

Electronic Theses and Dissertations, 2004-2019

2015

Body Mass Bias Mitigation for Females in Military Physical Readiness Testing through Load Carriage Implementation

Aaron Yeaton
University of Central Florida

 Part of the [Industrial Engineering Commons](#)
Find similar works at: <https://stars.library.ucf.edu/etd>
University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Yeaton, Aaron, "Body Mass Bias Mitigation for Females in Military Physical Readiness Testing through Load Carriage Implementation" (2015). *Electronic Theses and Dissertations, 2004-2019*. 1416.
<https://stars.library.ucf.edu/etd/1416>

BODY MASS BIAS MITIGATION FOR FEMALES IN MILITARY PHYSICAL READINESS
TESTING THROUGH LOAD CARRIAGE IMPLEMENTATION

by

AARON ALAN YEATON
M.S. University of Central Florida 2009
B.S. Thomas Edison State College 2007

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Industrial Engineering and Management Systems
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: Pamela McCauley

© 2015 Aaron A. Yeaton

ABSTRACT

The US Military requires specific fitness testing, known as the Physical Readiness Test (PRT), for its members to determine their overall fitness levels. The test currently being used has been shown to have bias towards heavier mass individuals of up to 20%. Prior research has been completed and several recommendations have been made to eliminate mass bias, but this has been conducted almost exclusively on males. There is very little data and research on military physical fitness testing for women besides combat specific evolutions exercises. A possibility exists to eliminate the bias for females through implementing load carriage during PRT events. A specified load is to be placed on women while performing the PRT and compared to a non-loaded control test. The results should show if the load carriage devised has a beneficial effect on current testing methods through eliminating the mass bias for women.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. McCauley for the continuous support of my Ph.D. study and related research. Her motivation, thorough knowledge, and guidance has helped me immensely and was invaluable in all phases of this and past projects. I could not have imagined a better advisor or mentor for my study.

I would also like to thank the rest of my thesis committee: Dr. Rabelo, Dr. Lee, and Dr. Cummings, for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

Last but not the least, I would like to thank my family; especially my wife Courtney for her unending support throughout this endeavor.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	v
LIST OF FIGURES	viii
LIST OF TABLES.....	x
CHAPTER ONE: INTRODUCTION.....	1
1.1 Introduction to Topic.....	1
1.2 Significance of Research.....	3
1.3 Theoretical Foundation of Work.....	4
1.4 Research Question.....	7
1.5 Contribution to the Body of Knowledge.....	8
CHAPTER TWO: LITERATURE REVIEW.....	9
2.1 History of the Military Physical Readiness Test.....	10
2.2 Current US Military Demographics	14
2.3 Penalties in PRT towards Body Mass, Little Occupational Relevance in Testing	22
2.4 Suggestions to Correct/Mitigate Bias in Testing and Increasing Occupational Relevance	28
2.5 Bias and Controversy in Specific Occupational Specialty Fitness Testing (Combat Testing)	
.....	38
2.6 Future Work and Ongoing Research in This Area.....	42
CHAPTER THREE: METHODOLOGY	44

3.1 Research Hypothesis	44
3.2 Proposed Subjects	45
3.3 Data Collection Environment.....	50
3.4 Data Collected	51
3.5 Experimental Design and Statistical Analysis Approach.....	54
CHAPTER FOUR: FINDINGS	57
4.1 Research Problem Statement Review	58
4.2 Research Participant Characteristics	58
4.2.1 Descriptive Statistics of Physical Characteristics.....	58
4.2.2 Participant Questionnaire Survey Data.....	64
4.2.3 Disqualification of Participants for the Physical Readiness Test	68
4.3 Participant’s Physical Readiness Performance Test Results.....	69
4.3.1 Descriptive Statistics of Test Results and Probability Distributions	69
4.3.2 Statistical Power and Margin of Error (Desired versus Actual).....	72
4.3.3 Power and Error (Push Ups).....	73
4.3.4 Statistical Power and Error (Curl Ups).....	74
4.3.5 Power and Error (Timed Run)	75
4.4 Statistical Differences between Treatments (Paired T-Test) Loaded vs Unloaded	76
4.4.1 Paired T- Test for Push-Ups	76

4.4.2 Paired T-Test for Curl-Ups.....	77
4.4.3 Paired T-Test for Timed Run.....	79
4.5 Analysis of Correlation between Treatments (Pearson’s Product Moment).....	81
4.5.1 Correlation Results for Push-Ups Unloaded versus Loaded	82
4.5.2 Correlation Results for Curl-Ups Unloaded versus Loaded.....	83
4.5.3 Correlation Results for Timed Run Unloaded versus Loaded.....	85
4.5.4 Research Question Hypothesis Test	87
CHAPTER FIVE: CONCLUSION.....	89
5.1 Contributions to Body of Knowledge	90
5.2 Future Research.....	91
APPENDIX A: RISK ASSESSMENT QUESTIONNAIRE.....	94
APPENDIX B: PARTICPANT AVAILABILITY QUESTIONNAIRE.....	96
APPENDIX C: PARTICPANT PRETEST PHYSICAL ASSESSMENT DATA SHEET	98
APPENDIX D: PARTICIPANT SELF FITNESS ASSESSMENT	101
APPENDIX E: PHYSICAL READINESS TEST DATA COLLECTION SHEETS	104
APPENDIX F: IRB OUTCOME APPROVAL LETTER.....	108
.....	109
LIST OF REFERENCES.....	111

LIST OF FIGURES

Figure 1: Gender of U.S Active Duty Members 2011 (DMDC, 2011)	2
Figure 2: Unloaded vs Loaded 2MR (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011)	6
Figure 3: Enlisted BMI Separations 1990-1998 (Hodgdon, 1999).....	20
Figure 4: Calculation of Theoretical Mass Bias (Vanderburgh, P. M., & Crowder, T. A., 2006)	28
Figure 5: Relationship between exercise tolerance time and body mass/lean body mass Bilzon, J. L. J., Allsopp, A. J., & Tipton, M. J. (2001).....	38
Figure 6: Power Curve for Sample Size Selection.....	47
Figure 7: PRT Health Risk Assessment Questionnaire	48
Figure 8: Tools used for Research & Data Collection.....	53
Figure 9: Proposed Methodology Flow	56
Figure 10: Ethnic Make Up of U.S. Active Duty Military (DMDC 2011).....	60
Figure 11: Chart of research Participant Make up	61
Figure 13: Active Duty Personnel Ages (DMDC 2011).....	62
Figure 14: Research Participant Age Group Overview	62
Figure 15: Probability Plot for Participant Weight.....	63
Figure 16: Current or Prior Affiliation of All Participants with the Military	64
Figure 17: Percent of Participants who have Ever Failed a Past PRT	65
Figure 18: Exercise Time (min) per Week for Research Participants	66
Figure 19: Method of Exercise for All Research Participants	66

Figure 20: Percentage of Research Participants who had Prior Experience with Weighted Vests	67
Figure 21: Probability Plot for Participant Performance	72
Figure 22: Power Curve for Desired Alpha Level	73
Figure 23: Required Sample Size for .8 Statistical Power in Push- Ups	74
Figure 24: Required Sample Size for .8 Statistical Power in Curl Ups	75
Figure 25: Required Sample Size for .8 Statistical Power for Timed Run	76
Figure 26: Histogram of Differences of Means for Push -Ups Unloaded vs Loaded.....	77
Figure 27: Histogram of Differences of Means for Curl-Ups Unloaded vs Loaded	79
Figure 28: Histogram of Differences of Means for Timed Run Unloaded vs Loaded	80
Figure 29: Scatterplot of Push-Ups vs Weight Correlation Unloaded and Loaded	82
Figure 30: Scatterplot of Push-Ups vs Weight Correlation Unloaded and Loaded (Overlaid Graph)	83
Figure 31: Scatterplot of Curl-Ups vs Weight Correlation Unloaded and Loaded.....	84
Figure 32: Scatterplot of Curl –Ups vs Weight Correlation Unloaded and Loaded (Overlaid Graph)	84
Figure 33: Scatterplot of timed-Run vs Weight Correlation Unloaded and Loaded	85
Figure 34: Scatterplot of timed Run vs Weight Correlation Unloaded and Loaded (Overlaid Graph)	86
Figure 35 : Scatterplot Overview of All Exercises Push-Ups, Curl-Ups, & Timed Run performance vs Weight.....	86

LIST OF TABLES

Table 1: Required exercise tests by branch.....	3
Table 2: Change in female population for Army & Navy 1960-2010 (US Census Bureau)	14
Table 3: Active Duty Military Personnel Gender per Branch (DMDC 2011).....	21
Table 4: Confidence Interval (percentage) of injury among female and male Army basic training recruits (Bell, Mangione, Hemenway, Amoroso, & Jones, 2000).....	21
Table 5: Equations generated to predict strength through triceps extension (Moschella, K. N. 2013)	37
Table 6: Descriptive Statistics for Participant Physical Characteristics	59
Table 7: Descriptive Statistics for Participant Performance Loaded and Unloaded.....	71
Table 8: Paired T-Test for Unloaded vs Loaded Treatments of Push-Ups.....	77
Table 9: Paired T-Test for Unloaded vs Loaded Treatments of Curl-Ups	78
Table 10: Paired T-Test Unloaded vs Loaded Treatments of Timed Run.....	80
Table 11: Summary Results of Correlation Testing	88

CHAPTER ONE: INTRODUCTION

1.1 Introduction to Topic

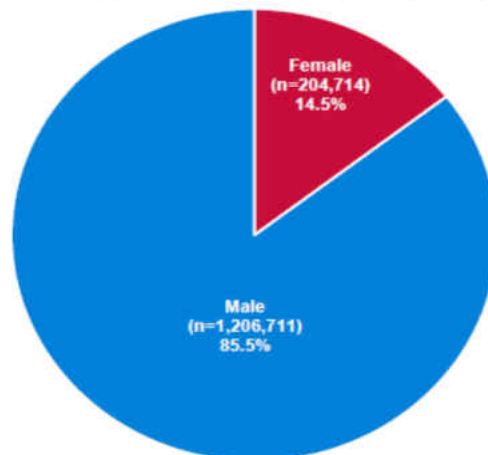
Conflict is an inevitable reality that has been continuously repeated throughout the history of humanity. There is little doubt that our unfortunate seeming need to war, which started with the earliest of our ancestry, will project itself upon future generations. Whether engagement is brought about by politics, religion, or desired conquest, states and nations have needed the ability to defend themselves from one another. Organized and well-structured armed forces were born out of necessity. These militaries have shaped world history from the conquest of Alexander the Great to the World Wars of contemporary generations. The nations of the world's armed services continue to play a major role in modern society. Military establishments have transformed throughout the centuries with the advancement of science, technology, sociology, and countless other catalyst of change. Although pronounced changes have occurred to militaries and even expansion of roles, such as disaster relief assistance, they are still at their cores a fighting force. Fighting forces must be mission ready and operationally effective in order to carry out required duties.

The foundation of a successful mission is directly contingent upon the overall health, training, and mental acuity of those tasked. Fitness has been shown to play a very important role in all of the aforementioned mission essential factors. Virtually every modern military organization realizes the importance of fitness and has designed specific tests for judging and maintaining the appropriate minimum fitness level of its individuals. Unfortunately, some of the fitness tests being used have met criticism for bias towards certain participants. Mass bias (bias towards heavier mass, not just higher BMI individuals) in the testing can result in a deleterious

and significant effect to the U.S. Military’s overall physical fitness programs and personnel. This research examines current U.S. Military physical fitness standards and the relevant work to the topic pertaining to bias in testing methods and mitigation of the problem. A test has been devised utilizing load carriage during normal physical fitness assessments that may eliminate mass bias through a simple means (weighted exercise vests). A consideration of bias significance and a historical review of the physical readiness testing (PRT) standards is done to understand the basis of original test design. This is jointly assessed with an inspection of the modern U.S. military population demographic and physiology trends. Prior research is studied concerning bias with both the “normalized” and “gender neutral” testing standards; completed along with recommendations from current research to mitigate such concerns. Research trends, ongoing work, and recommendations of such are also outlined.

2.08. Gender of Active Duty Members (N=1,411,425)

This pie graph presents the distribution of Active Duty members by gender. In total, there are 204,714 (14.5%) female Active Duty members and 1,206,711 (85.5%) male Active Duty members.



Note: Percentages may not total to 100 due to rounding.

DMDC Active Duty Military Personnel Master File (September 2011)

Figure 1: Gender of U.S Active Duty Members 2011 (DMDC, 2011)

1.2 Significance of Research

Physical fitness is an important part of military life. The U.S. Military requires its personnel to complete two physical fitness test annually, also known as physical readiness test or (PRT). Some branches, such as the Marines, require additional testing in combat readiness or Combat Fitness Test (CFT). For the most part, the PRT requirements are the same throughout the branches with minor exceptions.

Table 1: Required exercise tests by branch

Branch	Upper Body Muscular Strength/Endurance	Trunk Muscular Strength/Endurance	Aerobic Capacity
Army	2- minute push ups	2- minute curl-ups	2-mile run
Air Force	1- minute push ups	1-minute curl-ups	1.5 mile run
Navy	2- minute push-ups	2-minute curl-ups	1.5 mile-run

The scores of individuals taking the test are recorded by their respective local command personnel and entered into their evaluation files. Although overall job performance and occupational knowledge test scores are more heavily weighted in determining advancement, PRT scores are also considered and maintained on file throughout a service persons career. The promotion processes within the branches are highly competitive and a low PRT score could certainly sway the decision for one's individual advancement over another. Perhaps even more important, individuals can be discharged from the military if they fail a certain number of fitness tests within a specific period. For example, the U.S. Navy (USN) allows 3 failures within any given four year span. If a member fails one test, they are usually put on a probationary status that involves required health information sessions, monitoring, and mandatory exercise regimens. It can be seen that a failure in any PRT can lead to serious unwanted consequences. The overall

purpose of the physical readiness tests is to gauge the current fitness levels of its personnel. If the mass bias is not mitigated, the military is not accurately assessing its members and the PRT is not meeting its conceived goal. Several command level individuals responsible for fitness testing of various branches/divisions have been interviewed and agree this specific research is valuable to them and will be very beneficial to the military's fitness programs in general.

Summary of research significance/possible benefits/contribution

- Current military fitness testing bias is unfair to certain members who could lose out on career advancement or be discharged from the military altogether. A need exists to eliminate the mass bias found within the military PRT that is easy to implement and a real world solution that commanders will actually use. This research could provide that solution.
- Elimination of the mass bias could result in the military having an accurate view of overall fitness levels which is the ultimate goal of the PRT. In turn, it would improve operational readiness, management of future health programs, and the overall wellbeing of its members.

1.3 Theoretical Foundation of Work

The theoretical foundation of this research is based upon several past works which relate to allometric scaling, mass bias, and recommendations to eliminate such bias in military fitness. There is a moderate amount of work detailing the mass bias that exists within military fitness tests, but a limited amount of research offering solutions to the problem. Vanderburgh and Crowder state, "In exercise science, allometry can be used to understand how certain outcome variables of physical performance are related to one-, two-, or three-dimensional factors such as

height, cross-sectional area, or mass, respectively.” Allometric scaling laws are used to show the proportionality in these dimensional factors. In the 1RM test, 1RM is directly proportional with muscle cross sectional area which in turn is proportional to $[\text{mass}]^{2/3}$. The authors used these equations and applied them to four theoretical individuals taking the PRT. Two males of differing mass 60 kg and 90 kg and two females of 45kg and 75kg. The results show that bias towards body mass can range from 14.7% to 20% dependent on mass and the scoring system of the service (Vanderburgh & Crowder, 2006). This research outlines the bias that exists within the testing and shows the need for improvement because of bias. (Bilzon, J. L. J., Allsopp, A. J., & Tipton, M. J. 2001) pioneered research in the field of loaded carriage by experimenting with individuals and $\dot{V}O_2$ on treadmills. The authors were able to identify a mass bias towards heavier individuals in timed runs that are meant to predict fitness. They also concluded that these types of test were not a good indicator of job performance where load carriage is necessary. Positions that require a lot of material handling, for instance, should not be measured with timed distance run testing. This load carriage research was expanded upon by (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011) in which 56 male cadets were used in a loaded carriage study to determine if mass bias could be eliminated specifically for military fitness testing. Tests were conducted for 2 out of 3 events which the US Military requires scoring (timed run and push-ups). The research successfully showed loading cadets with a 30 lb back pack eliminated bias. There was correlation between mass and scores at -0.38 ($p = 0.0039$) for the pushups and 0.42 ($p < 0.0013$) for the timed distance run in the unloaded condition. The loaded run displayed different results with personnel obtaining scores at a reduction of 32.2% for the push-ups and 18.8% for the run. There was correlation between mass and scores at only -0.06 ($p = 0.661$) for the pushups and 0.06 ($p = 0.661$) for the timed distance run. The goal being a correlation of 0

between mass and repetitions or time in the case of the distance run. Research of this nature is the theoretical foundation of my work. It should be noted that all quantifiable testing was conducted on male subjects and the load carriage study involving military fitness testing was not conducted on all required events. To date, there is no study that offers evidence that adding load to female subjects would be a viable option to eliminate mass bias.

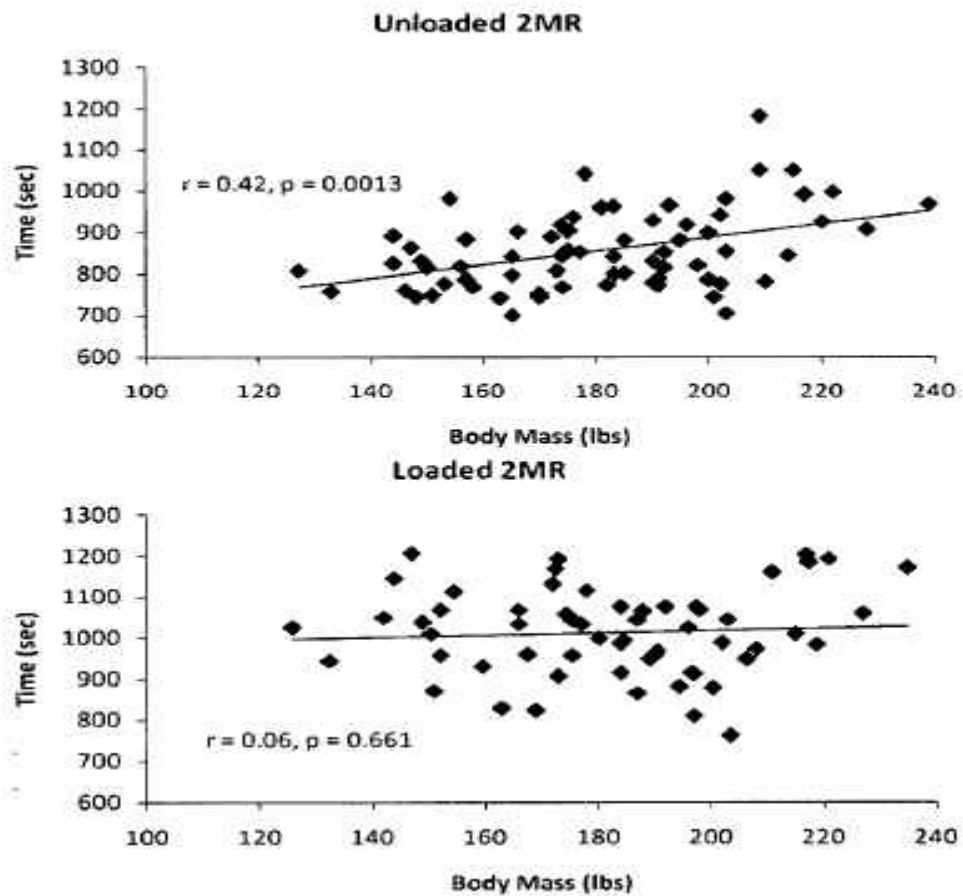


Figure 2: Unloaded vs Loaded 2MR (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011)

1.4 Research Question

The premise of mass bias lies within allometric scaling laws in that individuals with different masses will experience bias in the PRT when performing the required test exercise components. It has been revealed that service member's performance on physical fitness testing is correlated with their mass. Generally, the higher the mass for a member, the lower their PRT scores will be. Research shows adding load carriage for a male population can effectively eliminate the bias and add more occupational relevance to testing because general military duties often involve material handling different loads. Although there has been studies in this arena on males, no empirical proof exists where conclusions can be made for females in this testing or if loaded curl-ups eliminate bias as well. Curl-Ups were not tested because of the use of back packs for loading.

Research Question:

Does adding load carriage to all the separate components of the Navy Physical Fitness Testing eliminate body mass bias for women in each event?

Ho: The correlation coefficient is NOT significantly different from 0 (for each PRT event).

Ha: The correlation coefficient IS significantly different from 0 (for each PRT event).

