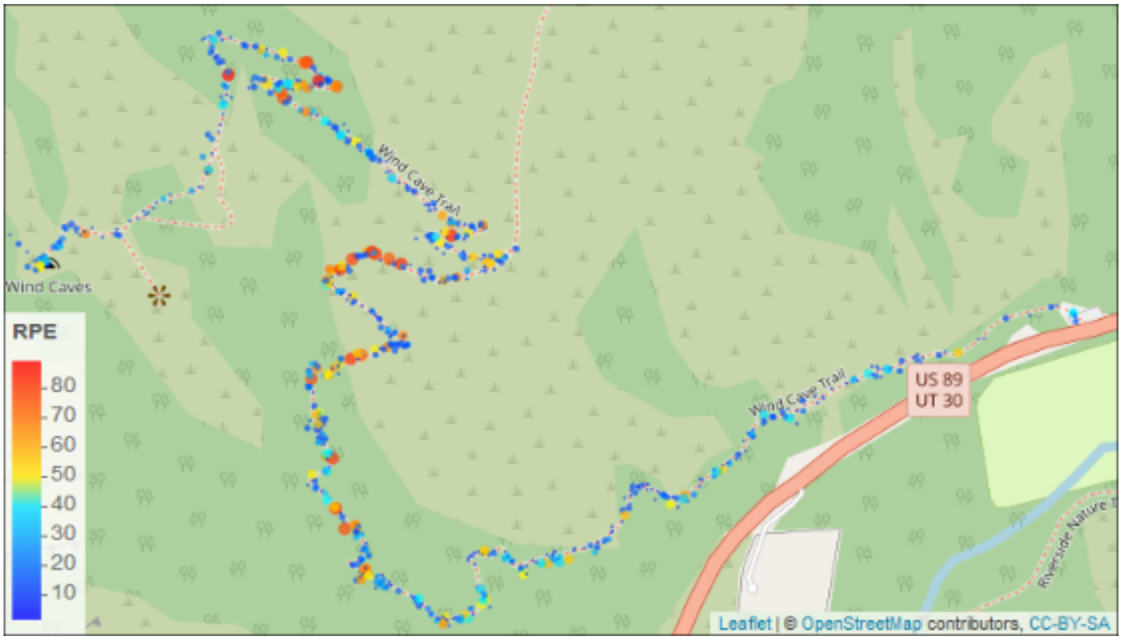
No Exertion At All

Rate your current perceived exertion.
(If you are taking a break, rate your perceived exertion right before you stopped hiking.)

Record the time from the fitness tracker
Time: ______________________

Turn to the Next Page

Appendix D. Wind Caves Hiking Trail Information



All RPE ratings provided by participants in the feasibility study plotted on the Wind Caves trail based on latitude and longitude at the time RPE was documented. Each circle represents a single RPE provided at that trail location. Circle color and size represent the value of the RPE rating. Larger circles indicate a higher RPE. The color to RPE value mapping can be viewed in the inset on the map.

Appendix E. Physical Activity Readiness Questionnaire (PAR-Q)

Regular physical activity is fun and healthy, and increasingly more people are starting to become more active every day. Being more active is very safe for most people. However, some people should check with their doctor before they start becoming much more physically active. If you are planning to become much more physically active than you are now, start by answering the seven questions in the box below. If you are between the ages of 15 and 69, the PAR-Q will tell you if you should check with your doctor before you start. If you are over 69 years of age, and you are not used to being very active, check with your doctor. Common sense is your best guide when you answer these questions. Please read the questions carefully and answer each one honestly: check YES or NO.

1. Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?
2. Do you feel pain in your chest when you do physical activity?
3. In the past month, have you had chest pain when you were not doing physical activity?
4. Do you lose your balance because of dizziness or do you ever lose consciousness?
5. Do you have a bone or joint problem (for example, back, knee or hip) that could be made worse by a change in your physical activity?
6. Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?
7. Do you know of any other reason why you should not do physical activity?