1.5 Contribution to the Body of Knowledge

The proposed research stands to contribute to the human body of knowledge in several ways that could impact military fitness testing. There remains a large gap of knowledge for female performance related to military fitness testing. No data or research on women & load carriage specific to the PRT exists. This research would contribute to the understanding of female fitness performance relative to load carriage and military testing. If a viable method exists that could make the PRT fair, it should be fully understood for both genders so that it could be implemented at commands.

Load carriage testing for mass bias elimination has never been done for all PRT events, to my knowledge. This is due to the approach taken for loading the weight. Back packs have been used in several scenarios which limit subject movement and restrict the curl-up event. The use of weighted vests will allow the testing to be complete on all components of the PRT giving a more accurate depiction for the test overall. This portion of the research will also contribute to the body of knowledge and provide data for future testing of this load carriage method.

Data can also translate to prevalent combat fitness testing and occupational relevance attainment as well. Women in combat is a hot and controversial topic in the headlines lately. The data and findings gained from this research could be used as part of a basis for future combat fitness testing and a PRT that is applicable to actual military duties. Integration of women into combat roles continues to be researched and this research provides additional insight into that body of knowledge as well.

CHAPTER TWO: LITERATURE REVIEW

A thorough review of literature pertaining to U.S. Military fitness testing shows there is certainly room for improving the current methods used. The tests were designed in a different era and have not evolved as needed. Military demographics have changed markedly over the past several decades including a major increase in female service members which should have catalyzed improvements to testing, but have not. Additionally, considerable bias towards heavier mass members exists within the physical readiness tests of each branch which has been quantified. Recommendations have been made to mitigate the bias and improve methods overall via physically changing tests, implementing correction factors, and incorporating more occupational relevance. Many of these suggestion still need validation through research of actual subjects though. There is a major lack of data when it comes to military physical fitness and women. No research was found pertaining to load carriage for elimination of mass bias for women in PRT testing. The lack of knowledge is likely due to it being a past male dominated organization with little need for this data and too few female research subjects. Current research trends are moving towards examining women in military combat fitness, but it is specific to women in infantry roles born from recent controversy. There is a need to further investigate any bias within the PRTs of each branch and develop methods to eliminate bias through balancing muscle strength, endurance, and occupational relevance. This holds especially true for research pertaining to women service members where, once again, there is a large gap of knowledge for in this type of testing. This research is relevant and will contribute to the human body of knowledge. Additionally, it could give commanders of US Military members a simple option to eliminate a major issue that effects over 200,000 women currently serving.

2.1 History of the Military Physical Readiness Test

It is important to understand the overall objective, requirements, creation, and evolution of the Navy PRT. A review of the fitness testing history will increase the possibility of obtaining a solution to the bias problem that is in alignment with the Navy's goals. One path for solution may be viable while others would likely not fit. For example, a solution to the issue which involves expansive change, such as high tech expensive equipment, would not be a realistic solution for the military. Through understanding why the current test was derived, we can conclude some basic constraints of the test. Evaluating the objectives and changes throughout history can show a path that the military is moving toward. Understanding the picture as a whole, including gaps in existing study, will make this research more useful in contributing to the body of knowledge and a solution for the mass bias in one.

Fighting forces must be mission ready and operationally effective in order to carry out required duties. The foundation of a successful mission is directly contingent upon the overall health, training, mental acuity, of those tasked. Fitness has been shown to play a very important role in all of the aforementioned mission essential factors. Work has been done to highlight and emphasize the importance of fitness to military and their need to have a way to accurately assess it (Conway, T. L., & Naval Health Research Center, S. D. C. A. 1987). It can be seen that fitness has a very real effect not only on overall operational readiness of the Armed Forces, but the individual service person as well. Prior research has been conducted in this field which helps in understanding what factors may play a role in members failing the physical fitness test and how individuals deal with the stress of taking the current test. This work shows relations between fitness, stress, mental health, and productivity. The authors used a cross sectional study involving

all branches with the secondary data sources of a 2005 DOD health related behaviors survey.

16,146 military members of all branches were surveyed on their perceived stress levels.

Descriptive statistics and multivariate logistic regression analysis were used to determine if stress played a statistical factor in passing or failing the PRT. It was found that individuals who were experiencing personal stress are 18% less likely to pass the PRT than individuals who are not.

Those experiencing work related stress are 27% less likely to pass the PRT than their unstressed counterparts (Johnson, V. V. T 2009). Fitness, stress, mental health, and productivity are related as discussed earlier. The military has great interest in ensuring their members are at the most optimal levels of these interrelated factors pertaining to wellness.

The United States Navy was one of the first and most benchmarked branches that entered the realm of quantifiable modernized testing with the introduction of a fitness program. In 1976 the Chief of Naval Operations (CNO) issued instruction OPNAVINST 6110.1 which concentrated on developing good cardio-respiratory fitness. The program was developed after a popular aerobics program of the time designed by Dr. Ken Cooper. OPNAVINST 6110.1A was released in 1980 and was the first time a mandatory quantifiable test was to be performed. This instruction was released after a request of the Commander in Chief to assess the services current fitness levels. A joint services study group determined there were major gaps in military physical fitness programs including no way to assess overall fitness. Almost immediately, a quantifiable physical fitness test was developed as sort of a kneejerk reaction to the study group's findings. The official instruction of the US Navy that sets the rules, regulations, and requirements of this branches PRT. It is important to review this instruction fully and understand the actual goals and regulations of the instruction. Not only the original instruction, but what has evolved or not evolved from the instruction's inception to the present day document (OPNAVINST 6110.1J

2011). The initial requirements were a mix of exercises including sit-ups, pushups, pull ups, flexed arm hangs, and a timed run. There were a number of inherent problems with the initial requirements including safety concerns and gender bias (Hodgdon, 1999). For instance, there was no differentiation in minimum time for the run between male and females and sit ups were required to be done in a position that was devastating to the lower back, neck, and shoulders. To this day, it is unknown how and why the specific original testing standards and scores were derived. It was also thought that the assembled exercise set was a good measure of cardio-respiratory, muscular endurance, and muscular strength of an individual. This work reviews some of the early requirements of the PRT and the thought process on its development, goals, and pitfalls.

It is important to examine and understand what negative factors are already known about the testing. The PRT instruction was criticized for having ethnic and gender bias within body fat measurement and minimum exercise standards respectively. Hodgdon's work stated that the instruction contradicted itself in saying, "Whatever new requirements that the military Services establish, there must be an 8 to 10 percent difference between male and female standards." This was completely contradictory to the very next line in the instruction for determining female standards. That line stated "Military Services shall not derive, extrapolate, or adjust female standards using data from male subject or vice versa." It also stated later in the instruction, "The standards for one gender shall not be extrapolated from the other gender's standard or be derived from the data base of the other gender." (Hodgdon, 1999). It can be seen that the development of the PRT has not been without errors, criticism, and, bias. The reviewed literature illustrates some of the first realizations of bias within the original, or early, instruction and consideration that it was contradictory to itself, confusing, and in need of revision to bring some fairness to the

female gender's requirements. Work has been done which discusses how the Navy stopped focusing specifically on one exercise test (PRT) and started to look at service member's health as a whole (Kennedy-Armbruster, 2012). The total Physical Readiness Program was also combined with a combination of other health conscious agendas to promote overall healthy living through exercise and good nutrition. These programs still exist today and focus on overall health of military members with physical fitness as one of the main components which has shown effective results. This was an effective step by the Navy in realizing their PRT should not be a standalone entity and needed changes.

The literature reviewed gives insight into the overall goals and original thought process behind development of the Navy's required fitness testing. Testing was developed as a knee jerk reaction after a presidential inquiry about current military fitness levels could not be answered. In haste, a program was put together that was thought to gauge fitness levels of its personnel several decades ago. The research displays how vital overall fitness is to the military because of operational readiness, as well as its importance to individual personnel. It specifies unintentional contradictory test requirements and several forms of bias in early test procedures. The effects of stress on testing, gender differences, ethnicity differences, age differences, and other factors which have been problems in test development are highlighted. The review of literature on the history of the PRT depicts many flaws in a test that was designed decades ago and changed little with times. It has helped identify the need for changes in the current PRT that could come through this research and explains the intended purpose of the PRT. This research was also used to identify the significance of my proposed work.

2.2 Current US Military Demographics

The US Military has seen major changes in demographics throughout the past several decades. Changes in the general population and make up of its forces should have been a catalyst for evolution in the fitness testing requirements. A thorough understanding of trends within the population demographics and current conditions are necessary to successfully develop testing methods that are not biased towards certain individuals and remain fair for all members. There is a major lack of data for women in military fitness testing which is evident in the literature. This gap will need to be closed as more females join active duty yearly.

Table 2: Change in female population for Army & Navy 1960-2010 (US Census Bureau)

Table 510. Department of Defense Personnel: 1960 to 2010
(In thousands (2,475 represents 2,475,000). As of end of fiscal year; see text, Section 8. Includes National Guard, Reserve Other officer candidates are included under enlisted personnel)

Year	Army					Navy ²					
	Total ^{1,2}	Male		Female		Total ¹	Male		Female		
		Officers	Enlisted	Officers	Enlisted		Officers	Enlisted	Officers	Enlisted	
1960.....	2,475	873	97	762	4.3	8.3	617	67	540	2.7	5.4
1965.....	2,654	969	108	846	3.8	8.5	670	75	583	2.6	5.3
1970.....	3,065	1,323	162	1,142	5.2	11.5	691	78	600	2.9	5.8
1975.....	2,128	784	98	640	4.6	37.7	535	62	449	3.7	17.5
1980.....	2,051	777	91	612	7.6	61.7	527	58	430	4.9	30.1
1983.....	2,123	780	97	602	9.5	66.5	558	62	444	6.3	40.8
1984.....	2,138	780	98	601	10.2	67.1	565	62	448	6.6	42.6
1985.....	2,151	781	99	599	10.8	68.4	571	64	449	6.9	45.7
1986.....	2,169	781	99	597	11.3	69.7	581	65	457	7.3	47.2
1987.....	2,174	781	96	596	11.6	71.6	587	65	462	7.2	47.7
1988.....	2,138	772	95	588	11.8	72.0	593	65	466	7.3	49.7
1989.....	2,130	770	95	584	12.2	74.3	593	65	464	7.5	52.1
1990.....	2,044	732	92	553	12.4	71.2	579	64	451	7.8	52.1
1991.....	1,986	711	91	535	12.5	67.8	570	63	444	8.0	51.4
1992.....	1,807	610	83	449	11.7	61.7	542	61	417	8.3	51.0
1993.....	1,705	572	77	420	11.1	60.2	510	58	390	8.3	49.3
1994.....	1,610	541	74	394	10.9	59.0	469	54	355	8.0	47.9
1995.....	1,518	509	72	365	10.8	57.3	435	51	324	7.9	47.9
1996.....	1,472	491	70	347	10.6	59.0	417	50	308	7.8	46.9
1997.....	1,439	492	69	346	10.4	62.4	396	48	290	7.8	44.8
1998.....	1,407	484	68	340	10.4	61.4	382	47	280	7.8	42.9
1999.....	1,386	479	67	337	10.5	61.5	373	46	271	7.7	43.9
2000.....	1,384	482	66	339	10.8	62.9	373	46	272	7.8	43.8
2001.....	1,385	481	65	337	11.0	63.4	378	46	273	8.0	46.6
2002.....	1,414	487	66	341	11.5	63.2	385	47	279	8.2	47.3
2003.....	1,434	499	68	352	12.0	63.5	382	47	276	8.2	47.3
2004.....	1,427	500	69	358	12.3	61.0	373	46	273	8.1	46.1
2005.....	1,389	493	69	353	12.4	57.9	363	45	266	7.8	44.5
2006.....	1,385	505	69	365	12.5	58.5	350	44	255	7.6	43.2
2007.....	1,380	522	71	379	13.0	58.8	338	44	244	7.6	42.2
2008.....	1,402	544	74	392	13.5	59.7	332	44	235	7.7	41.4
2009.....	1,419	553	76	399	14.3	59.4	329	44	231	7.9	42.2
2010.....	1,431	566	79	407	15.1	60.3	328	44	228	8.2	43.4

Research has been conducted that gives an insight into what percentage of the U.S. population of civilians ages 17-20 met the standards required to pass physical fitness tests in

the military (Nolte, Franckowiak, Crespo, and Andersen (2002)). This data can be observed to get a baseline of the overall fit of PFT standards to the U.S. population from which it draws. Since the PRT standards were originally developed based on a majority percentage of the population averages, at least body fat, we can conclude that most of the population should be eligible for duty within the given study age group. This is not what has been concluded in the research. The authors used a secondary data source of the most recently completed National Health and Nutritional Examination Survey to match against the height and weight charts of the military services. They used the national data to calculate the amount of men and women in several subgroups to find what percentage would meet PRT standards. It was found that 13-18% of men and 17- 43% of women did not meet standards. The authors concluded a large part of the U.S. population from which the military recruits is not able to meet standards. They also concluded there is a major discrepancy between the standards of men and women that should be assessed and corrected. The research suggests there was a shift in population body fat over the decades. This research shows additional flaws within the design of the fitness test, particularly within the body mass index requirements.

Body fat percentage has been controversial in the fitness standards of the military and there is a wealth of research on it. A high body fat content has been linked to overall sickness which in turn lowers unit productivity and effectiveness (Kyröläinen et al., 2008). A study was conducted on 7,179 Finish male military personnel with an average Body Mass Index (BMI) of 26% and a range of 18-59%. The goal was to include a wide range of body types and compare the association of sick days with their BMI. The results showed that personnel with higher BMI had the most sickness absences (>7 days) with ($P < .001$) (Kyröläinen, H., Häkkinen, K., Kautiainen, H., Santtila, M., Pihlainen, K., & Häkkinen, A. 2008). The study indicates the

importance of assessing body mass index for the military. Not just to US Military, but others as well. It is important to design tests that correctly measure without gender or ethnic bias.

Higher BMI is also associated with lower physical fitness scores as shown in research conducted on the Navy. 2002 PRT cycle was used to analyze the association of BMI and PRT score results. A wide cross sectional sample of 22,314 women and 131,287 men was used to determine higher BMI scores did indeed reflect a lower PRT outcome. The study also exhibited higher BMI is correlated with increased self-reported risk factors such as elevated cholesterol and joint problems. BMI trends per gender and age were tracked to show an increase with age for men beginning at age 18 years, the constant value was 25.6, with an increase of 0, 0.765 per year ($t = 62$, $p = 0.000$), For women beginning at age 18 years, the constant was 24.5 with an increase of 0.0.159 per year ($t = 5.48$, $p = 0.000$) (Bohnker, Sack, Wedierhold, & Malakooti, 2005). The study indicates lower scores and higher risk factors with increasing BMI. This information has been taken into consideration for my test design as a safety concern. The work also indicates the negative effects of gaining weight to try and improve scores through correction factors and allometric scaling which has been a complaint of those against certain types of bias mitigation options.

Flanagan examined the relationship between BMI and waste circumference in Canadian military forces with similar conclusions to related studies (Flanagan, 2008). The author's hypothesis was that increasing BMI and waste circumference related to lower performance in Canadian physical fitness testing. Analysis for 127 males with a BMI=27.7 kg/m² (± 3.4) and mean waste circumference of 96.99 cm (± 9.8 cm)) indicated increasing BMI and waste circumference correlated ($P < 0.05$) with lower results for tests including, Low High Crawl,

Sandbag Carry, Land Evacuation V02max, Push-Ups and Sit-Ups. Flanagan's study also noted a growing trend in BMI. This research shows that BMI is not just a factor for the selected tests of the US military (run, push-up, and curl-up) but for other combat fitness based exercises. These findings suggest my research data could be translated to other tests, and that body mass index does play a role on Comb at Fitness Testing as well. A disturbing increase in a trend that has such negative implications on the overall wellbeing of the members of the military and its mission has been brought to the attention of the heads of the Department of Defense (DOD). In a study of overweight and obesity in active duty U.S. Military members it was shown that obesity grew from 25,776 members in 1998 sharply to 86,186 members in 2010 (Armed Forces Health Surveillance, 2011). The trend continues today and will certainly hinder the PRT evaluations of those who fall in the high BMI obese category. These continued changes in body type trends will exacerbate the mass bias effects on test scoring.

The rise in the branches' female population has grown markedly over the past 5 decades. For example, in 1960 there was a total of 31,700 females serving in all the branches combined. A DOD Census completed in 2010 showed a total of 205,500 women serving active duty. When speaking in terms of total personnel, during 1960 only 1.3% of active duty personnel were women. In contrast of today's forces, females comprise 14.3% of the total population (U.S. Census Bureau, 2012). An incremental increase of the females over the past decades has intensified the spotlight on gender fairness within the ranks. The information from the 2012 census gives a good overview of changes o demographics of the military and the need for evolution on testing. It supports the relevance if this research in that there is very little data available for all these women serving. The trend of women joining continues and the gap of knowledge is growing.

A woman's body has definite physiological differences than a man's. Studies have shown that women have a substantially greater chance of getting an injury during sports or training (Bell, Mangione, Hemenway, Amoroso, & Jones, 2000). Eight hundred and sixty one subjects were observed during basic military recruit training for eight weeks. Key variables such as demographic characteristics, body composition, and physical fitness were measured prior to training. Injuries to female recruits were more than double during the training period relative risk [RR] = 2.1, 1.78 –2.5) and experienced serious time-loss injuries were close to 2.5 times more often than men ([RR] = 2.4, 1.92–3.05). The physical fitness levels of women entering training were lower than that of their male counterparts, but the women's relative fitness levels increased more than that of the men's increase over the eight weeks. Women are more prone to injury during training and other factors must be taken into consideration as well. This research indicates the difference between genders and gives some insight on fitness levels after extended training periods. The importance to acclimate test candidates to loaded carriage testing is critical. The research was used as a safety factor in experimental design because we know injury rates are up to 2 times higher than men.

Additional circumstances also need to be considered when viewing military fitness and gender. Pregnancy can be a concern when an individual holds a mission critical position that cannot be easily occupied by another due to training and skills. Time out for pregnancy can be significant and women are usually allotted 6 months after birth to return back to physical fitness testing. Recent research has shown this time may be insufficient (Weina, 2006). Fifty –four women took part in a test to measure their fitness scores prior to and after pregnancy. Two post pregnancy standard fitness tests were given at 6 months and 1 year after child birth. All women for this part of the research has passed the PRT cycle prior to pregnancy. The mean decrease in

test scores were 40 points for the six month test with ($p=0.000$). The second test administered showed an improvement in mean results with just an 18 point decrease in scores prior to pregnancy ($p = 0.000$). Recommendations from the research were to increase the time women are waived the fitness test and full physical duty. This is relevant to my research in that individuals selected who were pregnant and gave birth in the past year may have an effect on raw scoring. This type of information helps guide questionnaires needed for getting test demographic correct.

Questions of female physical abilities have come into the spotlight even more in recent times. Traditional non-combat support type positions are where females are usually occupied. It is thought that physical fitness is less of a critical factor in most of these positions. In fact (Burland & Lundquist, 2013) argue that there is a duality within military forces between combat and support type positions that has been quantifiably proven. Silos such as these can make gender bias for combat and other physically tolling positions even greater to overcome for females. It is important to eliminate any gender bias in physical fitness testing that could add to the delay of breaking this barrier. A study that was completed on body fat standards of 1,038 male and 374 female members at 3 separate Army posts originally showed 11% of men and 17% of women overweigh/over fat by standards. New body fat equations, thought to be a more valid estimator, were then used to recalculate the original data. The change included an extra abdominal circumference measurement which significantly changed the results. The amount of women who were estimated to be above limits dropped to 12% which is in line with the males tested. The newer measurement technique has been recommended by the author going forward. This research supports the importance of eliminating bias that exists in testing so that silos within the military can be eliminated resulting in a more effective, optimized, and coherent military. Additionally, it shows the importance of selecting the correct assessment methods to eliminate

bias. A 12% drop in failure rate is certainly significant and testing data can help equalize male vs female BMI limits.

The literature reviewed pertaining to current US Military demographics and trends identify several changes that cannot be ignored when evaluating the population for test design. Changes occurred over the past several decades in population body mass indexes, gender, and ethnicity which are major factors acknowledged in the research. Shifts in BMI were especially evident when standards were changed in 1994 to one max %. Ethnicity bias in BMI measurement played a controversial role during this time period and measurement techniques were reevaluated. Changes like these are required to bring fairness for all members.

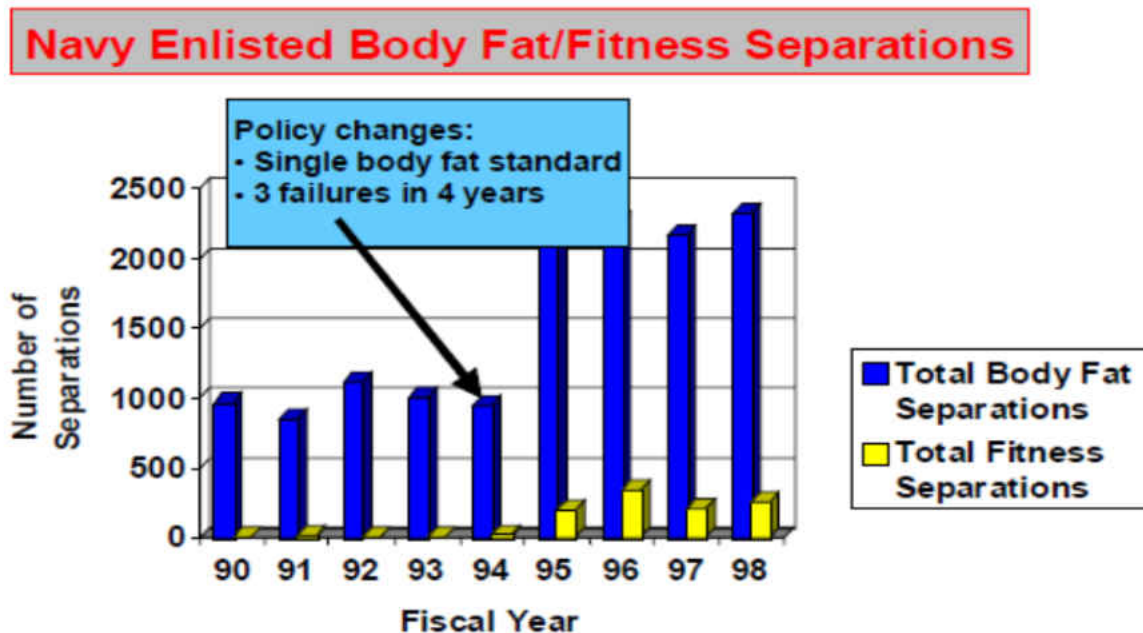


Figure 3: Enlisted BMI Separations 1990-1998 (Hodgdon, 1999)

Unfairness is still shown to exist in fitness testing for women ranging from physiological safety concerns for CFT, BMI measurements, and pregnancy policies, which have been found within the literature. Census numbers clearly paint the picture of change in the population and past research agrees there is much room for improvement in some of the policies towards the current population.

Table 3: Active Duty Military Personnel Gender per Branch (DMDC 2011)

Service Branch	Officers		Enlisted		Total		Ratio of Officers to Enlisted	
	Male	Female	Male	Female	Male	Female	Male	Female
Army	81,791	15,760	403,631	60,255	485,422	76,015	1 to 4.9	1 to 3.8
Navy	44,689	8,520	223,036	43,896	267,725	52,416	1 to 5.0	1 to 5.2
Marine Corps	20,537	1,328	166,798	12,363	187,335	13,691	1 to 8.1	1 to 9.3
Air Force	53,187	12,291	213,042	50,301	266,229	62,592	1 to 4.0	1 to 4.1
Total DoD	200,204	37,899	1,006,507	166,815	1,206,711	204,714	1 to 5.0	1 to 4.4
	238,103		1,173,322		1,411,425		1 to 4.9	

DMDC Active Duty Military Personnel Master File (September 2011)

The literature also recognizes the importance of fairness to female members in order to break down existing silos as barriers to equality. This past work is used to identify the need for my research because of the changing demographic. It is also used in part for test design in methodology. It is important to match the test subjects to the current demographics of the Navy female force so that the wrong data is not gathered. It also brings up safety concerns and other conditions (such as time after pregnancy) specific to female testing that have major consequences on test design in methodology.

Table 4: Confidence Interval (percentage) of injury among female and male Army basic training recruits (Bell, Mangione, Hemenway, Amoroso, & Jones, 2000)

Injury Type	Men	Women	Relative Risk	95% CI
One or more injuries	27%	57%	2.1	(1.78-2.50)
Time-loss injury	17%	41%	2.4	(1.92-3.05)

2.3 Penalties in PRT towards Body Mass, Little Occupational Relevance in Testing

A review of literature pertaining to the design of fitness testing has revealed how body mass bias has become a major concern. Additionally, there are problems inherent in the exercises chosen by the military because they do not accurately represent general tasks performed by military personnel; they are not occupationally relevant. It may be difficult to balance a test between occupational relevance and overall fitness, but some literature shows it may be attainable through selecting a variation of predictive tests which must be prescreened.

Vanderburgh and Crowder state, “In exercise science, allometry can be used to understand how certain outcome variables of physical performance are related to one-, two-, or three-dimensional factors such as height, cross-sectional area, or mass, respectively.” Allometric scaling laws are used to show the proportionality in these dimensional factors. In the 1RM test, 1RM is directly proportional with muscle cross sectional area which in turn is proportional to $[\text{mass}]^{2/3}$. As a person’s mass increases, their strength increases as well, but not as you might think. An exact replica of a person that is 20% larger cannot lift 20% more weight, but only 13% more weight. In order to eliminate the bias in a 1RM test you can simply make the exponent negative yielding the equation $[1\text{RM} \times \text{mass}]^{-2/3}$ (Vanderburgh & Crowder, 2006). Equations similar to the one used for eliminating bias in the 1RM test can also be derived from allometric scaling laws and used on repetition based endurance exercises as well. The allometric scaling equation used to correct bias in sit ups and pushups is $[\text{repetitions} \times \text{mass}]^{1/3}$. We can see that the exponent in this case is positive because repetition endurance based exercises are negatively correlated with mass. The theoretical equation used for $[\text{VO}]_{2\text{max}}$ on the timed run is $[\text{time} \times \text{mass}]^{-1/3}$. Note that even though the exponent is negative in this

correction factor, mass is still negatively correlated because a smaller time is better. The authors used these equations and applied them to four theoretical individuals taking the PRT. Two males of differing mass 60 kg and 90 kg and two females of 45kg and 75kg. The results show that bias towards body mass can range from 14.7% to 20% dependent on mass and the scoring system of the service. The authors used allometric scaling equations to determine body mass bias and this is important in determining possible limits of bias. My testing differs in using these equations as a comparison only to actual fitness testing results. Whereas the authors of this research did not actually perform PRT and theoretically calculated results from existing well established equations. These equations remain a relevant way to check testing results against what is expected in mass bias, but the data cannot be translated directly to female participants.

The PRT component exercises have also been criticized for not truly representing the typical work that soldiers, sailors, and airmen complete during their typical day. That is, the endurance based exercises are not occupationally relevant to normal military tasks that generally require lifting, load carriage, and material handling. Moreover, although heavier mass individuals are better suited to these tasks and tend to perform better in terms of this work related fitness, they have a career advancement disadvantage compared to their lighter coworkers. (Vanderburgh, 2008). In Vanderburgh's research he elaborates on the biological scaling laws of allometry and how these are once again applicable in occupational relevance of the PRT. The muscular and aerobic testing of the military fitness test are performed using body weight as the only resistance. Part of the attractiveness of using body weight only as resistance is because there is no equipment required to perform the tests. Although this is a great from a logistical perspective, the military is not going a true view of its service member's ability to perform work related tasks. Persons performing military work very often must move themselves and the added

weight of a load. There is a very poor correlation of PRT exercises based mostly in repetitive endurance to tasks that require more absolute strength and power. Load carriage is a very common task in the military, especially for those stationed onboard Navy ships. This work shows how the current PRT exercises do not represent the bulk of military duties on a daily basis. There is very little occupational relevance. Vanderburgh argues using body weight only for PRT does not assess a true performance of actual work related tasks. This is relevant to my research in that adding load carriage can increase the occupational relevance of the testing. This will make the exercises more like a material handling situation, where added weight is moved.

Twelve male individuals were assessed during a volunteer testing exercise for steady state Oxygen uptake ($\dot{V}O_2$) on a treadmill (Bilzon, Allsopp, & Tipton, 2001). The participants were measured both with and without an 18kg loaded backpack. On a following test, at least seven days later, the same participants were asked to run to exhaustion with the 18kg pack. The results showed that there was a strong relationship between ($\dot{V}O_2$), body mass, and lean body mass. Statistical analysis showed a ($r = -.087, P < 0.01$) for body mass and ($r = -.074, P < 0.01$) for lean body mass. There was also a moderate to strong relationship between body mass and exercise tolerance time when carrying the extra load weight ($r = .069, P < .05$). There was very little correlated relationship between loaded weight tolerance time and ($\dot{V}O_2$) ($r = 0.12$). Additionally, load bearing tolerance increased with heavier body mass. The authors concluded that aerobic exercises that measure aerobic units relative to body mass, such as timed runs used in PRT events, are unsuitable for measuring ability in tasks that require load carrying. They also concluded that these types of tests, timed distance runs, are bias to heavier personnel. This research adds upon the evidence that a bias exists within the PRT testing. The

authors used load carriage on treadmills and compared to unloaded treadmill tests to determine distance run exercises were not good indicators for load carrying and bias to heavier mass individuals. This testing differs in that only males were tested and concentration was on identifying bias rather than eliminating it. Work was also only completed on treadmills and for one evolution (distance run only). The research could be extended to all PRT components.

In a separate study, 94 women (age = 27.4 +/- 6.7 years, with a body mass of BM = 60.3 +/- 8.4 kg) were asked to perform $\dot{V}O_{2\max}$ testing on a motorized treadmill through indirect calorimetry (Vanderburgh, P. M., & Katch, F. I. 1996). Lean body mass (LBM) was found for each test subject through hydrostatic weighting. The authors utilized allometric scaling for $\dot{V}O_{2\max}$ basing their initial expectations on prior “theory of similarity” research where the body mass exponent is proposed to be 2/3 or 0.67. Their allometric scaling equations yielded a different exponent of .61 +/- .27 (95% CI) for total body mass, but when LBM was used the exponent changed to 1.04 +/- .26 (95% CI). The research validated the bias towards heavier individuals using body mass for the expression $mL \cdot [BM]^{-1} \cdot [min]^{-1}$. The more accurate equation was found to be $mL \cdot [BM]^{-0.7} \cdot [min]^{-1}$ thereby eliminating the bias. The work also found that the ideal expression for women using lean body mass to be $mL \cdot [BM]^{-1.04} \cdot [min]^{-1}$. The expression is very close to the currently used standard and indicates body mass bias only exists in a population with heavier mass due to body fat and not LBM. The U.S. Military could benefit from such bias neutralizing expressions, but it is important to have an accurate way to measure body mass indexes to properly incorporate them into the physical readiness programs. This work is some of the very limited data available on women pertaining to a portion of military fitness testing. It generates additional evidence of body

mass bias in heavier mass individuals and suggests different exponents to scaling equations for using lean body mass versus LBM. This work differs from my research because it deals again with equation based elimination of bias. It also only concentrates on distance run and fails to account for the bias found within the other exercises in the PRT. Regardless, it is one of the only research papers available pertaining to women and military fitness. No load carriage was used in determining the mass bias in women.

Studies pertaining to measuring occupational relevance testing have also been conducted on organizations that are paramilitary in nature (Sell, K. M. 2006). A team of 81 firefighters from the same fire department participated in a study to determine what physical fitness exercises were the most occupationally relevant and best predictors for passing a simulated firefighting task scenario test. Body mass was obtained through 3-site skinfold measures and tests were completed in the 1.5-mile-run, 1 RM press, vertical jump, grip strength, sit-ups, push-ups, and sit-and-reach test. The fire fighters also completed a simulated firefighting task scenario test for occupational relevance. The fitness test scores were analyzed along with the occupationally relevant test. It was found through multiple regression techniques that 50.03% of the variance in the occupationally relevant test could be explained by the physical fitness variables ($p < 0.05$). Although there is still a lot of variation to explain, it was found that flexibility, grip strength, and percent body fat held heavy weight within the testing. Work such as this can help identify which standard physical fitness exercises could be the best predictors of military occupation relevance and still be able to measure overall fitness levels. This research suggest that occupationally relevant tests can show a wide variance in overall fitness correlation. It is imperative that any tests suggested are well vetted and compared prior to implementing. There was a combination of testing that did yield good predictive results for fitness. Research like this can help identify if

other OR exercises within the military would be a good predictor of fitness as well. This work differs from my testing because it was performed on males, nonmilitary organization and testing, and concentrated on finding correlated vents rather than mass bias elimination.

It can be seen through the literature reviewed that there are several methods to calculate the bias that exists including theoretically by allometric scaling equations (up to 20% bias). Several past tests all conclude bias exists although different methods were used to calculate it. This is significant work incorporated into the methodology of my research validating the bias for the initial unloaded PRT. This literature also provides more support of the need for research in this area to make the test fair. The lack of fitness data in reviewed literature on women once again represents the gap of knowledge and the need to close it. Work reviewed pertaining to occupational relevance exposes the inappropriate form of current testing and helps guide the theoretical framework for loaded carriage studies by showing lessons learned.

CALCULATION OF EQUIVALENT POINTS AND RESULTING PENALTIES BASED ON BODY MASS DIFFERENCES OF 60 TO 90 KG FOR MEN AND 45 TO 75 KG FOR WOMEN AND AN AGE OF 28 YEARS

Event	Male		Female	
	60 kg	90 kg	45 kg	75 kg
	Maximal Performance (Equivalent Points)	Equivalent Performance ^a (Equivalent Points)	Maximal Performance (Equivalent Points)	Equivalent Performance ^a (Equivalent Points)
Army				
Push-ups, 2 minutes	77 (100)	67 (89)	50 (100)	42 (90)
Sit-ups, 2 minutes	82 (100)	72 (89)	82 (100)	69 (86)
2-mile run time	13:18 (100)	15:14 (78)	15:48 (100)	18:44 (74)
Total points (% difference) ^b	300	256 (14.7%)	300	250 (16.7%)
Air Force				
Push-ups, 1 minute	57 (10)	50 (8.75)	41 (10)	35 (9.0)
Abdominal crunches, 1 minute	53 (10)	46 (8.5)	47 (10)	40 (8.5)
1.5-mile run time	9:36 (50)	11:00 (42)	11:24 (50)	13:31 (40.5)
Total points (% difference) ^b	70	59.25 (15.4%)	70	58 (17.1%)
Navy				
Push-ups, 2 minutes	84 (100)	73 (85)	46 (100)	39 (80)
Curl-ups, 2 minutes	101 (100)	88 (80)	101 (100)	85 (75)
1.5-mile run time	8:55 (100)	10:12 (85)	10:17 (100)	12:11 (85)
Total points (% difference) ^b	300	250 (16.7%)	300	240 (20%)

^a Equivalent performance is the raw score for the heavier individual to be physiologically equivalent to that of the lighter individual. For example, the 60-kg Army male subject's 77 push-ups are equivalent to 67 push-ups for the 90-kg male subject. Similarly, equivalent points are the corresponding points from the scoring tables.¹⁻³

^b This percentage difference is the lighter vs. the heavier individual's total points. For example, the 60-kg Army male subject's performance is equivalent to 256 points for the 90-kg male subject, a 14.7% difference.

Figure 4: Calculation of Theoretical Mass Bias (Vanderburgh, P. M., & Crowder, T. A., 2006)

2.4 Suggestions to Correct/Mitigate Bias in Testing and Increasing Occupational Relevance

Literature was reviewed on research suggesting options to remove mass bias and add occupational relevance. These methods differ greatly in approach and range from changing the actual test methods to implementing correction factors to scoring. Some research also suggests changing traditional ways to calculate fitness (such as Vo2max calculations) which require updating measurements through statistical means relative to individual mass via regression.

The Rockport Walk test has been validated as an accurate predictor of [VO] 2max in an Air Force population (Weiglein, Herrick, Kirk, & Kirk, 2011). The Air Force started allowing

the 1 mile walk in place of the 1.5 mile timed run in 2010 for personnel who are medically unable to participate. The Rockport Walk test had not been assessed for use in an Air Force population prior to the work of the researchers. 24 active duty volunteer participants were taken from Scott Air Force base in Illinois. 8 officer and 16 enlisted servicemen were used to get a ratio representation of the Air Force. The men were asked to perform the 1 mile Rockport walk test, 1.5 mile timed run, and a [VO] 2max expired gas test. The run and walk were completed on an indoor track and the [VO] 2max test was completed indoor on a motorize treadmill. Statistical analysis of the results were done including a paired samples t-test, regression analysis, and Person product moment correlation. The results of the analysis found there were no significant differences between measured and predicted [VO] 2max using the 1-mile walk equation ($p = 0.177$) There was also a strong correlation between predicted and measured [VO] 2max Max where Pearson's product moment correlations were high ($r = 0.817$). The 1.5-mile run time and measured VO2 Max also produced a high negative correlation ($r = -0.890$). Perhaps most importantly, there were no significant differences between points scored on the 1-mile walk and 1.5-mile run using the Air Force scoring tables. A high correlation between the tests were also found ($p = 0.573$, $r = 0.827$). The authors also compared the results of their testing to national averages of American College of Sports Medicine (ACSM) data. The comparison showed the 1.5-mile run and 1-mile walk against ACSM percentile rankings had a high correlation ($r = 0.975$). The research suggests that the 1 mile walk test is a valid indicator of predicted [VO] 2max in an Air Force population. Tests such as the Rockport walk test can be a useful tool in deciphering service members aerobic fitness levels since they predict [VO] 2max as a quantifiable value. This work was completed as an attempt to identify another method for assessing aerobic fitness of Air Force personnel. The Rockport Walk test was shown to be a

valid indicator of VO₂max and this was more accurate than current methods because it gives a quantifiable value. The test was completed on tracks and treadmill for VO₂ testing. Good indicator that there are other reliable methods to accurately assess service members. This work once again differs from my research in that it only examines the distance run and not the pushups or curl ups. It does not address bias elimination, but rather offers a more accurate measure of cardio-fitness. It was also only conducted on male members and continues the trend of a gap in knowledge for data on women.

There are also researched methods to more accurately record [VO] 2max readings (Savonen et al., 2012). The current method for aerobic fitness estimation, [VO] 2max, is done by dividing maximum oxygen consumption by body weight (per wait standard). Savonen et al., submit that a more statistically accurate technique can be used. The authors used data available through the Dose-Responses to Exercise Training Study (DR's EXTRA) which was a randomized controlled four year trial on the effects of regular physical exercise and diet. The study included data for 635 men and 638 women including body mass and direct measure of [VO] 2max. The research shows a more statically correct method can be used to counterbalance the effect of weight on maximal testing. [VO] 2max can be adjusted for body weight through regression or (adjusted standard). The research demonstrates an increase of [VO] 2max during the study was found to be biased. Heavier individuals showed lower results (negative mean residuals) due to the statistical techniques used, while lighter individuals where shown to have an overestimated intake level (positive mean residuals). The weight standard showed a [mL min] ⁻¹ increase during the study for women of 20.9 [mL min] ⁻¹ min and men at 26.4[mL min] ⁻¹. The actual increase per kg was shown to be only 7.0[mL min] ⁻¹ (95% confidence

interval: 5.3--8.8) for women and 8.0 [mL min]⁻¹ (95% confidence interval: 5.3--10.7) for men. Methods like these could be a more accurate way to truly determine [VO]_{2max} within a given population. It should be noted that this research was conducted on a specific age group of 57-78 years which is nowhere representative of a military forces demographic. Savonen et al, have also determined a more accurate way to estimate cardio-fitness. Their approach was different in that they used data from existing medical studies and change the actual equations for calculating through adjusted standards in utilizing regression. This work could be used to eliminate some of the mass bias seen in the distance run. This testing differs from my research because of several reasons. It was a distance run only, equation based mitigation, and the wrong age group and demographic (57-58 years). Additionally, existing data from nonmilitary fitness testing was also used. The methods used could be applied to a different age group and may be a better means to isolate mass bias within these group ranges.

The curl-up, which is used in all branch PRTs, has also been examined through research. It has been shown that anchored curl-up used in the PRT can put as much as 452 lbs. of compression on the spine which exceeds National Institute for Occupational Safety and Health occupational standards (Peterson, 2013). Although it has not been proven, there is suspicion that the curl-up is responsible for part of the Navy's increase in lower back injuries. The curl-up is not an occupationally relevant exercise either. Although it has not been proven, there is suspicion that the curl-up is responsible for part of the Navy's increase in lower back injuries. The curl-up is not an occupationally relevant exercise either. A majority of military workday tasks include lifting, carrying, and pulling of loads. Variations of the plank exercise have been researched by the Navy and sources independent of the military. 1078 Canadian fire fighters and 181 college

students participated in a plank study over a period of 3 years. The Navy used this research data and also conducted a study in 2011 to obtain data on the plank as a suitable replacement for curl-ups. Although the data was not released, the Navy went as far as developing testing standards through analysis of past research and their own. They established test scoring with the data and also determined it could be a viable test replacement. To this date, the plank has not been implemented in any military branch. This research brings to light the possible negative effects of curl-ups. This emphasizes the importance my research must put on safety while conducting the test. The research recommends replacing a component of the test (curl-ups) to increase occupational relevance. This is accomplished by utilizing the plank maneuver in place of the traditional curl-up. A large study was done confirming the viability of the plank as a replacement. The Navy was said to examine the plank, but never implemented its use. This research differs from my work because it is isolating one part of the test and changing it. It does not necessarily increase occupational relevance and has not been studied for mass bias.