Appendix F. International Physical Activity Questionnaire (15-69 years)

We are interested in finding out about the kinds of physical activities that people do as part of their everyday lives. The questions will ask you about the time you spent being physically active in the **last 7 days**. Please answer each question even if you do not consider yourself to be an active person. Please think about the activities you do at work, as part of your house and yard work, to get from place to place, and in your spare time for recreation, exercise or sport.

Think about all the **vigorous** activities that you did in the **last 7 days**. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Think *only* about those physical activities that you did for at least 10 minutes at a time.

1. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?*

_____ **days per week**

No vigorous physical activities ***Skip to question 3***

2. How much time did you usually spend doing **vigorous** physical activities on one of those days?

_____ **hours per day**

_____ **minutes per day**

Don’t know/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

3. During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

_____ **days per week**

No moderate physical activities ***Skip to question 5***

4. How much time did you usually spend doing **moderate** physical activities on one of those days?

_____ **hours per day**

_____ **minutes per day**

Don't know/Not sure

Think about the time you spent **walking** in the **last 7 days**. This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.

5. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time?

_____ **days per week**

No walking ***Skip to question 7***

6. How much time did you usually spend **walking** on one of those days?*

_____ **hours per day**

_____ **minutes per day**

Don't know/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

7. During the **last 7 days**, how much time did you spend **sitting** on a **week day**?*

_____ **hours per day**

_____ **minutes per day**

Don't know/Not sure

* Denotes items selected for use in Chapters 3 & 4.

Appendix G. Baecke Fitness Inventory

Work Index

1. What is your main occupation?

low activity, moderate activity, high activity

((1) low activity including clerical work, driving, shop keeping, teaching, studying, housework, medical practice, and occupations requiring a university education; (2) middle activity including factory work, plumbing, carpentry, and farming; (3) high activity includes dock work, construction work, and professional sport.)

2. At work I sit:

never, seldom, sometimes, often, always

3. At work I stand:

never, seldom, sometimes, often, always

4. At work I walk:

never, seldom, sometimes, often, always

5. At work I lift heavy loads:

never, seldom, sometimes, often, always

6. After working I am tired:

very often, often, sometimes, seldom, never

7. At work I sweat:

very often, often, sometimes, seldom, never

8. In comparison of other of my own age I think my work is physically:

much heavier, heavier, as heavy, lighter, much lighter

Sport Index

9. Do you play sports?

Yes, No

10. In comparison with others of my own age I think my physical activity during leisure time is:

much more, more, the same, less, much less

11. During leisure time I sweat:

very often, often, sometimes, seldom, never

12. During leisure time I play sport:

never, seldom, sometimes, often, very often

13. What sport do you play most frequently?

low intensity, medium intensity, high intensity

((1) low level examples: billiards, sailing, bowling, golf; (2) middle level examples: badminton, cycling, dancing, swimming, tennis; (3) high level examples: boxing, basketball, football, rugby, rowing)

14. How many hours do you play a week? (For your most frequently played sport)

< 1 hour, 1-2 hours, 2-3 hours, 3-4 hours, > 4 hours

15. How many months do you play in a year? (For your most frequently played sport)

< 1 month, 1-3 months, 4-6 months, 7-9 months, > 9 months

16. What sport do you play second most frequently?

low intensity, medium intensity, high intensity

((1) low level examples: billiards, sailing, bowling, golf; (2) middle level examples: badminton, cycling, dancing, swimming, tennis; (3) high level examples: boxing, basketball, football, rugby, rowing)

17. How many hours do you play a week? (For your second most frequently played sport)

< 1 hour, 1-2 hours, 2-3 hours, 3-4 hours, > 4 hours

18. How many months do you play in a year? (For your second most frequently played sport)

< 1 month, 1-3 months, 4-6 months, 7-9 months, > 9 months

Leisure Index

18. During leisure time I watch television:

never, seldom, sometimes, often, very often

19. During leisure time I walk:

never, seldom, sometimes, often, very often

20. During leisure time I cycle:

never, seldom, sometimes, often, very often

21. How many minutes do you walk and/or cycle per day to and from work, school, and shopping?

< 5 minutes, 5-15 minutes, 15-30 minutes, 30-45 minutes, > 45 minutes

* All BFI items were included in Chapters 3 and 4.