In prior research conducted 56 Reserve Officer Training Corps (ROTC) cadets participated in 2 separate tests involving load carriage (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011). Mathematical modeling from prior test data indicated a 30 lb. load would effectively eliminate bias. The purpose of the research was test the model with actual subjects. The first test was the 2 mile distance run (Army) loaded and unloaded with the 30 lb. back pack. The second test was pushups loaded and unloaded with the same load. I was determined there was significant mass bias within the current unloaded standard in both timed run and push-ups. This is not surprising given the evidence that already exists determining there is mass bias in these tests. There was correlation between mass and scores at -0.38 ($p = 0.0039$) for the pushups and 0.42 ($p < 0.0013$) for the timed distance run. The loaded run displayed different results with

personnel obtaining scores at a reduction of 32.2% for the push-ups and 18.8% for the run. There was correlation between mass and scores at -0.06 ($p = 0.661$) for the pushups and 0.06 ($p = 0.661$) for the timed distance run. The loaded run and pushups test seems to eliminate the bias shown with no load. Implementing a loaded run and pushups may require more equipment for personnel performing the PRT, but it has been shown to eliminate bias and improve occupational relevance of testing since most duties include moving more than just body weight. This testing is the most relevant to my research because it involves load carriage to mitigate mass bias in actual military fitness testing. The results concluded that load carriage successfully eliminated mass bias, but testing was only conducted for men. Additionally, the use of back packs for the load would not allow for performing testing on the curl-up portion of the PRT. Back packs require more training and familiarity for maneuvering during exercises. My research differs in that I will be using weighted vests that can be used during the curl-up portion of the PRT to gather data. This may successfully allow bias mitigation through the full range of testing required by the military. I am also concentrating study on women due to the lack of data that proves this method could be viable for both genders. If proved a viable method of bias elimination, the test would be proven out for both genders and all portions of the PRT. Essentially, it could be implemented for full use.

A study conducted by Clark introduces the fact that female recruits performing basic training are 4-6 times more likely to suffer from this injury (Clark, H. D. 2013). 12 people, 7 male and 4 female, from ages 18-29 were asked to perform a loaded march on a treadmill for 80 minutes with 20kg packs. The participant's stride length, frequencies, steps, and gates were all observed. The study concluded the increase in stress fractures for females could be attributed to gender differences in gait mechanics. In particular, deficits in hip abductor strength result in

inadequate absorption of impact force. The fact that women's strides are generally shorter than a males exacerbates the problem in increasing the impact frequency during any raise in cadence of a march. Care must be taken to incorporate events into the PRT which are safe and eliminate bias both. Load carriage along with the other suggestions found through past research may be the answer, but should be thoroughly vetted. Exercises that may be suitable replacements or additions to the PRT must be evaluated thoroughly for safety concerns. For example, the 1RM bench test discussed prior, is an excellent indicator of muscle strength, but is prone to cause injury. Load carriage has been shown to cause stress fractures within the military ranks. This work is relevant to the safety of the study and should will need to be investigated further to determine loaded carriage would not cause increased injury during PRT evaluations. Female gates are generally shorter than males increasing the amount of impact over the same length of distance traveled. This could contribute further to the possibility of stress factors. The PRT is not the same as the researched because the PRT does not require 80 min of loaded time to complete.

A good indicator of muscle strength is the 1RM bench test (1 repetition max weight), but this test can be dangerous as history has shown. An ideal solution has been found to eliminate the risk through predicting the 1RM bench through another less dangerous exercise (Moschella, 2013). The military PRT could obtain an accurate measure of absolute strength (upper body) to counterbalance the endurance based exercise bias while limiting risk of injury. 50 women were selected for the study who consistently trained at least 2-3 days per week for 3 months. Data was collected for various body measurements, body mass, and lifting abilities then analyzed statistically. The author used stepwise multiple regression analysis to formulate three equations useful to predicting 1RM bench with a 4RM triceps lying extension and two other variables to increase prediction accuracy if needed ($p < .001$). There were three separate prediction equations

formulated each increasing in R^2 to explain variance. The first equation which included only triceps extension yielded an R^2 value of .62. The second prediction equation included triceps extension and chest circumference and yielded an R^2 value of .74. The greatest explanation of variance was seen with the third equation which included triceps extension, chest circumference, and % body fat with an R^2 value of .88. The PRT also includes component tests that are used to measure muscle strength and endurance (pushups, sit ups). It has been established that these types of exercises measure muscle endurance more than muscle strength partly because of its repetition based requirement during performance. There has been studies done which include the types of exercises that are more an accurate measurement for muscle strength and endurance. The goal is to neutralize the bias that exists toward heavier individuals for endurance based exercises. Adding exercises that measure absolute strength can be difficult because of the equipment usually required. This testing gives options to testing that could replace some of the existing PRT exercises. Additionally, the exercises do not require a lot of equipment and are safer than bench press. This is a great example of how exchanging some components of the PRT could bring more occupational relevance balance through testing strength versus endurance. This test differs from my research because it is nonmilitary related testing. It would substitute actual parts of the PRT and does not include the distance run.

Another means of mitigating the bias within the PRT is to simply correct the scores after members complete their required current PRT exercises. Data from past studies involving exercises used by the military has been analyzed and used to develop scoring correction factors (Paul V., 2007). Lower mass limits were selected of 125 and 150 lbs. for several reasons, but most importantly these masses hold to be the most optimal for performing PER tests and consistently correlate to higher scores relative to their sub groups. Mass increasing above these

limits would be score corrected using factors formulated by the author deviated from past researched allometric scaling equations. Upper limits are decided through body mass per gender and age group; maximum limits are set so that high body fat individuals are not rewarded. The serviceman or women's obtained score would be multiplied by a factor to eliminate mass bias. There is an advantage of simplicity in this technique above ratio scaling which can involve confusing units and or equations. The existing PRT scoring tables would not require a change and body mass bias could be eliminated. This research examines the use of scoring correction factors to eliminate the existing mass bias with the PRT. The author states that an inherent problem with using correction factors is that they are mathematically equivalent to scaled scores, but they have don't generally have the advantage of preserving the original units of the raw score. They still require tables and are not intuitively attractive to non-exercise scientists. This author's research seeks to correct that issue and make scaled scores more easily interpretable. Opponents to the use of mathematical corrective factors say they rely on theoretical data to eliminate the counter effect and that they reward higher BMI individuals rather than actual performance of exercises. One could try and gain weight to obtain better scores. This work differs from my intended research because it once again utilizes equation based mitigation rather than actual tests. It is difficult to gain confidence from commanders conducting the PRT because of the theoretical vs quantifiable basis of final score results.

This group of research from the literature review shows the capability of different methods to eliminate mass bias from testing and improve occupational relevance in some scenarios. All the authors take different routes in reaching the goal of fairness in testing and there is some disagreement in methods used. Disagreement exist in methods incorporating allometric scaling equations alone because it is thought it provides incentives to individuals to gain mass for

better scores. There are those who do not agree with correction factors because the theoretical basis is difficult to comprehend for non-exercise scientist. Changes to the actual tests have been recommended, but they often only fixate on one component of the PRT leaving bias in the remaining events. Some of these tests have been approved for official use such as the Rockport Walk Test as an indicator of cardio-fitness, but only to be used in scenarios where the member is injured. Literature shows many methods exist, but they do not agree on procedures to get the job done. These studies are relevant and contributes to my research in several ways. They are used in the methodology to build upon proposed testing procedure. It identifies the possibility to mitigate mass bias for the PRT in males adding to the validation of the methodology. Furthermore, the review of this literature also addresses some safety concerns in my methodology concerning female load carriage over long distances and times.

Table 5: Equations generated to predict strength through triceps extension (Moschella, K. N. 2013)

Variables	Equation
(#1) 4RM Triceps Extension	$4RM\ CP = 11.139 + 1.105(4RM_T)$
Weight	
(#2) 4RM T + Chest	$4RM\ CP = -41.282 + .870(4RM_T) + .681(Chest)$
Circumference	
4RM T + Chest C + % Body	$4RM\ CP = -46.7 + .802(4RM_T) + .838(Chest) - .254$
Fat	(Body fat)

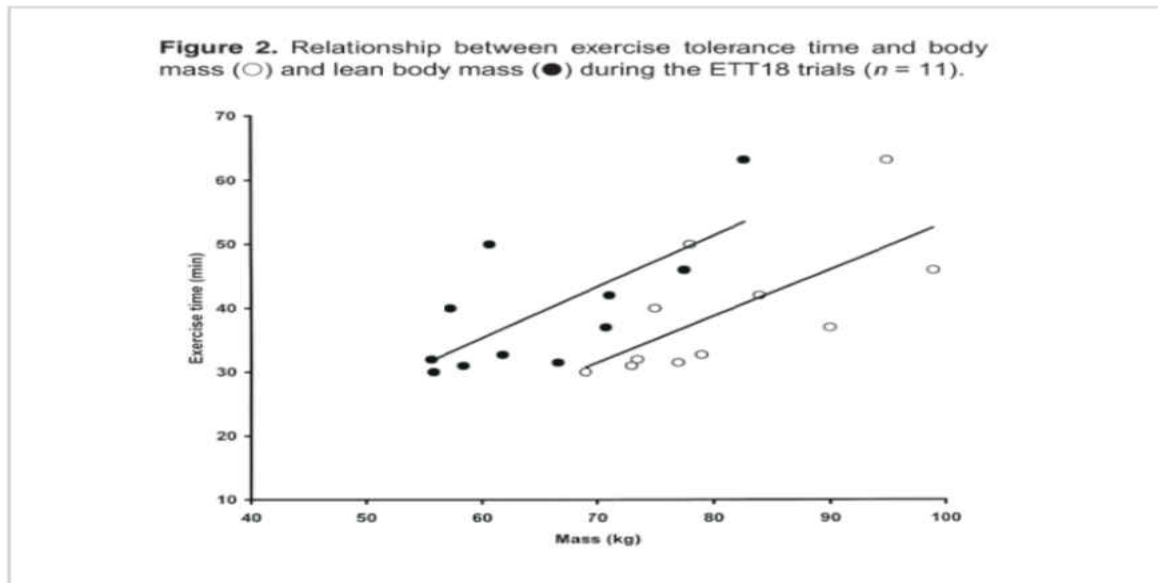


Figure 5: Relationship between exercise tolerance time and body mass/lean body mass Bilzon, J. L. J., Allsopp, A. J., & Tipton, M. J. (2001)

2.5 Bias and Controversy in Specific Occupational Specialty Fitness Testing (Combat Testing)

Literature has been reviewed which shows the need for occupational relevance in testing for more specific combat testing as well. Bias exists in several forms within the traditionally male based US Military organization. Mass bias is evident in the PRT testing and some literature suggests the combat fitness testing has bias towards the female gender as well. Several works exist that determine detail to attention must be observed when trying to create tests to assess job performance. A test can be designed which measures some occupational aspects, but is a poor indicator of overall fitness. A balance must be reached between the two. A controversial topic which is making headlines is equality for women within the ranks. Orders have been issued to begin integration of women into combat support roles and this is breaking barriers within the military. Unfortunately, some literature shows gender bias towards women specific to fitness and

performance in these positions. A gap of knowledge remains for analyzing tests that are meant to determine combat fitness and maneuvers expected of individuals in fighting roles.

Demographics report show a significant increase in female service members over the past several decades. The introduction of females into a male dominated organization has been deemed as successful, but was not an easy task. Sexual harassment and fraternization have been of concern as well as gender bias (Gooch, 2001). Gooch's research was a thorough literature review of life in the Navy for both men and women following a large push for integration of genders. Although integration has been called a success it is necessary to eliminate gender bias to fully unlock the potential of our military. Research indicates gender bias within the traditionally male dominated organization. Any bias can be detrimental to the integration of women in to roles they are capable of performing r outperforming men in. This will help optimize the military. This is relevant to my research in understanding the dilemma and inhibition behind letting women in combat based front line specialties. Bias not just based on quantifiable physical performance, but ignorance as well. In a study conducted at the U.S. Air Force academy, 41 cadets were surveyed to help understand attitudes on gender and training. 78% of the respondents were male and 22% female. A covert eight item instrument known as the (Modern Sexism Scale) was built into the questionnaires to find bias. 69% of men who participated in the test responded negatively through the MSS towards their thoughts on women and fitness (Do, Samuels, Adkins, Clinard, & Koveleskie, 2013). It should be noted that this test was only done at the Air Force academy and does not fully represent all U.S. Military branches. This research more evidence of a negative attitude of servicemen towards their female counterparts relative to physical performance. Test conducted covertly. Relative to my research once again in understanding the barriers that have and continue to exist for women moving towards combat roles. My research could help break

some of these inhibitions by gathering data on female performance. The data from my research could potentially be translated because it relative to overall fitness and occupational relevance. Load carriage of some nature is often a major component of combat fitness testing.

A controversial topic relating to gender bias has been the role of women in combat. Until recently, women have not been allowed in combat roles in the U.S. Military. A review of literature has shown that many other countries do allow women in combat roles, but women still make up a small percentage (Barry, 2013). Israelis women are required to serve in the military along with men have been offered combat roles, but they include only 3% of combat soldiers. In France, women make up 19% of the army, but only 1.7% of infantry where they are allowed to join. Canada's women in combat forces are only 3.8%, New Zealand has 2%, and Norway below 1%. It is thought these numbers remain so low because of the rigorous physical tests required. Not all combat roles have opened up to women in the U.S., but with women integrating into these roles comes the question of combat fitness testing. Barry's research shows the amount of women in combat type positions of other countries. Although these positions are open in these countries a small percentage of women make up the actual occupational percentage. Research suggest this may be due to the fitness levels/testing required to qualify. This is an indicator that that even if women are allowed in US, the percentage may be very low. Raises the question, should physical training be as rigorous as it is? Relevant to my research in considering using the data as a source for future combat fitness studies. Will quantify how a military demographic group of women perform under load carriage testing.

A recent topic in research and news is the ability of women to perform certain exercises as compared to their male counterparts. One exercise in question is the pull up which was

brought to the forefront in the revised and then reversed fitness testing policy for female Marines. The reversal was due to a significant failure rate during recruit training for females at boot camp. There is contradicting research that shows women perform better than men in pull ups with training (Johnson et al 2009). In this study 58 college students 35 men and 23 women participated in a study to investigate various performance differences between men and women, who are familiar with strength training, on both lateral pulls and pull ups. For the pull up repetitions portion of the test, women out performed men with a ratio of M/W = .77. Men performed 8.1(+/- 1.9) pull ups on average while women performed 10.5 (+/- 2.2) on average. Marine Core instituted a pull up requirement in fitness testing and then removed it due to the overwhelming rate of female failures. It was incorrectly concluded that females did not have the strength to perform pull ups. This paper indicates that was an incorrect assumption where women who regularly trained doing pull ups outperformed their male counterparts. Relative to research because it shows misconceptions of female performance when it comes to fitness testing. What other exercises need evaluated? My research will provide additional data for exercises that could be used for combat fitness research.

The group of literature reviewed concludes gender bias can exist even with occupational fitness testing and that there is a general bias towards females in combat fitness roles. Additionally, it determines the inability of some occupational relevance tests to correctly judge fitness levels. Some exercises that were thought to be very difficult for females to perform, such as passing the pull up requirements in the Marines' CFT, have been proven inverse when comparing women and men that have acclimated to the exercise over time. The literature displays the overall need for further testing in the realm of combat fitness. It has contributed to this research through helping identify the weak points of knowledge and highlighting some pitfalls of

testing design which is used in my methodology. This group of work also brings about the possibility of incorporating my proposed testing methodology into future related research for such testing scenarios.

2.6 Future Work and Ongoing Research in This Area

Fitness remains a very important part of the military's ability to perform its missions. The PRT is designed to evaluate service men and women's physical fitness levels and should be a fair test as it is weighted in career promotions. There is moderate current research involving the fitness tests of the military, but not enough examining the bias within. There is a need for research to further investigate and empirically quantify the bias within each force's fitness tests. There is a minor amount of research that involves making changes to the military fitness testing programs through theoretically based improvements, but these need to be actually validated in ongoing studies. Almost all of the knowledge available pertaining to military fitness testing includes data on men only. A common theme within the literature review indicates future research should focus on gathering data on women. Fortunately, there is a trend of research increasing in this area due to the controversy of women's physical ability in combat. Although this will be a great benefit in gaining some data for military fitness, it will likely focus on a small percentage of personnel specific to combat roles. There is a greater need to research the more common physical fitness test and scoring systems of the military that all members must participate in. Most of the work concentrates on the actual ability of women to perform specific exercises devised for occupational specialty testing. For example, the ability of females to lift ammo cans repetitively or perform pull ups. The demographics of the military will continue to change over time and a continuous stream of research will only benefit the US military and

hopefully continue, even if it is born through controversial news headlines due to the strive for gender equality.

CHAPTER THREE: METHODOLOGY

3.1 Research Hypothesis

In order for the military to successfully and efficiently defend the nation against all enemies, it must be in mission ready status at any given moment. The components of each military branch must be functioning completely and in harmony. In today's technologically advanced force, it is sometimes easy to forget the most important asset of the US military; its people. Military service members are absolutely essential and it is in the best interest of the military to retain and promote certain individuals who have shown good service. The process in which service members are promoted is very competitive and almost all aspects of a member's records are examined and weighted for promotion. The PRTs are an important part of this process and all individuals should be evaluated fairly amongst the branches. Furthermore, it is important for commanding officers and leaders to accurately know the overall physical operational readiness of its men and women. A thorough understanding of the forces physical fitness must be known in order to guide the entire health and wellness programs in the most beneficial directions. Without an accurate system of measurement, there is no way to truly understand fitness levels, trends in wellness, and operational readiness as it pertains to the fitness component. Fairness and equality to all service members also plays a role in measuring for the fitness tests. Each member, regardless of differentiating characteristics, should be scored fairly and without bias. The bias in question is of course not intentional, but a consequence of a fitness program that has remained primarily unchanged throughout decades. Technology, new understanding of fitness level measurements, military demographics, bio-mechanics research, and changes in military occupational specialties throughout the past decades are several

compelling reasons for reviewing the physical fitness tests. Past research has shown a theoretical bias in the current PRT towards members who have a higher mass; calculated up to 20% in some scenarios. The objective of this research is to determine if load carriage is a viable option to eliminate mass bias for female service members in fitness testing. Prior load carriage research conducted on male cadets showed that adding 30 pounds of weight (in a back pack) to the run and push- up parts of the PRT effectively eliminated the mass bias for heavier individuals. This research hypothesis is that adding weight to all PRT events will also effectively eliminate the mass bias for females. Additionally, a different vehicle for loading the weight (weight vests) will allow for performing the curl-ups portion of the fitness test and increase the overall scoring accuracy.

Problem Statement/Research Question:

Does adding load carriage to all the separate components of the Navy Physical Fitness Testing eliminate body mass bias for women in each event?

Ho: The correlation coefficient is NOT significantly different from 0 (for each PRT event).

Ha: The correlation coefficient IS significantly different from 0 (for each PRT event).

3.2 Proposed Subjects

A comparison of PRT events has been done (loaded carriage versus unloaded) to identify any statistically significant variance in existing bias. The scope of the research was limited to women due to the fact that there is very little data available on their performance in military

physical fitness testing (gap of knowledge). Also, mass bias could have a greater effect on them due to greater scoring system variances.

The pool of voluntary personnel has been taken from multiple military and non-military affiliated organizations including the Veterans Affairs Office, UCF ROTC candidates of the Army/Air Force, armed forces recruiting centers, Student Veteran Association (SVA), and local (Florida) military bases. Additional individuals were gathered from outside military affiliated sources such local fitness clubs and centers. The research was limited to the testing methods of one military branch (Navy) because of resource constraints and the fact that all military branches have very similar PRT testing with exception to the Marine's combat fitness test. The Marine's combat fitness test is more akin to an occupational specialty test like those of infantry divisions, Special Forces, or pilots. It is not necessarily intended to gauge overall wellness, but performance under combat stress scenarios. There was also a scope limitation on the number of women tested in this research. A sufficient amount of individuals have been tested to ensure statistical accuracy (35), but it was restricted to within limits of resource constraints. A priori power analysis determines a sample size of 34 yields a power of 80%. It should be noted that there is a significant amount of research being conducted on the combat fitness test due to the recent interest of women in combat roles. Although relevant, this research is not focused on that specialized testing, but is more concerned with the services overall fitness test which is represented well with the Navy's method. This testing is relevant as it may mitigate bias within any type of military fitness test which uses non-weight biased events.

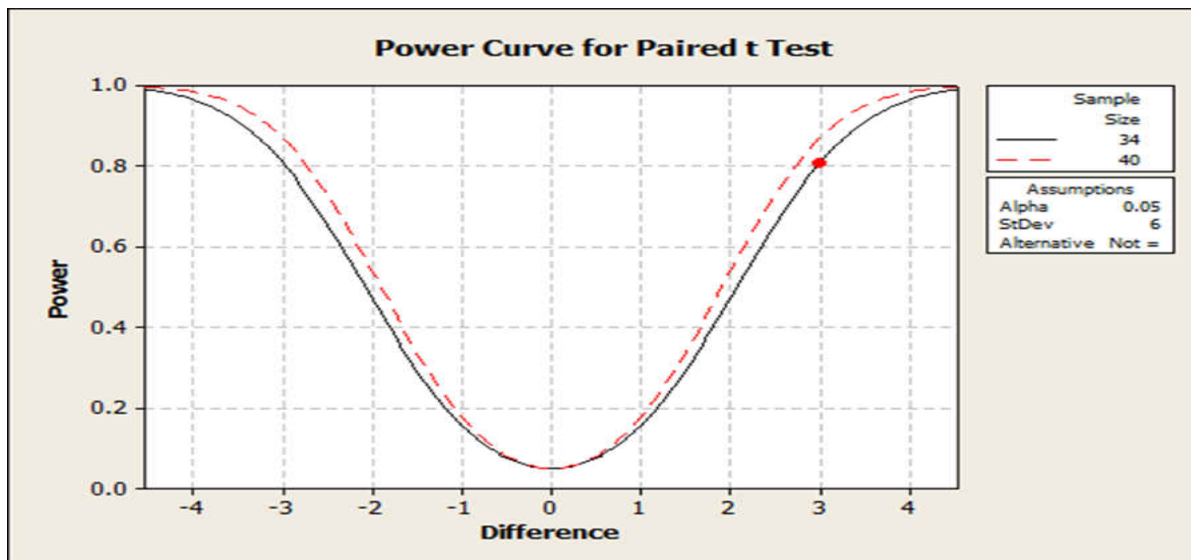


Figure 6: Power Curve for Sample Size Selection

Identification of the proper candidates is critical to the research. The candidates asked to participate must have the ability to perform the tasks with safety in mind foremost. All candidates were screened with a risk assessment questionnaire similar to the one the Navy uses prior to its PRT in addition to all required IRB (Institutional Review Board) and University assessments.

Physical Readiness Test Risk Assessment Questionnaire

(1) Do you have a current PHA? If no, you may not participate today.

(2) Do you have chest pain (with or without exertion), bone or joint pain, high blood pressure or high cholesterol? If yes, have you been cleared, by your medical provider, to participate in PT?

(3) Have you had a change in your medical status since the last time you were asked these questions?

(4) Are you ill today or know of any medical condition that may prevent you from participating in physical activity today?

(5) (For PRT Only) Did you answer yes to any questions? If yes, do you have a PFA medical waiver or clearance form on file? If no you may not participate today?

Signature: _____ Date: _____

Figure 7: PRT Health Risk Assessment Questionnaire

The candidates selected were able to participate in both tests and load carriage acclimation exercises over a given period of time so that all the data could be compared and evaluated correctly. This was imperative because individual performance was compared between all testing

and cannot be analyzed if any of the tests were missing for an individual. Individual personal schedules were considered in selection to include work, class, and family responsibility loads over the testing periods.

Past research has shown that even a minimal amount of time away from required fitness standards starts affecting the service member's performance (Krebsbach, et al, 2012). It is also shown that active duty members tend to be fitter than the average civilian population and this must be taken into consideration (Nolte et al, 2002). The persons asked to perform this testing should be in a similar fitness condition of an active duty sailor. A short physical fitness screening was given to select candidates to estimate their current levels of overall fitness. Participants were asked to stretch prior to exercises and a summary of objectives was given. Resting heart rate was taken for each selected participant. Each person was asked to arm cycle using a portable body cycle for a short period of time in order to raise their heart rate to 70% of max $(220 - \text{age})$ and then given 80 seconds to rest. Heart rates were monitored with the Polar Electro E600 heart rate monitors and equipped with chest straps to discover whose heart rates return to near resting over 80 seconds. This test will be used as a rough indicator of current fitness.

1. Calculate 70% Max Heart Rate for start time $\sim t_0$

$$(220 - \text{Participants Age}) \times .70 = \text{HR}_0$$

2. Calculate decrease % in heart rate

$$(\text{HR}_0 - \text{HR}_{80}) / \text{HR}_0 = \% \text{ decrease}$$

**Time will be considered = 0 at the 70% of heart rate mark and the clock will start concurrently with exercise stopping. The participants heart rate will be recorded at time equals 80seconds (HR80) of resting. The percent decrease in heart rate will be evaluated and recorded.

3.3 Data Collection Environment

Test data was collected through the same process required by OPNAVINST 6010J Guide 5. Conditioning and test safety was followed per the required instruction including operational risk management and the correct environmental conditions. The testing evolution was conducted with at least one trained Cardio Pulmonary Resuscitation (CPR)-certified monitor and at least one assistant per every 25 individuals being tested. A minimum of two monitors were present for each test. An emergency assistance plan was in place which included telephone numbers and procedures for emergency situations.

The test site was a level pre-measured 1.5 mile course along with sufficient flat ground to perform the static events. Course distance was measured with walking wheel and handheld GPS to verify accuracy. Weather was monitored for safety and observed prior to start of testing. The observation includes forecasted weather conditions prior to events and monitoring during testing as well. The PRT was not to be conducted under harsh environmental conditions. As required, the test was not be performed outdoors when wind chill is 20 F or lower, or when “black flag” conditions exist (wet bulb globe temperature (WBGT) of 90 F or higher. No hazards to subjects from traffic or other obstacles occurred.

Data and information was recorded on the Pretest Fitness Indicator sheet (APPENDIX C) and used along with other measurements and self-evaluations (APPENDIX D) as an indicator of current overall fitness for participant selection. It was important for participants not to be at physical risk and fall within current Navy demographics for fitness. The pretest were not used to take the best results in heart rate scores, but used to validate selected individuals who are above average fitness levels and verify extenuating safety concerns do not exist.

3.4 Data Collected

It was desirable to start with a large pool of volunteer candidates for the research and then narrow down the participants to a number which was statistically acceptable (35-40 at .80 Power Level) given limited resources and other selection constraints. These included safety, availability to complete all testing, current physical fitness condition and matching Navy demographics. At the same time, it was important not limit and narrow the window of fitness types or the bias may not be observable. There must be some variance in mass for the testing. The following measurements and data were recorded and critical to research analysis.

- Body characteristics including height, weight, age, and BMI.

Initial body composition and weight was recorded for candidates this was completed with an Omron HBF-516 Full Body Composition Monitor and traditional height/weight charts utilized by the Navy. The impedance meter was utilized in order to go through a larger pool of candidates without performing multiple tape measure calculations as well as verify height/weight charts. A different method of body composition analysis was used after for candidates who did not meet regulatory requirements (measuring tape). The measurements taken with the tape were input into the designated “female”, equation for calculation.

Females:

$$\% \text{ body fat} = 163.205 \times \text{Log}_{10} (\text{waist circ.} + \text{hip circ.} - \text{neck circ.}) - 97.684 \times \text{Log}_{10} (\text{height}) - 78.387$$

($N = 202$, $R = 0.856$, $SEE = 3.61$ % fat)

- Pre fitness test data which includes heart rate monitor of resting heart rate and heart rate after arm cycling test completion.

Participants were asked to stretch prior to exercises and a summary of objectives given.

Resting heart rate will be taken. Each person will be asked to arm cycle using a portable body cycle for a short period of time in order to raise their heart rate to 70% of max (220-age) and then given 80 seconds to rest. Heart rates were monitored with the Polar Electro E600 heart rate monitors equipped with chest straps to discover whose heart rates return to near resting over 80 seconds. This test will be used as a general indicator of current fitness.

- PRT unloaded test raw scores (Distance Run, 2 Min Max Curl-Ups, 2 Min Max Push-Ups)

The PRT was performed in the unloaded condition prior to weight vests being used. The test performed were conducted per OPNAVINST 6010 instruction. Scoring was recorded manually on sheets designed to accurately, quickly, and simply record the results (APPENDIX E).

Distance run times were measured with standard sports stop watch counters and associated candidates identified with a bib pin numbering system such as those used in marathons. Curl-Ups and Push-Ups were recorded manually by paired test partners. These are the same methods of counting used per the Navy instructions. All scores were recorded on the same raw score sheets.

- PRT loaded carriage test raw scores (Distance Run, 2 Min Max Curl-Ups, 2 Min Max Push-Ups)

The PRT was performed in the loaded condition (weight vests) only after an acclimation period (conducted prior to any recorded results taken). The test was performed per OPNAVINST

6010J instruction. Scoring was recorded manually on sheets designed to accurately, quickly, and simply record the results. Distance run times were measured with standard sports stop watch counters and associated candidates identified with a bib pin numbering system such as those used in marathons. Curl-Ups and Push-Ups are to be recorded manually by paired test partners. These are the same methods of counting used per the Navy instructions. All scores are to be recorded on the same raw score sheets.



BMI Measurement Points



Weight Vests



Myotape Body Measurement Tape



Omron HBF-516 Full Body
Composition Monitor



Portable Body Cycle



Polar Electro E600 heart rate monitors

Figure 8: Tools used for Research & Data Collection

3.5 Experimental Design and Statistical Analysis Approach

The research analysis was approached through gathering a representative pool of individuals who accurately depict the current demographics of women service members. The pool of voluntary personnel was taken from multiple military affiliated organizations including the Veterans Affairs Office, UCF ROTC candidates of the Army/Air Force, armed forces recruiting centers, Student Veteran Association (SVA), and local (Florida) military bases. Additional individuals were gathered from outside military affiliated sources such as fitness clubs and gyms.

Volunteer's needed to meet all pertinent safety requirements as directed by the University and standard military fitness testing protocol. This involved several risk and health assessment questionnaires prior to any testing. Once it was determined safe testing can was performed. Selected subjects were asked or volunteered to perform a short fitness assessment including heart rate monitor during exercise and all participants had their body mass index recorded and took a written fitness survey. This information was used to make sure the candidates selected were within the desired matching fitness levels US Military majority. Slightly above average fitness levels than general population. The selected individuals were asked to perform the current PRT test per Navy instruction OPNAVINST6010.x and performance data was recorded. A second test was administered under the same instruction, but with a specified load to be carried by each individual during the event (20 lbs.). Full test procedures are provided in Appendix E. There was a rest period between testing for safety and so that exhaustion did not interfere with test data as well. The data was compiled and then validated through comparison with general summary statistics of like demographic groups (Navy Fitness Scoring Scales), documented allometric

scaling research, and analyzed for correlation to detect bias. The data gathered from the two separate tests was then analyzed with Minitab statistical software and evaluated through Hypothesis testing, Paired T-test, and Pearson Product Moment Correlations. Results and recommendations for the loaded carriage test as mitigation of mass bias will be given. This research has been discussed with a subject matter expert, Dr. Paul Vanderburgh, who agrees the methodology outlined for testing with weighted vests should answer the problem statement/hypothesis devised. It was also noted that there is a need for this work due to the current gap in knowledge in female PRT fitness data.

Research Methodology Summarized Steps

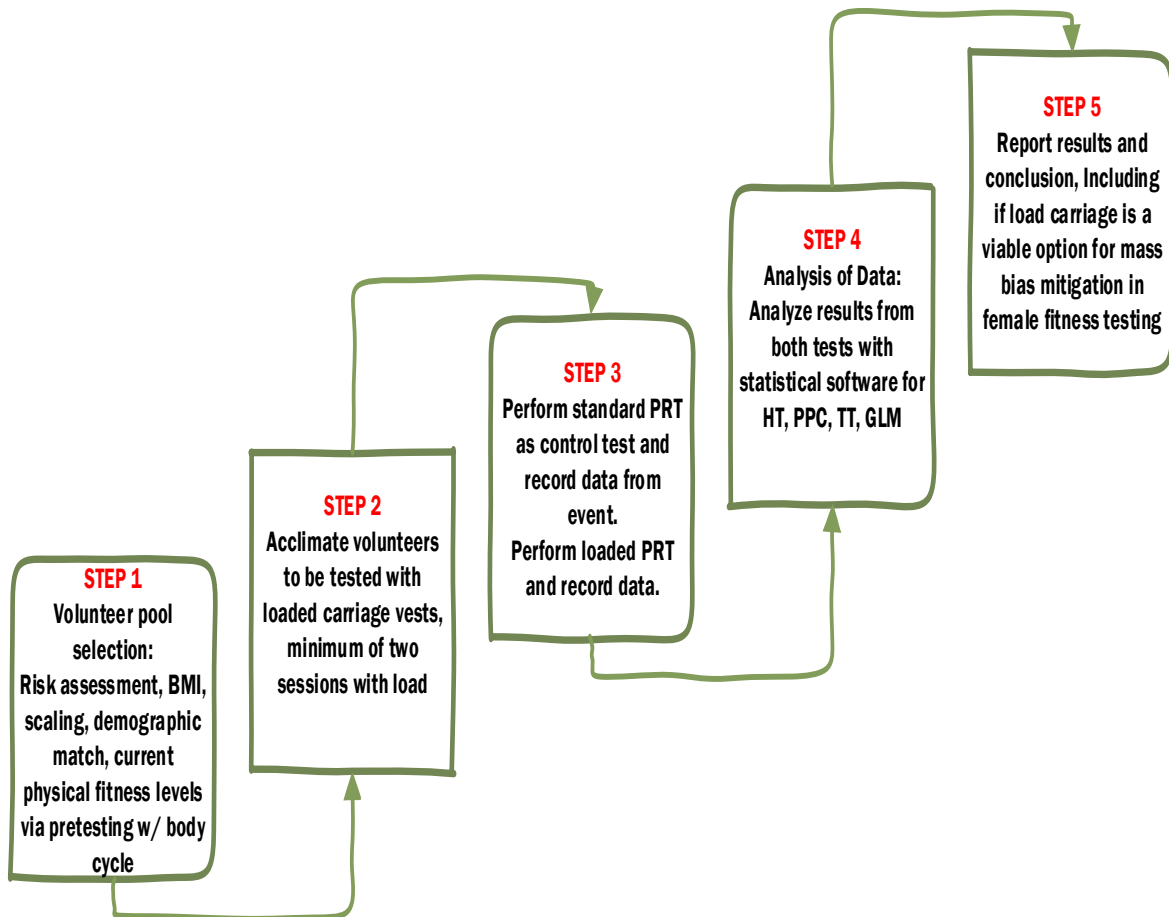


Figure 9: Proposed Methodology Flow

CHAPTER FOUR: FINDINGS

The research data showed a significant decrease in bias when utilizing the weight vests (loaded carriage). This phenomenon was observed in all components of the Physical Readiness Test at different degrees. The testing has confirmed loaded carriage tests do indeed mitigate at least part of the bias seen in the traditional fitness testing due to mass of the individual. To my knowledge there is no comparative study for either male or female related to this measurement for the curl ups portion of testing. The push-ups and timed run components of the fitness test have not been tested for females, but the observations and conclusions made are analogous to prior testing performed on males (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011). I have concluded the use of load carriage in the PRT would be a viable and realistic way to make the test fairer for all member's required to participate. The vests are not as cumbersome as other exercise equipment such as stationary bikes or weight benches and make them more logistically reasonable to use. This is especially true for members deployed to locations around the world or stationed on smaller vessels that would limit space. Acclimation to load carriage could pose a problem and took longer than expected during my testing (3-4 versus 2 session). In order for commanders to implement load carriage they should allow sufficient time for service men and women to get used to the equipment. Additionally, I found the traditional scoring tables used for calculating the overall participant PRT scores to be incapable of use for the loaded test. The testing results were understandably lowered due to the load added during the exercises. This resulted in a high failure rate under the current scoring system. Additional work would need to be performed in order to determine a fare revised scale for the loaded PRT.

4.1 Research Problem Statement Review

The objective of this study is to determine whether adding weight to individuals while they are performing the Navy's Physical Readiness Test would make a statistically significant change in the outcome relative to an existing bias due to mass. The research question is written in the following hypothesis statement and analyzed through Paired T-Tests and Pearson Product Moment Correlation.

Does adding load carriage to all the separate components of the Navy Physical Fitness Testing eliminate body mass bias for women in each event?

Ho: The correlation coefficient IS NOT significantly different from 0 (for each PRT event).

Ha: The correlation coefficient IS significantly different from 0 (for each PRT event).

4.2 Research Participant Characteristics

Research participant characteristics are important to observe in that we want to mirror, to the best extent possible, the current demographic of the U.S. Armed Forces. Because this study concentrates on women, we are not concerned with the gender make up, but have concentrated on ethnicity, age, and weight/BMI of the research participants.

4.2.1 Descriptive Statistics of Physical Characteristics

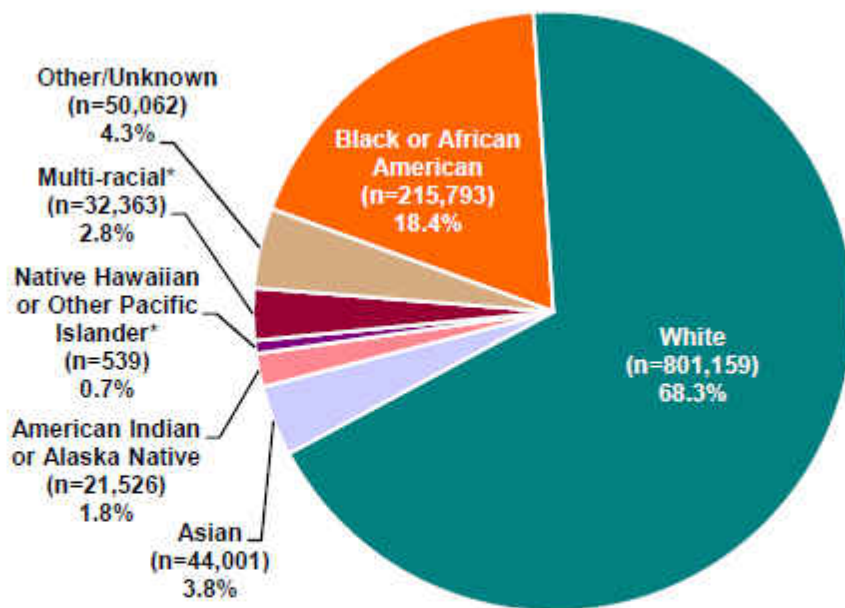
The overall descriptive statistics of the research participants are shown in Table 7. The total number of participants who completed all phases of the study was 35. Measurements for height, weight, impedance, and BMI were recorded for further analysis relative to performance.

Table 6: Descriptive Statistics for Participant Physical Characteristics

Variable	N	Mean	SE Mean	St Dev	Minimum	Q1	Median	Q3	Maximum
Age	35	26.971	0.800	4.731	18.000	24.000	26.000	30.000	38.000
Height	35	63.200	0.670	3.962	56.000	59.000	63.000	66.000	71.000
Weight	35	152.52	2.76	16.32	124.20	139.32	154.44	162.00	183.00
BMI	35	26.911	0.470	2.781	22.381	24.893	26.507	28.933	32.716
Impedance	35	27.201	0.492	2.912	22.932	25.178	26.329	29.577	33.073

The target ethnic makeup of the participant sample is 68% white, 18% black or African American, 3.8% Asian, 2.8% multiracial, 1.8% Alaskan native or American Indian, and .7% native Hawaiian or Pacific Islander. The research study's actual sample ethnic make-up was 48.2% white, 34.8% black or African American, 1.4% Asian, 7.5% multiracial, 0% Alaskan native or American Indian, and 0% native Hawaiian or Pacific Islander. Although this is outside the target goal, it is acceptable when the BMI is brought into view. The main reason for targeting ethnicity was because it has played a factor in the past when calculating BMI and could make the PRT standards unfair. The maximum allowable BMI for the Navy is 33% for females and all participants fall below this requirement. Additionally, height and weight BMI table calculations were backed up by electrical impedance measurements to make sure they were in alignment. A T-Test for BMI height and weight calculations versus electrical impedance yielded results showing no statistical difference between means $\mu = 27.2$, $\sigma = 2.91$ Impedance and 26.91 $\sigma = 2.78$ respectively with P Value of 0.671. Therefore we can conclude that both test measurements are in alignment. Since we have no participant failures on BMI standards it is unnecessary to drill down into ethnic diversity of the sample and the test remains valid. The test has a range of

calculated BMI from 22.4% to 32.7%. Recent studies have shown an increase in BMI with the overweight population being 12.7% for females (Reyes-Guzman, C. M. 2015). The target for 35 total candidates who completed all phases of testing would be 5 participants who are considered overweight by PRT standards. As previously stated, there are no individuals who failed the BMI calculation and 1 individual over on the impedance measurement. This is not within the target range, but 4 individuals are at the threshold of the BMI standard at 32-33%. The average BMI for the group was 26.911, $\sigma = 2.781$.



Note: Percentages may not total to 100 due to rounding.