Appendix H. Pain Catastrophizing Scale

Everyone experiences painful situations at some point in their lives. Such experiences may include headaches, tooth pain, joint or muscle pain. People are often exposed to situations that may cause pain such as illness, injury, dental procedures or surgery.

We are interested in the types of thoughts and feeling that you have when you are in pain. Listed below are thirteen statements describing different thoughts and feelings that may be associated with pain. Using the following scale, please indicate the degree to which you have these thoughts and feelings when you are experiencing pain.

0 - not at all

1 - to a slight degree

2 - to a moderate degree

3 - to a great degree

4 - all the time

1) When I'm in pain, I worry all the time about whether the pain will end.

2) When I'm in pain, I feel I can't go on.

3) When I'm in pain, it's terrible and I think it's never going to get any better.

4) When I'm in pain, it's awful and I feel that it overwhelms me.

5) When I'm in pain, I feel I can't stand it anymore.

6) When I'm in pain, I become afraid that the pain will get worse.

7) When I'm in pain, I keep thinking of other painful events.*

8) When I'm in pain, I anxiously want the pain to go away.

9) When I'm in pain, I can't seem to keep it out of my mind.

10) When I'm in pain, I keep thinking about how much it hurts.*

11) When I'm in pain, I keep thinking about how badly I want the pain to stop.

12) When I'm in pain, there's nothing I can do to reduce the intensity of the pain.

13) When I'm in pain, I wonder whether something serious may happen. *

*Denotes PCS items included in Chapters 3 & 4.

Appendix I. Demographic Questionnaire

Age: _________

Sex Assigned at Birth: _______________

Gender Identity: _______________

Race/Ethnicity: _______________

Height: _________ ft _________ in

Weight: _________ lbs

Have you had an injury to any of the following body regions in the last 12 months?

Foot	**No**	**Yes**
Ankle	**No**	**Yes**
Knee	**No**	**Yes**
Quadriceps / Hamstring	**No**	**Yes**
Hip	**No**	**Yes**
Neck	**No**	**Yes**
Back	**No**	**Yes**
Shoulder	**No**	**Yes**
Head	**No**	**Yes**

If you answered yes to sustaining a lower body injury in the last 12 months, please briefly describe the injury, any current impairment due to the injury, and provide a general date of injury:

____

Appendix J. Pre-Hike Survey

Have you hiked this trail previously? **No** **Yes**

Have you recently travelled from an area of lower altitude? **No** **Yes**

How long have you been at this altitude? _______________

Are you hiking alone today? **No** **Yes**

If **no**, number of others hiking with you (do not include yourself in this total):

____

Will you be using trekking poles today during the hike? **No** **Yes**

Will you be carrying a pack during the hike? **No** **Yes**

If **yes**, weight of pack: _______________

Approximately how many hikes have you completed in the last month?: _______________

How difficult do you expect this hike to be (select one)?

Very Easy **Easy** **Fair** **Moderate** **Difficult** **Severe** **Extreme**

Appendix K. Traditional Hiking Difficulty Rating Scale

How difficult do you expect this hike to be?

Very Easy

Easy

Fair

Moderate

Difficult

Severe

Extreme

CURRICULUM VITAE

Kelci B. Hannan
Department of Kinesiology and Health Science (573) 846-7496
Utah State University kelci.hannan@usu.edu
Logan, UT 84322

EDUCATION

Utah State University, Logan, Utah Anticipated May 2024
- Ph.D. in Disability Disciplines – Pathokinesiology
- Advisor: Christopher Dakin

Texas Christian University, Fort Worth, Texas 2018
- M.S. in Kinesiology – Motor Control Specialization
- Advisor: Adam King

Truman State University, Kirksville, Missouri 2015
- B.S. in Athletic Training

PUBLICATIONS

Peer-Reviewed Journal Articles

Hannan, K.B., & King, A.C. (2022). Lower limb ground reaction force and center of pressure asymmetry during bodyweight squats. *Int J Sports Phys Ther.,* 17(6), 1075-1082. https://doi.org/10.26603/001c.37861.