DMDC Active Duty Military Personnel Master File (September 2011)

Figure 10: Ethnic Make Up of U.S. Active Duty Military (DMDC 2011)

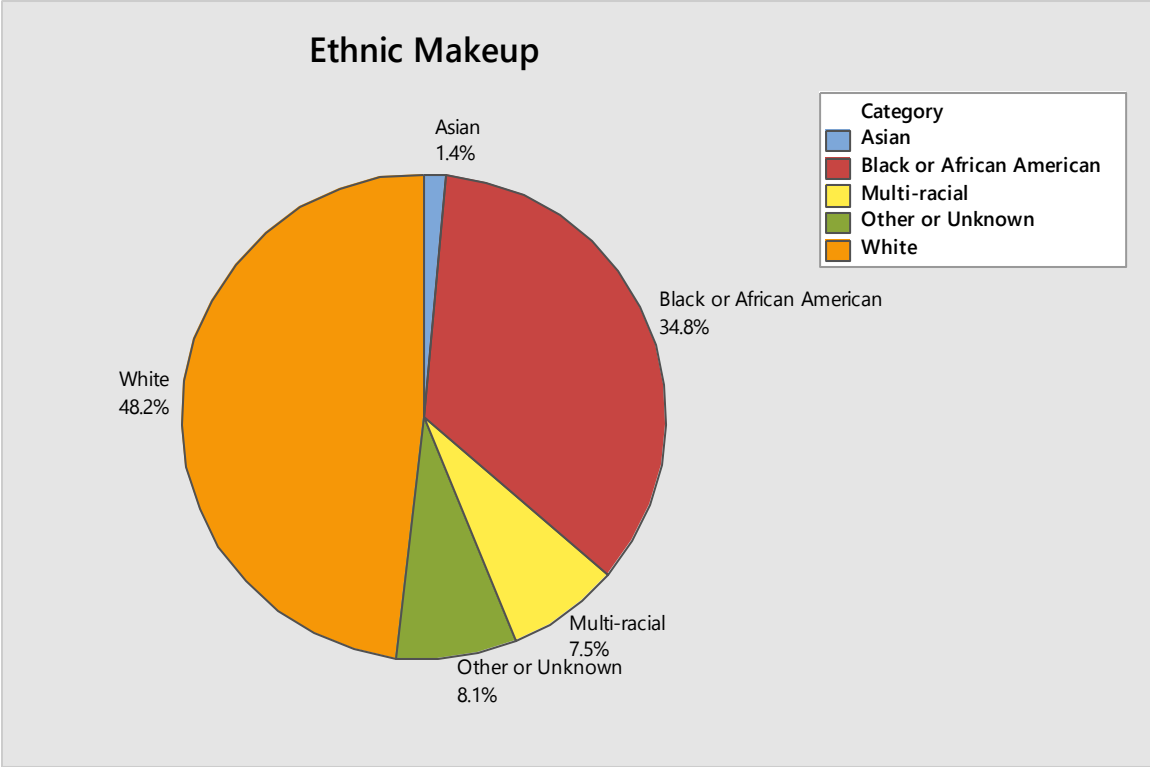
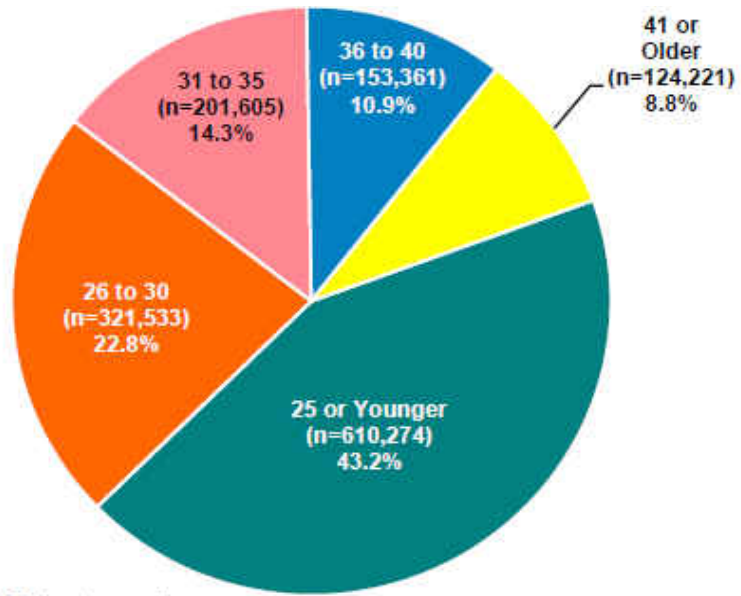


Figure 11: Chart of research Participant Make up

The target age groups for this research study break down as shown in Figure 12 with a majority (43%) of the participants desired to be 25 or younger at the time of the study. The next largest group is to be 26-30 year old females (23%). This is followed by 31-35 year old group at roughly 14%, 36-40 year old at 11%, and 41 and older at 8%. The largest differences in age demographic from the targeted were in the 26-30 age group with 17.2% difference and the 41 and older group with an 8.8% difference because the group had no individuals over 40 years of age. The largest portion of participants came from the 26-30 age group followed closely by the 25 and younger age group. We are limited by the number of volunteers for this study and future research work can concentrate even further on current demographics.



Note: Percentages may not total to 100 due to rounding.

DMDC Active Duty Military Personnel Master File (September 2011)

Figure 12: Active Duty Personnel Ages (DMDC 2011)

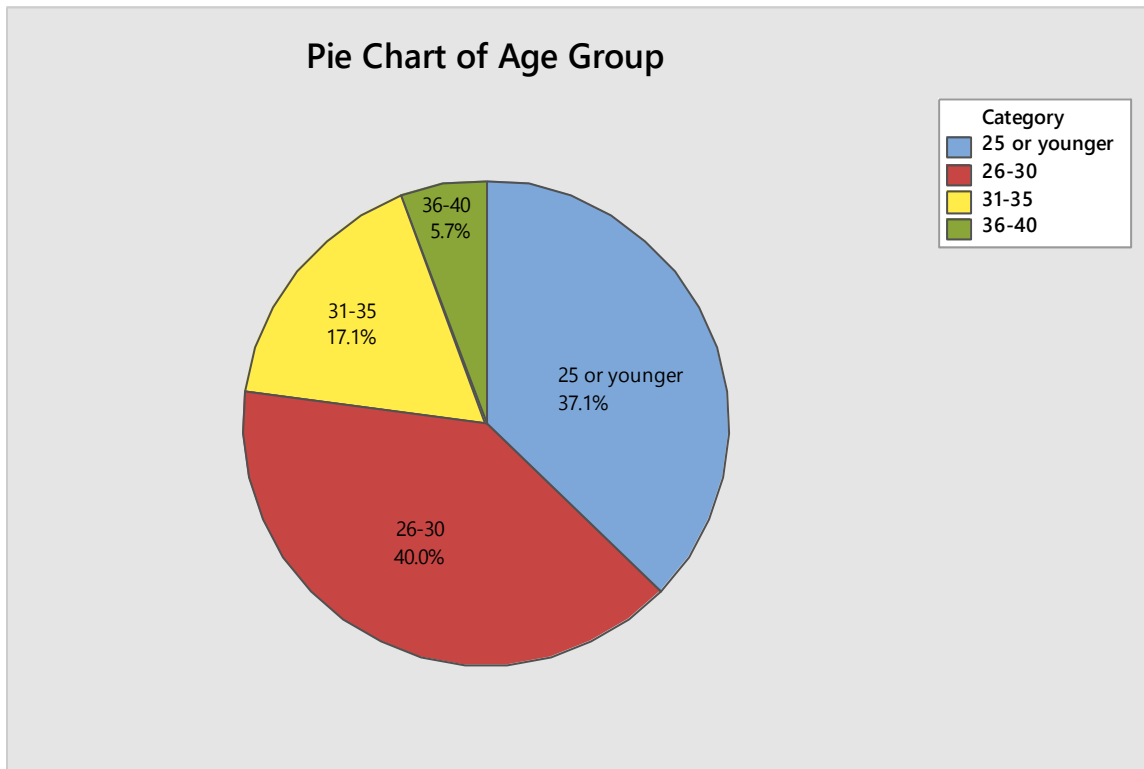


Figure 13: Research Participant Age Group Overview

Participant weight is a critical factor in the study for determining the relationship between it and actual performance for the loaded and unloaded portions of the PRT. The mean weight for all participants was 152 lbs, $\sigma = 16.32$ which is lower than the U.S. national average of 166.2. The weights recorded fell within a normal distribution probability plot with a small Anderson-Darling statistic .337 and a P-Value of .486 which is $>.05$ indicating the data is normally distributed. This distribution assumption is important for further statistical analysis of weight for the study.

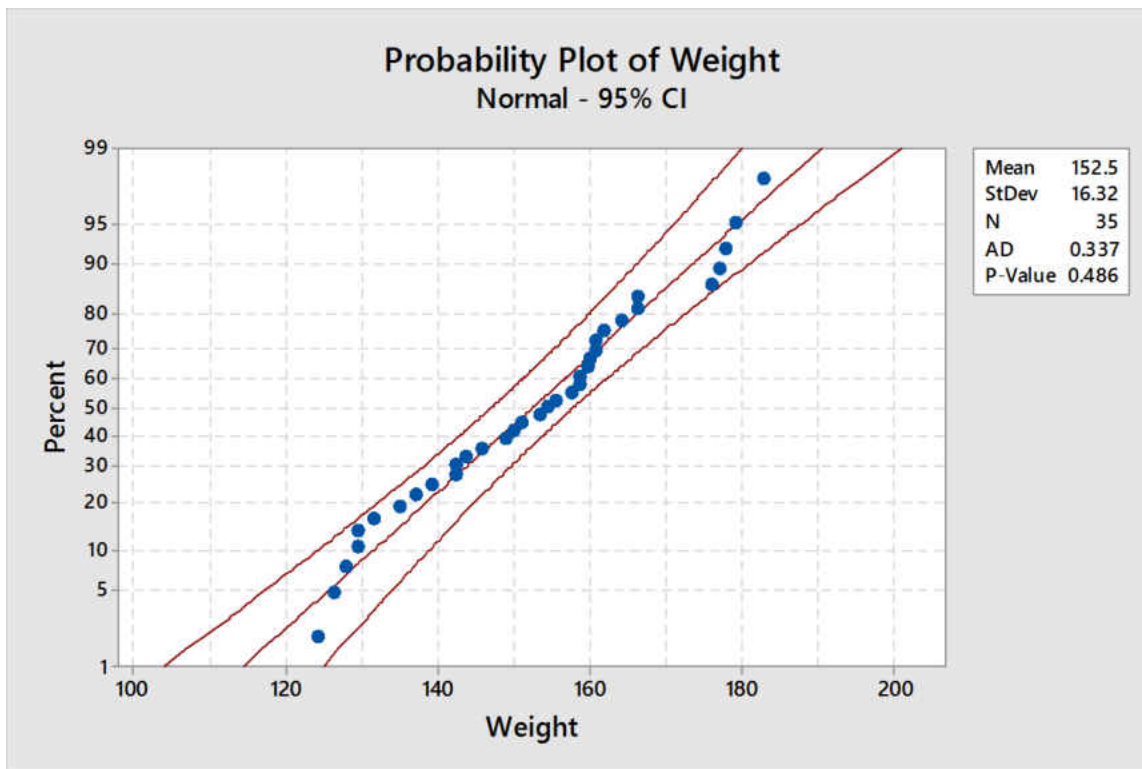


Figure 14: Probability Plot for Participant Weight

4.2.2 Participant Questionnaire Survey Data

A Physical Fitness Self-Assessment Survey (Appendix D) was given to the research participants in order to gauge their weekly workout routines, experience with military fitness testing, and familiarity with weight vests. The assessment yielded results showing 25 of the 35 candidates who completed all phases of the research have never been affiliated with the military or subsequent fitness tests. This required additional explanation of the rules associated with the testing, discussion of the military's purpose for testing, and what would be asked of them. It was found the highest amount of individuals who had been affiliated with the military were veterans of the Navy and familiar with the PRT. This was followed by the Air Force, ROTC, and Army respectively. These individuals still went through a refresher training to understand the requirements of the study, especially the loaded version of the test.

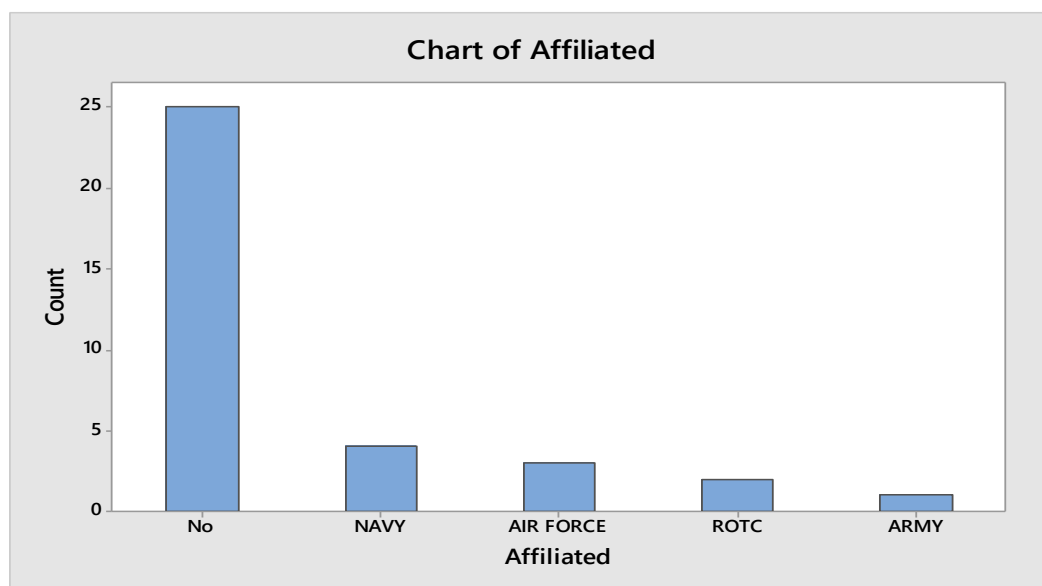


Figure 15: Current or Prior Affiliation of All Participants with the Military

Those individuals who were associated with prior military experience or testing were asked if they had ever failed a military fitness test. One individual, 3% of entire sample, had failed a test

in the past during initial recruit training. All other individuals who participated had never failed a PRT. This represents 10% of individuals who had been affiliated with the military and is consistent with military failure rates.

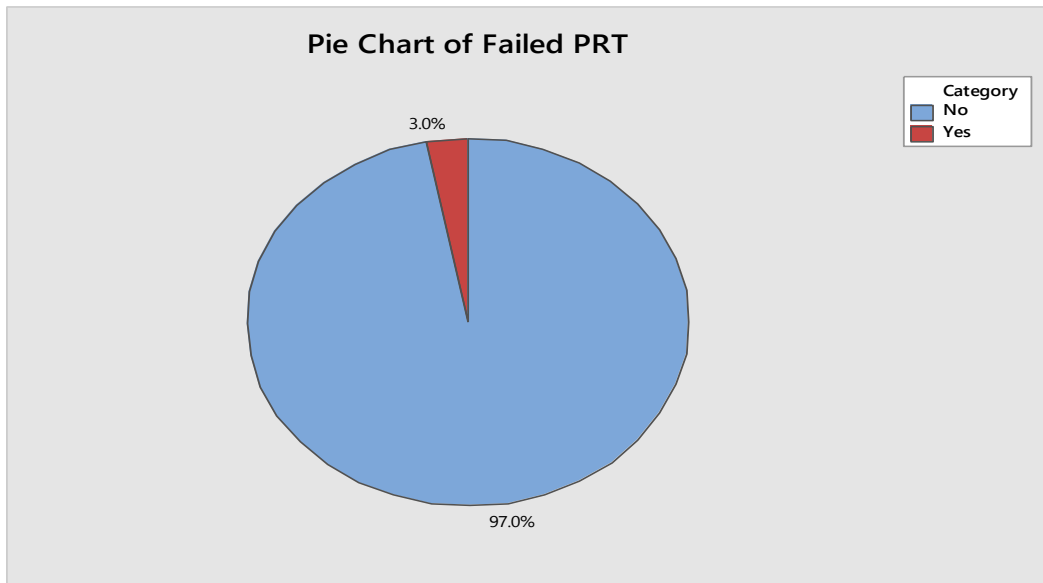


Figure 16: Percent of Participants who have Ever Failed a Past PRT

In order to understand the participant's fitness and exercise habits, the self-assessment asked about their weekly workout routines. The group's weekly time participating in fitness activities varied greatly with many individuals exercising several times per week. It was found the average exercise time in minutes per week for the group was $\mu = 227.3$, $\sigma = 87.02$. This is generally higher than the amount of time for most Naval occupations, specifically ship's force personnel, who undergo Physical Training 2 times per week for 60 min. Although it should be mentioned many service men and women exercise more than the 2 times per week allotted for divisional Physical Training (PT). The method of exercise also varied between individuals with most (9) identifying their exercise routine with running/jogging. This may be due to the fact those who volunteered understood one of the major components of the research study would be

the 1.5 mile timed run which attracted them. Six participants identified with exercises that were not listed on the self-assessment, for example hiking, kayaking, or Cross-Fit. Others participated in utilizing stationary equipment, weight lifting, organized sports, cycling, dance, and push/curl-ups.

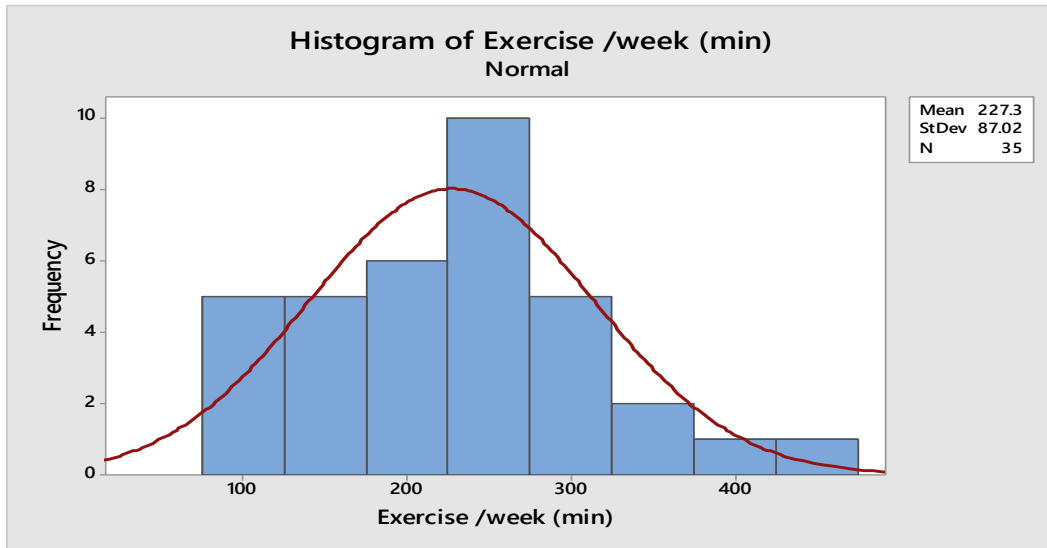


Figure 17: Exercise Time (min) per Week for Research Participants

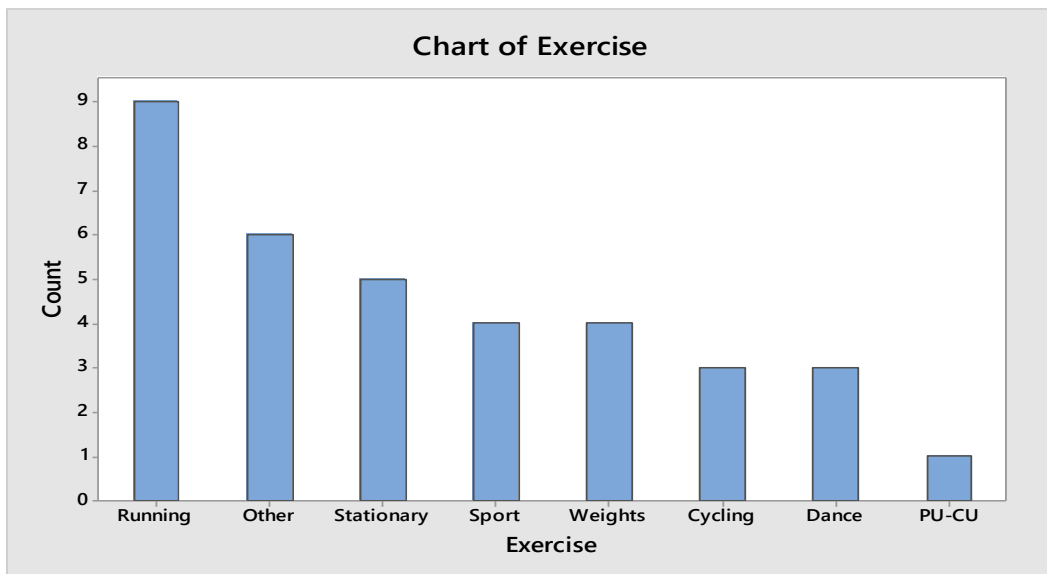


Figure 18: Method of Exercise for All Research Participants

Research participants were asked if they had any experience in exercising with weighted vests in the past. It was found that 8 (22.9%) of research participants had prior experience with weighted vests. A majority of these participants had minimal experience with vests only utilizing them a few times during work out sessions.

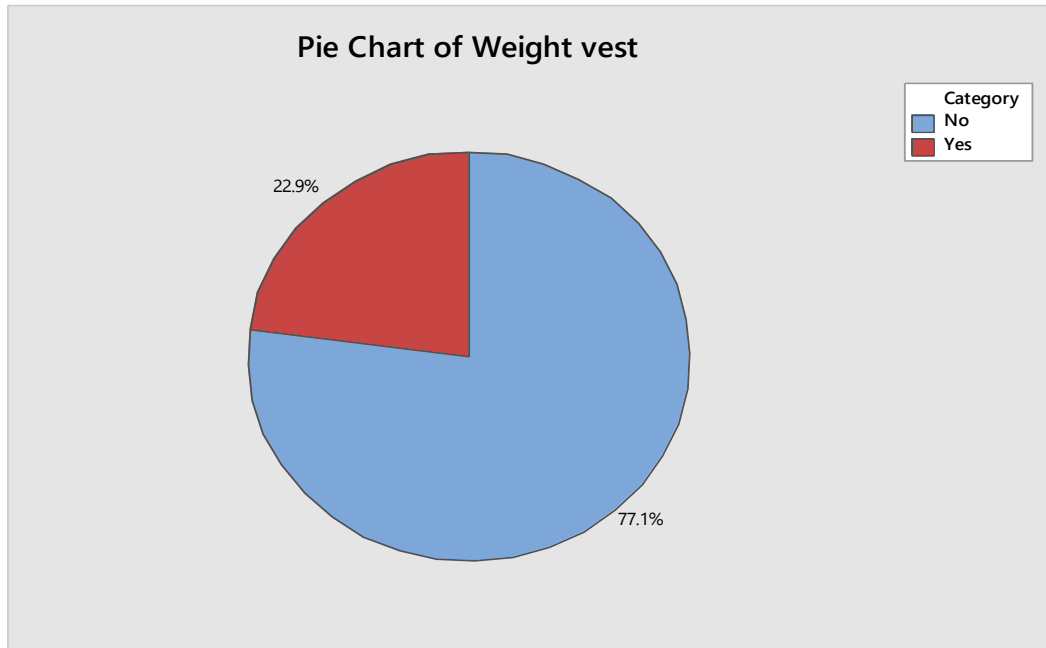


Figure 19: Percentage of Research Participants who had Prior Experience with Weighted Vests

The intention of the self-assessment survey was to gauge participant's fitness levels, work out habits, and prior experience with certain components of the study. This allowed me to estimate a learning curve for acclimation to the loaded portion of the study as well as provided a possible explanation to any outliers found within the data. For example, if a participant routinely used weighted vests for exercise while running and scored abnormally high on the test. It also played a role in safety to assess individuals who were sedentary in fitness.

4.2.3 Disqualification of Participants for the Physical Readiness Test

A total of 8 participants were disqualified from testing because of various reasons. One participant was disqualified because of a pregnancy within the last year. Prior research has shown recent pregnancies can negatively affect military fitness test scores (Weina, S. U. 2006). A majority of those disqualified for the research study (4) completed the first phase of testing, but had limited availability or didn't show up to 2nd assessment. This disqualified them because there could be no comparison between loaded and unloaded test treatments with only one data set. The remaining 3 individuals who were disqualified from the testing were because of a self-perceived inability to complete all phases of the testing. One individual requested to end the test during the first phase (unloaded) and 2 other individuals were uncomfortable with performing exercises with the weighted vest and felt they could not get acclimated. Participants disqualified or opting out of the study represented 19% of the total candidate pool with roughly 5% due to weight vest acclimation. The 5% who were uncomfortable with the weight vest acclimation could pose a problem if this type of test was implemented into the U.S. Military fitness testing. Although this represents a small percentage, all individuals may not feel comfortable with weighted carriage and should have additional unbiased testing options.

4.3 Participant's Physical Readiness Performance Test Results

Research participant Physical Readiness Test result data has been collected for both the loaded and unloaded components of the research. This data was statistically analyzed to answer the research question posed. A table of descriptive statistics (Table 8) is provided to gain an overview of the data sets. Assumptions for normality of performance data was observed to satisfy any requirements for statistical analysis conducted. Thirty-five individuals participated in the study which was found to be an acceptable sample size through power analysis reviewed in section 4.3.2. This was followed by a paired T-Test to verify there was a statistical difference between the treatment means of loaded versus unloaded performance measures. An exploration of and comparison of correlation (weight vs performance) was conducted to specifically answer the research hypothesis that loaded carriage effectively reduces the mass bias found in the PRT.

4.3.1 Descriptive Statistics of Test Results and Probability Distributions

A review of descriptive statistics shows the overall results for all components of PRT including loaded/unloaded push-ups, curl-ups, and the timed distance run. The tables and figures following have been labeled using the term formatted as (exercise 1 as the unloaded test) and (exercise 2 as the loaded test). For example, curl-ups 1 was the initial test given in the unloaded condition without weighted vests while curl-ups 2 labels the second test given in the loaded condition. The unloaded portion of the test resulted in understandably higher score values than the loaded test which was expected and understandable. Only one individual actually performed better on one exercise in the loaded condition (the timed run).

The unloaded push-ups mean repetitions showed $\mu = 22.43$, $\sigma = 10.13$ and loaded showed $\mu = 12.36$, $\sigma = 5.003$. This represents a decrease of 44.9% in overall repetitions for push-ups. In a

prior study conducted with loaded 30 lb back packs instead of 20 lb vests on 56 male cadets, the reduction in push-ups repetition was 32.2% as a comparison (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011). The unloaded push-ups test result for my study $\mu = 22.43$ fits the “Good” performance scale of repetition scoring for the mean age group of 25-29 year olds of 19-37 repetitions. This could be considered the average result in PRT testing when relating to the current scoring system.

The unloaded curl-ups mean repetitions showed $\mu = 63.23$, $\sigma = 21.86$ and the loaded curl-ups showed $\mu = 37.40$, $\sigma = 10.25$. This represents a decrease of 40.9% in overall repetitions for curl-ups by the group. The unloaded curl-ups test result for my study $\mu = 63.23$ fits the “Good” performance scale of repetition scoring for the mean age group of 25-29 year olds of 54-84 repetitions. This could be considered the average result in PRT testing when relating to the current scoring system. Unfortunately, there seems to be no prior loaded curl-up test performed on military fitness for males or females to compare.

The unloaded timed run mean time in seconds showed $\mu = 967$, $\sigma = 192.2$ and loaded showed $\mu = 1287$, $\sigma = 112.8$. This represents an increase of 33.1% in overall completion time for the 1.5 mile run. In a prior study conducted with loaded 30 lb back packs instead of 20 lb vests on 56 male cadets, the increase in the timed run was 18.8% as a comparison (Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. 2011). The unloaded timed run test result for my study $\mu = 967$ fits the “Satisfactory” performance scale of timed run scoring for the mean age group of 25-29 year olds of 960-900 seconds. This could be considered below average result in PRT testing when relating to the current scoring system.

Table 7: Descriptive Statistics for Participant Performance Loaded and Unloaded

Variable	N	N*	Mean	SE Mean	St Dev	Minimum	Q1	Median	Q3	Maximum
Push-ups 1	35	0	22.43	1.71	10.13	6.00	15.00	20.00	29.00	48.00
Push up 2	35	0	12.360	0.846	5.003	2.000	8.000	13.000	15.000	25.000
Curl ups 1	35	0	63.23	3.70	21.86	27.00	45.00	62.00	81.00	105.00
Curl ups 2	35	0	37.40	1.73	10.25	12.00	32.00	37.00	43.00	66.00
run (s)	35	0	967.4	32.5	192.2	641.0	813.0	918.0	1146.0	1438.0
Run 2 (s)	35	0	1287.4	19.1	112.8	1089.0	1212.0	1279.0	1388.0	1474.0

Multiple probability plots were developed to determine the normality of all performance data from the group. This was done to insure all assumptions for data normality for performance data were met prior to further analysis. Although this was not necessarily required for the Paired T-Test due to its “robust” nature related to distributions, assumption of normality is needed for Pearson’s Correlation Testing. A multigraph probability plot is shown in Figure 20 demonstrating the normal distributions according to the Anderson- Darling Statistic and P-Value $>.05$ alpha levels.

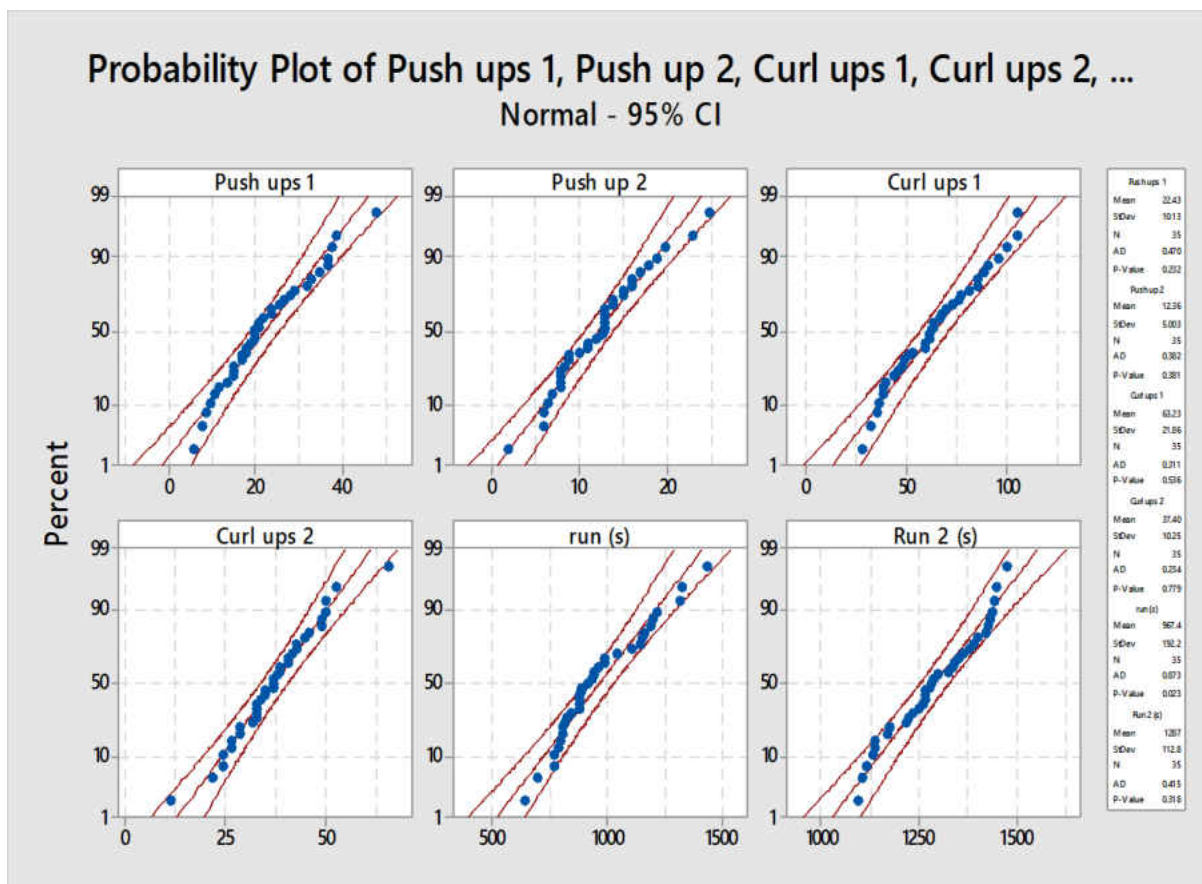


Figure 20: Probability Plot for Participant Performance

4.3.2 Statistical Power and Margin of Error (Desired versus Actual)

An analysis of statistical power was conducted prior to gathering research participants for the study. It was determined a sample size of 34 would be statistically acceptable for an alpha level .05 and a statistical power of .80. This is based upon the overall scoring method of the Navy PRT testing. A target goal of 40 participants was sought, but as discussed several candidates were disqualified from testing. The power and estimated required sample size was also determined for each PRT exercise component. A recommendation sample size of 35 was also made by a subject matter expert (Paul Vanderburgh) for my study.

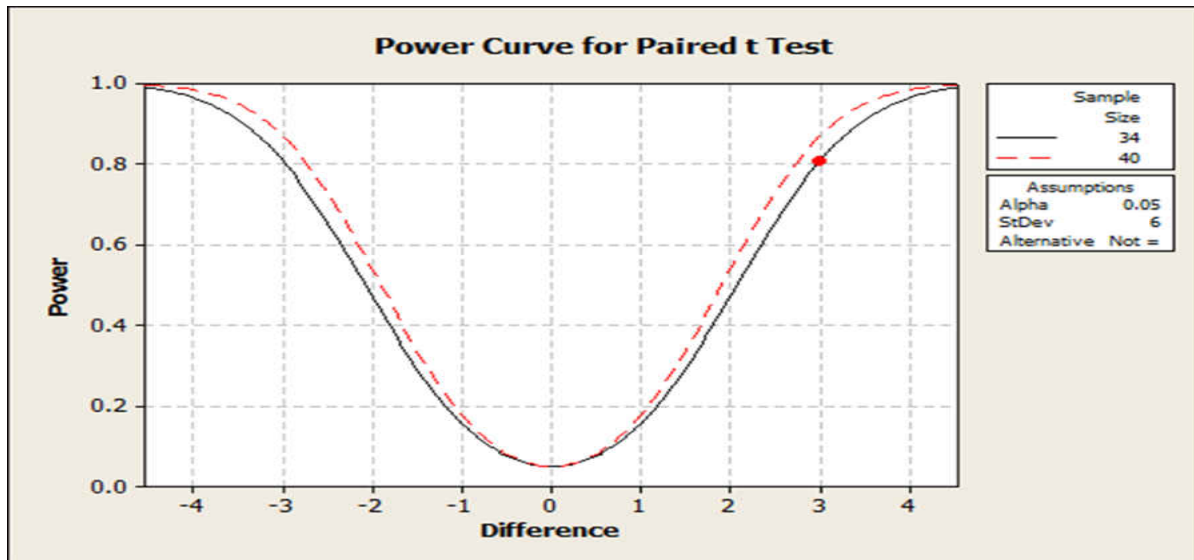


Figure 21: Power Curve for Desired Alpha Level

4.3.3 Power and Error (Push Ups)

It was discovered an acceptable sample size to obtain a statistical power of .8 with an alpha level of .05 for the push-ups component of the research was 9. Total participation sample size for this PRT component was 35 which exceeds recommended sample size by 26. Inputs for this analysis were gathered from the standard deviation of pushups and differences between means for loaded and unloaded performance. A graph is provided for depiction in Figure 22. Actual power with the sample size of 35 was 1.0.

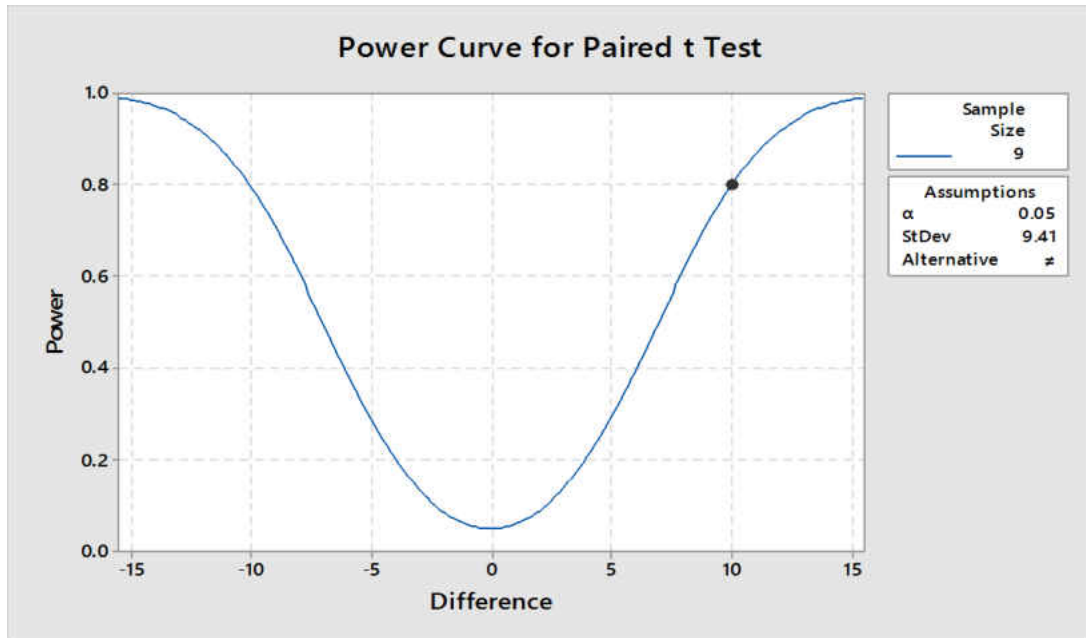


Figure 22: Required Sample Size for .8 Statistical Power in Push- Ups

4.3.4 Statistical Power and Error (Curl Ups)

It was discovered an acceptable sample size to obtain a statistical power of .8 with an alpha level of .05 for the push-ups component of the research was 8. Total participation sample size for this PRT component was 35 which exceeds recommended sample size by 27. Inputs for this analysis were gathered from the standard deviation of pushups and differences between means for loaded and unloaded performance. A graph is provided for depiction in Figure 23. Actual calculated power for the study is .99.

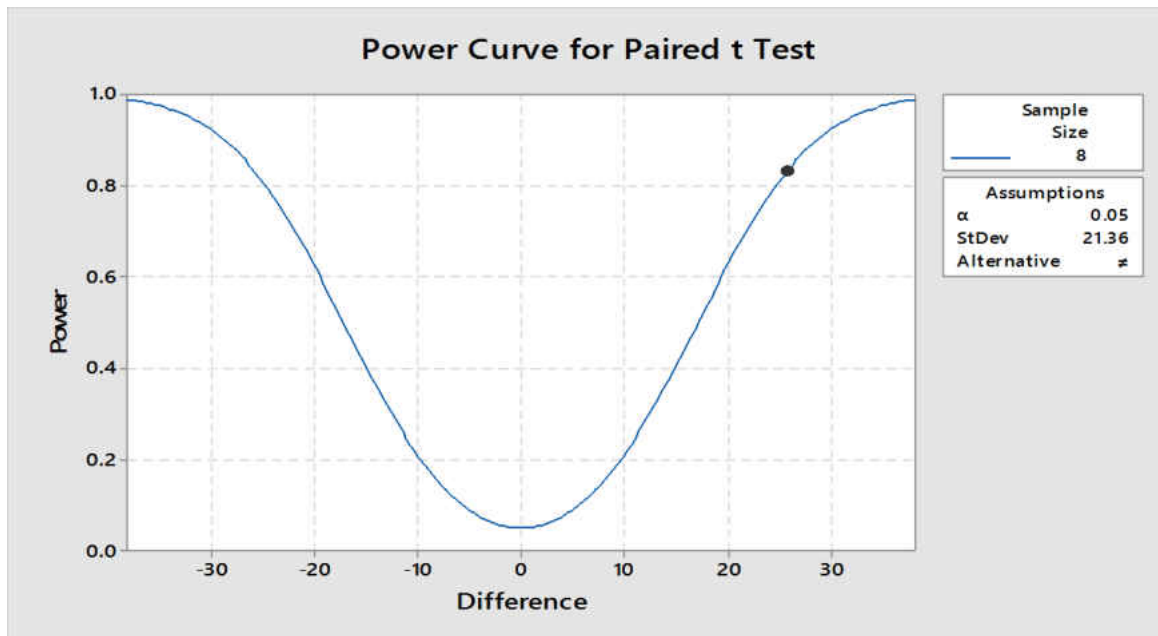


Figure 23: Required Sample Size for .8 Statistical Power in Curl Ups

4.3.5 Power and Error (Timed Run)

It was discovered an acceptable sample size to obtain a statistical power of .8 with an alpha level of .05 for the push-ups component of the research was 7. Total participation sample size for this PRT component was 35 which exceeds recommended sample size by 28. Inputs for this analysis were gathered from the standard deviation of pushups and differences between means for loaded and unloaded performance. A graph is provided for depiction in Figure 24. Actual calculated power for the study is .99.