Hannan, K.B., Todd, M.K., Pearson, N.J., Forbes, P.A. & Dakin, C.J. (2021). Absence of nonlinear coupling between electric vestibular stimulation and evoked forces during standing balance. Front. Hum. Neurosci. 15:631782. doi: 10.3389/fnhum.2021.631782

Hannan, K.B., Todd, M.K., Pearson, N.J., Forbes, P.A. & Dakin, C.J. (2021). Vestibular attenuation to random-waveform galvanic vestibular stimulation during standing and treadmill walking. *Sci Rep* 11, 8127. https://doi.org/10.1038/s41598-021-87485-4

King, A.C., & **Hannan, K.B.** (2019). Segment coordination variability during double leg bodyweight squats at different tempos. Int J Sports Med, 40(11), 725–731. https://doi.org/10.1055/a-0965-7358.

In Preparation

Hannan, K.B., Rice, N., & Dakin, C.J. Subthreshold noisy galvanic vestibular stimulation as a postural control intervention: A systematic review.

Rice, D., **Hannan, K.B.**, & Dakin, C.J. No evidence of stochastic resonance in postural sway response to noisy galvanic vestibular stimulation in healthy young adults.

GRANT WRITING

Funded

Graduate Research and Creative Opportunities, Utah State University 2019
Award: $1000

Grant-In-Aid of Scholarship and Research, Truman State University 2014
Award: $750

Not Funded

PhysioQ and Labfront Summer Research Grant 2022
Amount Requested: $2695

PRESENTATIONS

Poster Presentations

Hannan, K.B., Todd, M.K., Pearson, N.J., Forbes, P.A., Schwartz, S.E. & Dakin, C.J. (2021). Secondary Sensations Associated with Random-Waveform Electrical Vestibular Stimulation. Student Research Symposium, Utah State University.

Hannan, K.B., Todd, M.K., Pearson, N.J., Forbes, P.A., Dakin, C.J. (2019). Society for Neuroscience, Chicago, IL. Limited Habituation to Prolonged Electrical Vestibular Stimulation During Treadmill Walking, Society for Neuroscience (SFN), Chicago, IL.

Todd, M.K., **Hannan, K.B.**, Pearson, N.J., Dakin, C.J. (2019). Habituation and Secondary Sensations that Arise from Stimulation of the Vestibular System. Student Research Symposium, Utah State University.

Turner, S., Lindley, H., **Hannan, K.B.**, Goto, S., Bothwell, J.M., Garrison, J.C., Hannon, J.P. & King, A.C. (2019). Kinematic Variability of Female ACL Reconstruction and Healthy Athletes During the Drop Landing Task. Journal of Sport & Exercise Psychology.

Hannan, K.B., Stone, J., Arndts, D., Anzalone, A., Oliver, J.M., Bothwell, J.M., King, A.C. (2018). Segment Coordination Variability Changes During Back Squats with Biofeedback. American Society of Biomechanics (ASB), Rochester, MN.

Hannan, K.B., Mata, J.D., Oliver, J.M., Bothwell, J.M., King, A.C. (2018). Lower Extremity Coordination Patterns Between Traditional and Cluster Training During Back Squat. North American Society for the Psychology of Sport and Physical Activity (NASPSPA), Denver, CO.

Besand, K.B., Bothwell, J.M., Garrison, J.C., Hannon, J.P., Goto, S., Grondin, A. & King, A.C. (2017). Lower Extremity Segment Coordination Following Anterior Cruciate Ligament Reconstruction. American Society of Biomechanics (ASB), Boulder, CO.

Besand, K.B., Bothwell, J.M., Garrison, J.C., Hannon, J.P., Goto, S. & King, A.C. (2017). Effect of Different Sports on Hip and Knee Biomechanics in Adolescent Females During a Jump-Landing. American College of Sports Medicine (ACSM), Denver, CO.

Goto, S., Hannon, J.P., Grondin, A.N., **Besand, K.B.**, Abowd, M.E., Bothwell, J.M., & Garrison, J.C. (2017). Bilateral Lower Extremity Energy Absorption Patterns and Muscle Strength in Adolescent Males and Females During Jump-Landing at Return to Sport Following Anterior Cruciate Ligament Reconstruction. Journal of Athletic Training, Dallas, TX.

Backes, M., **Besand, K**., Stephenson, A. & Walker, D. (2012). Correlation Between Numerical Activity Factor and Bone Mineral Density in College-Age Females. Presented at Missouri Association for Health, Physical Education, Recreation, and Dance (MOAHPERD), Lake Ozark, MO.

TEACHING

Huntsman Scholars Lab, Utah State University, Guest Lecturer – Tech in Sports	2023
Introductory Biomechanics, Truman State University, Primary Instructor	2022-Present
Intro to Modern Data Analytics, Utah State University, Guest Lecturer	2022
Topics in Biomechanics, Utah State University, Guest Lecturer – Injury Biomechanics	2020-2022
Topics in Biomechanics, Utah State University, Primary Instructor	2020
Topics in Biomechanics, Utah State University, Teaching Assistant	2018-Present
Human Anatomy Cadaver Lab, Truman State University, Teaching Assistant	2012-2015
Concepts of Biomechanics Lab, Truman State University, Teaching Assistant	2013-2014

DISTINCTIONS/HONORS

Doctoral Student Researcher of the Year Nominee, Utah State University 2023
Presidential Doctoral Research Fellow, Utah State University 2018-2022
Outstanding Graduate Student, Kinesiology and Health Science Department 2022
Outstanding Thesis Nominee, Texas Christian University 2018
Clint Thompson Excellence in Athletic Training Award, Truman State University 2015
Outstanding Undergraduate Student in Athletic Training, Truman State University 2015
National Athletic Trainers' Association Foundation Scholarship – Graduating Undergraduate 2015
Harry S. Truman Leadership Scholar, Truman State University 2011-2015

CERTIFICATIONS

Teaching Online Certification – Quality Matters December 2022
Board Certified Athletic Trainer (ATC) Since June 2015
American Heart Association, BLS for the Healthcare Provider Expires January 2024

EMPLOYMENT HISTORY

Director of Research & Analytics, St. Louis Area Business Health Coalition 2023-Present

- Oversee reporting of healthcare claims analytics to regional medical group leaders

Lecturer, Department of Health and Exercise Sciences, Truman State University 2022-2023

- Taught asynchronous, online section of Introductory Biomechanics

Neuromechanics Laboratory Research Assistant, Utah State University 2018-2023

- Conducted research related to vestibular involvement in postural control

Ben Hogan Sports Medicine Graduate Assistant, Texas Health Resources 2016-2018

- Conducted data collection for various investigations related to lower extremity injury and rehabilitation
- Generated ACL injury risk reports used to inform patients of their relative risk of ACL injury

Biomechanics Tutor 2016

- Worked one-on-one with students to improve their understanding and application of biomechanics

Dee Johnson, Personal Training LLC Seasonally, 2014-2016

- Designed exercise programs for clients 13 to 65 years of age
- Implemented and oversaw cardiovascular and resistance training protocols for clientele

Graduate Assistant Athletic Trainer, University of Central Arkansas, Conway, AR 2015

- Oversaw the medical coverage of NCAA Division I Cross Country/Track and Field
- Served as clinical instructor to assigned athletic training students
- Developed and wrote university mononucleosis policy

PROFESSIONAL SERVICE

Hackathon Judge

HackUSU, Health & Fitness Category, Utah State University 2022, 2023

Reviewer

Undergraduate Research Fellowship, Utah State University 2022
Graduate Enhancement Award, Utah State University 2019
Undergraduate Research and Creative Opportunities (URCO) Grant Program 2018, 2019

Editorial Board

Curiosity: Utah State University's Undergraduate Research Journal 2020-2021

Research Presentation Judge

Student Research Symposium, Utah State University 2019, 2020