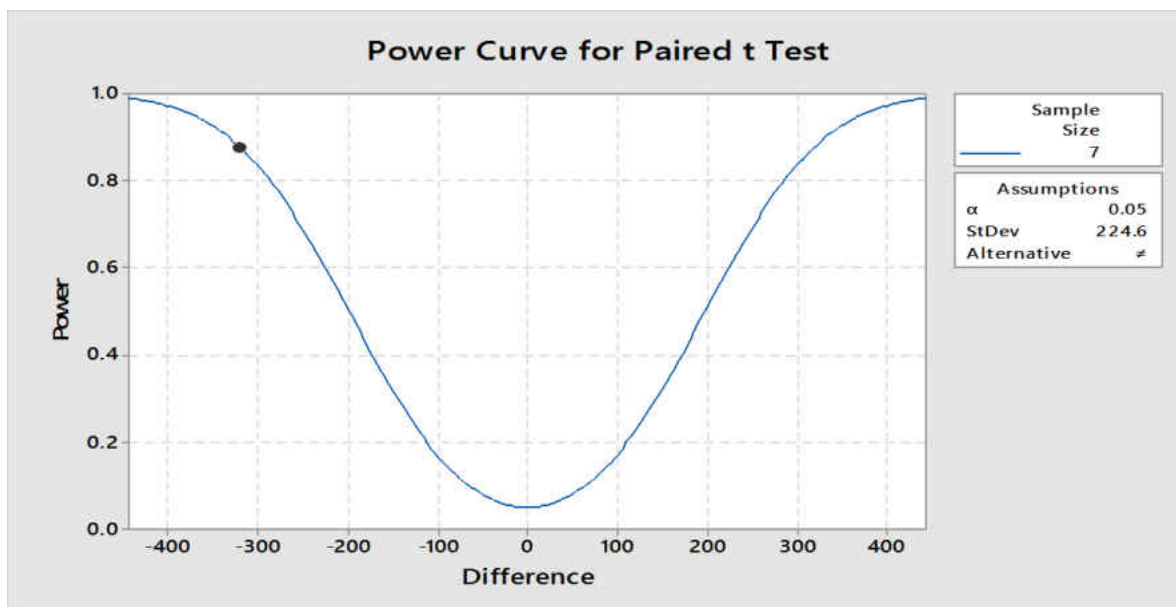


Figure 24: Required Sample Size for .8 Statistical Power for Timed Run

4.4 Statistical Differences between Treatments (Paired T-Test) Loaded vs Unloaded

In order to determine if there is a statistical difference between treatments of the unloaded versus loaded components of the PRT it is imperative to perform Paired T-Tests between data sets. This statistical analysis was chosen because it effectively calculates the difference between before and after measurement pairs and determines if there is a statistical difference. It was determined that a difference does indeed exist between all test means when evaluated for the loaded weight vest treatment of each event. As with prior tables and figures, the (exercise 1 event is for the unloaded condition) the exercise 2 event is for the loaded condition.

4.4.1 Paired T- Test for Push-Ups

A paired T-Test was conducted for the data sets containing performance information for research participant push-ups. The means of the two tests ($\mu = 22.43$, $\sigma = 10.13$ and $\mu = 12.36$, $\sigma = 5$ unloaded versus loaded respectively) yielded a 95% confidence interval of 7.86-12.28 (T-

Value 9.27 with a P-Value of 0.00. This concludes that there is indeed a statistical difference between the data sets after treatment of weight vests. Figure 25 depicts a Histogram of the confidence interval for the mean beyond the 0 difference mark.

Table 8: Paired T-Test for Unloaded vs Loaded Treatments of Push-Ups

	N	Mean	St Dev	SE Mean
Push-ups 1	35	22.43	10.13	1.71
Push up 2	35	12.36	5.00	0.85
Difference	35	10.07	6.42	1.09

95% CI for mean difference: (7.86, 12.28)

T-Test of mean difference = 0 (vs \neq 0): T-Value = 9.27 P-Value = 0.000

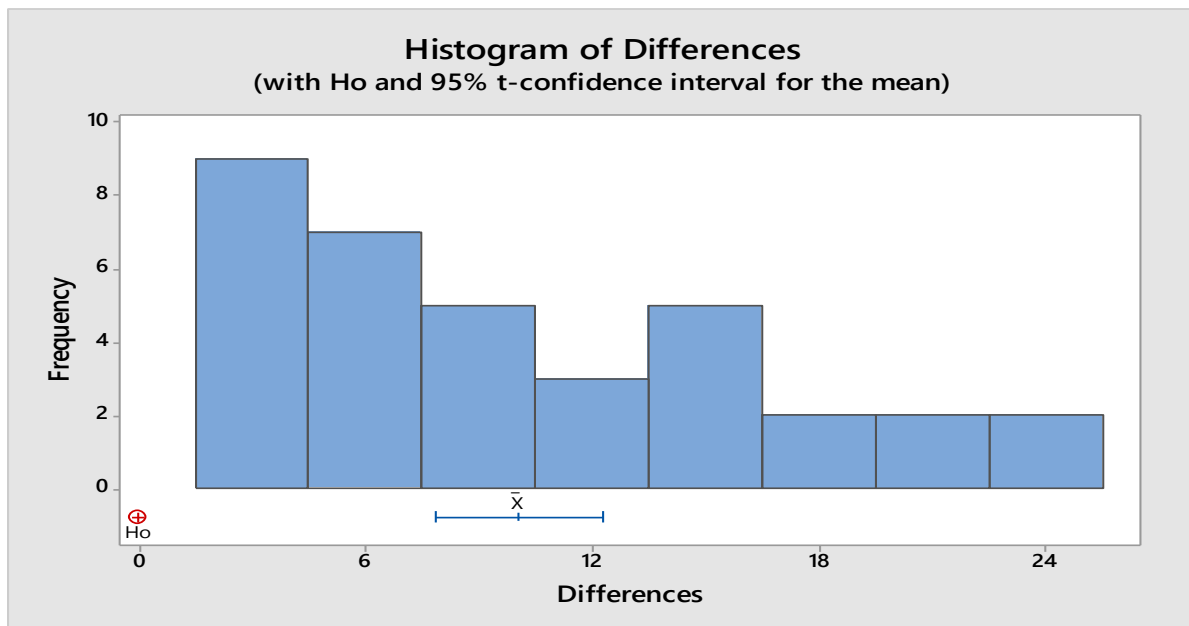


Figure 25: Histogram of Differences of Means for Push -Ups Unloaded vs Loaded

4.4.2 Paired T-Test for Curl-Ups

A paired T-Test was conducted for the data sets containing performance information for research participant push-ups. The means of the two tests ($\mu = 63.23$, $\sigma = 21.86$ and $\mu = 37.40$, $\sigma = 10.25$ unloaded versus loaded respectively yielded a 95% confidence interval of 20.81-30.8 repetitions (T-Value 10.48 with a P-Value of 0.00. This concludes that there is indeed a statistical difference between the data sets after treatment of weight vests. Figure 26 depicts a histogram of the confidence interval for the mean beyond the 0 difference mark.

Table 9: Paired T-Test for Unloaded vs Loaded Treatments of Curl-Ups

	N	Mean	St Dev	SE Mean
Curl ups 1	35	63.23	21.86	3.70
Curl ups 2	35	37.40	10.25	1.73
Difference	35	25.82	14.58	2.46

95% CI for mean difference: (20.81, 30.83)

T-Test of mean difference = 0 (vs \neq 0): T-Value = 10.48 P-Value = 0.000

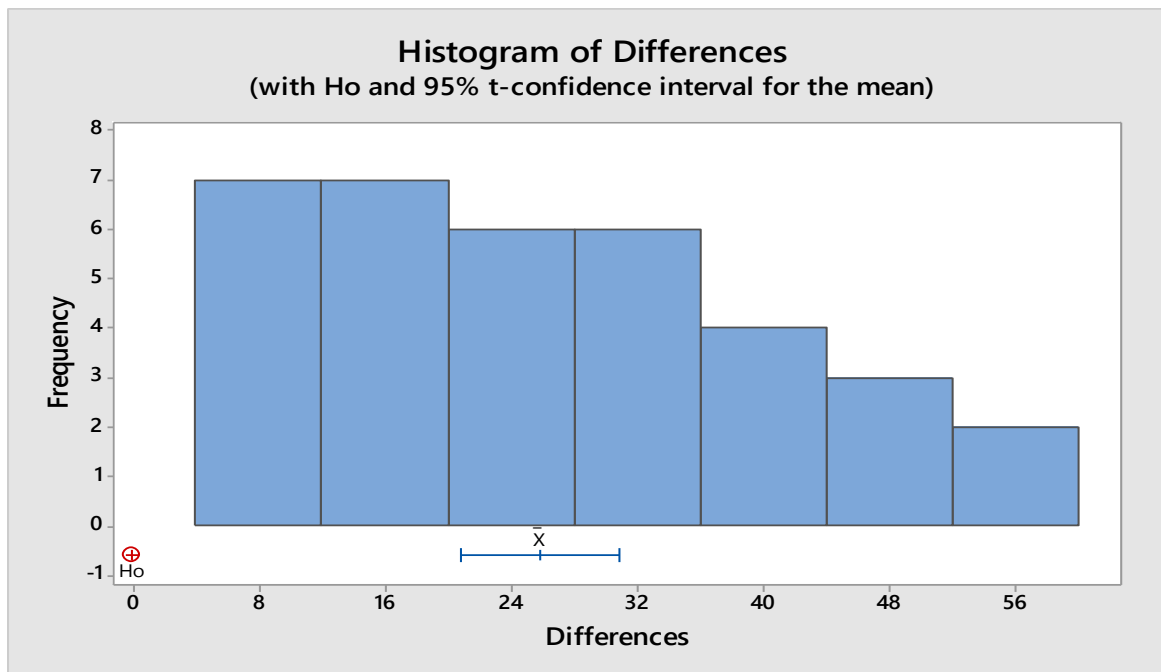


Figure 26: Histogram of Differences of Means for Curl-Ups Unloaded vs Loaded

4.4.3 Paired T-Test for Timed Run

A paired T-Test was conducted for the data sets containing performance information for research participant push-ups. The means of the two tests ($\mu = 967.4$, $\sigma = 192.2$ and $\mu = 1287.4$, $\sigma = 112.8$ unloaded versus loaded respectively yielded a 95% confidence interval of -369.7 to -270.2 seconds. (T-Value -13.08 with a P-Value of 0.00. This concludes that there is indeed a statistical difference between the data sets after treatment of weight vests. Figure 27 depicts a histogram of the confidence interval for the mean beyond the 0 difference mark. It can be seen the difference here is negative because the timed run is based on minimum time run versus maximum repetitions. All Paired-T Test analysis concluded a statistical difference between treatments.

Table 10: Paired T-Test Unloaded vs Loaded Treatments of Timed Run

	N	Mean	St Dev	SE Mean
run (s)	35	967.4	192.2	32.5
Run 2 (s)	35	1287.4	112.8	19.1
Difference	35	-319.9	144.7	24.5

95% CI for mean difference: (-369.7, -270.2)

T-Test of mean difference = 0 (vs \neq 0): T-Value = -13.08 P-Value = 0.000

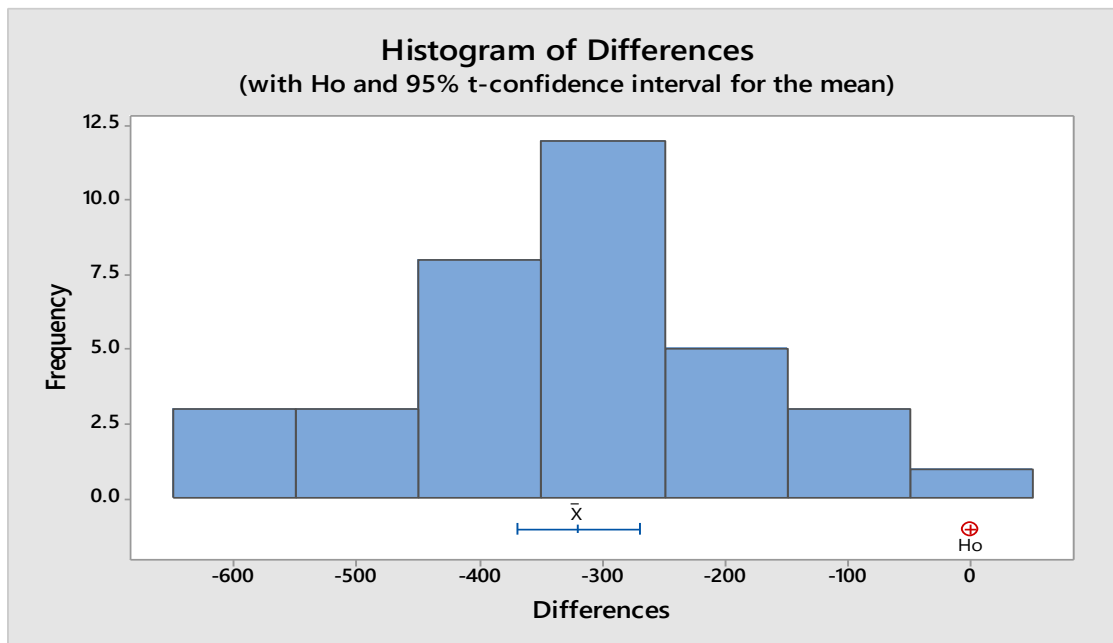


Figure 27: Histogram of Differences of Means for Timed Run Unloaded vs Loaded

4.5 Analysis of Correlation between Treatments (Pearson's Product Moment)

An analysis of correlation between the weight of each individual and their performance in each PRT exercise has been conducted. This is arguably the most critical factor of the research study in that the outcome answers the research question through hypothesis testing. Data has been collected for all research participants including physical characteristics and their performance for both loaded and unloaded test phases. This data has been evaluated for normality and Paired T-Tests have shown there is a significant difference between treatment means for all PRT exercises. Pearson Product Moment Correlation (r value) testing was conducted to determine the relationship between the variables weight and participant performance for both loaded and unloaded conditions. The r value (-1 to 1) has been determined for each event prior to and after weight vest usage. There is a moderate to strong correlation for all exercises prior to treatment, either positive or negative depending on test, which is significantly reduced as compared to evaluating the r value after the loaded tests. Although a test of this nature has never been conducted for military fitness testing on females, some testing has been conducted with loaded back packs and males (Vanderburgh, P. M. et al 2011). Those test showed a correlation of $r = -.38$ with $p = .0039$ for pushups and $r .42$ with $P = .0013$ for the timed run. The authors successfully showed mitigation of the correlation by significantly reducing the correlation $r = -.06$ with $p = .661$ for pushups and $r .06$ with $P = .661$. There are no comparisons for the curl- up test before loaded carriage for either male or female subjects at this time. The goal of the loaded carriage is to get the r value as close to 0 as possible which would indicate no correlation between weight and performance scores.

4.5.1 Correlation Results for Push-Ups Unloaded versus Loaded

Correlation analysis has been completed for both the unloaded and loaded push up tests of the PRT performed by each research participants. The correlation was observed between the participants weight and push-up repetitions before and after the use of the 20 lb weighted vest. There was a moderate negative correlation coefficient of $r = -.442$, $P = .008$ which indicates as the participant's weight increases their repetitions of push-ups decreases. The correlation coefficient changes on observation of the loaded push-ups test with $r = -.096$, $P = .583$. This indicates there is still a very weak negative correlation, but it is reduced by 78%. Figure 28 and 29 depict both the unloaded and loaded correlation between weight and push-ups repetitions. It can be seen that the overall repetition mean for push up decreases as well.

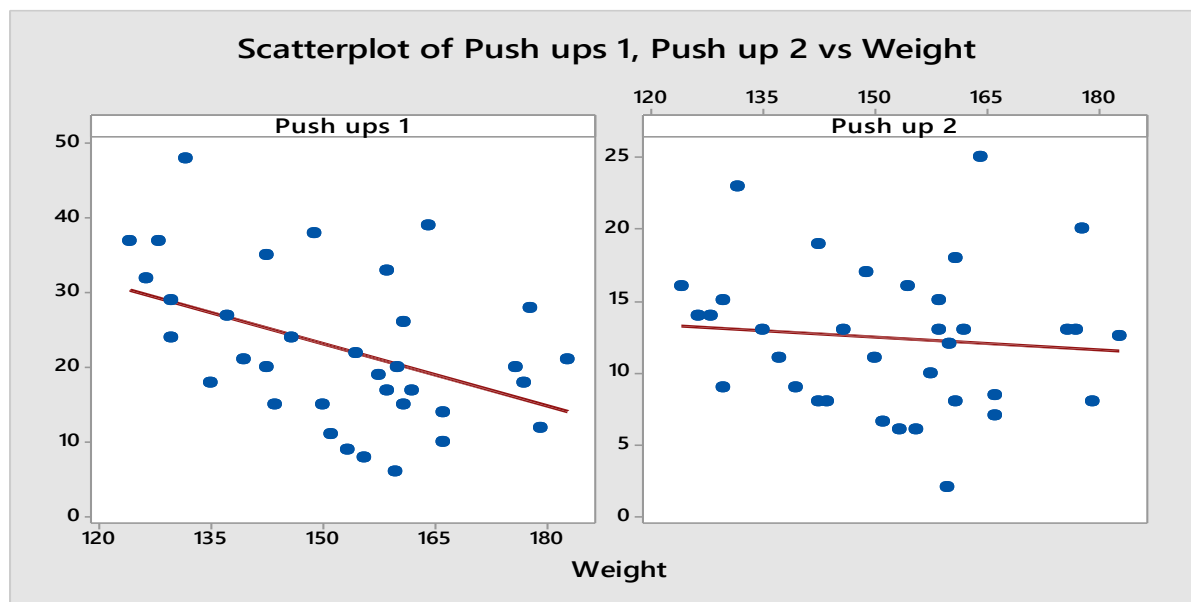


Figure 28: Scatterplot of Push-Ups vs Weight Correlation Unloaded and Loaded

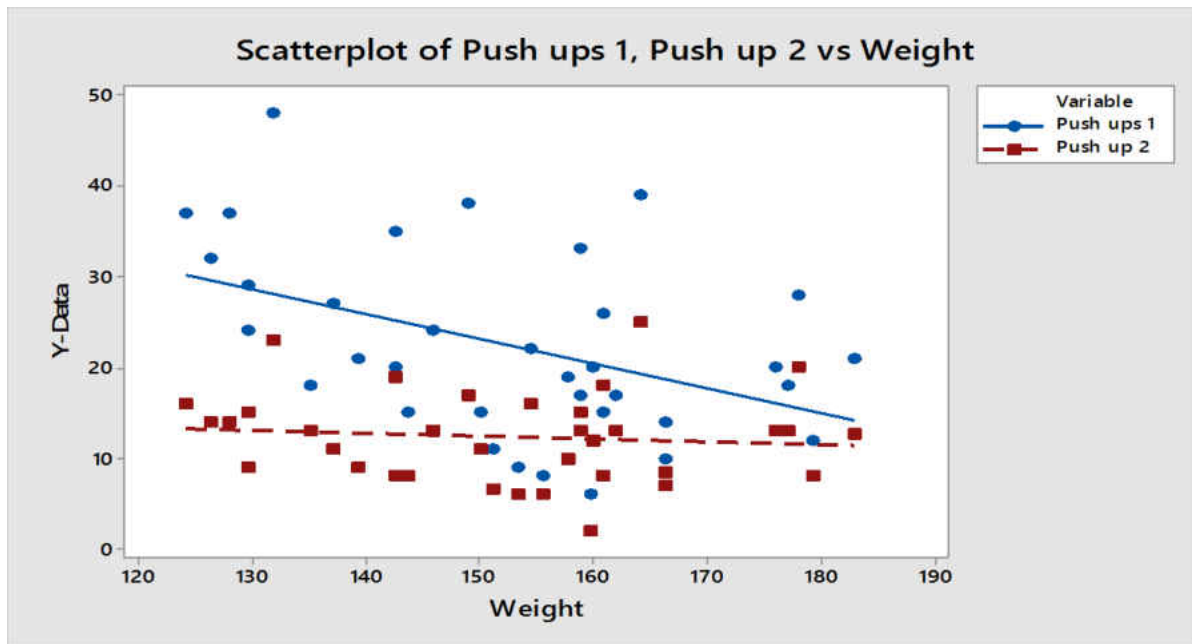


Figure 29: Scatterplot of Push-Ups vs Weight Correlation Unloaded and Loaded (Overlaid Graph)

4.5.2 Correlation Results for Curl-Ups Unloaded versus Loaded

Correlation analysis has been completed for both the unloaded and loaded curl-up tests of the PRT performed by each research participants. The correlation was observed between the participants weight and curl-up repetitions before and after the use of the 20 lb weighted vest. There was a moderate negative correlation coefficient of $r = -.458$, $P = .006$ which indicates as the participant's weight increases their number of curl-up repetitions decreases. The correlation coefficient changes on observation of the loaded curl-ups test with $r = -.070$, $P = .688$. This indicates there is still a very weak negative correlation, but it is reduced by 85%. Figure 30 and 31 depict both the unloaded and loaded correlation between weight and push-ups repetitions. It can be seen that the overall repetition mean for push up decreases as well.

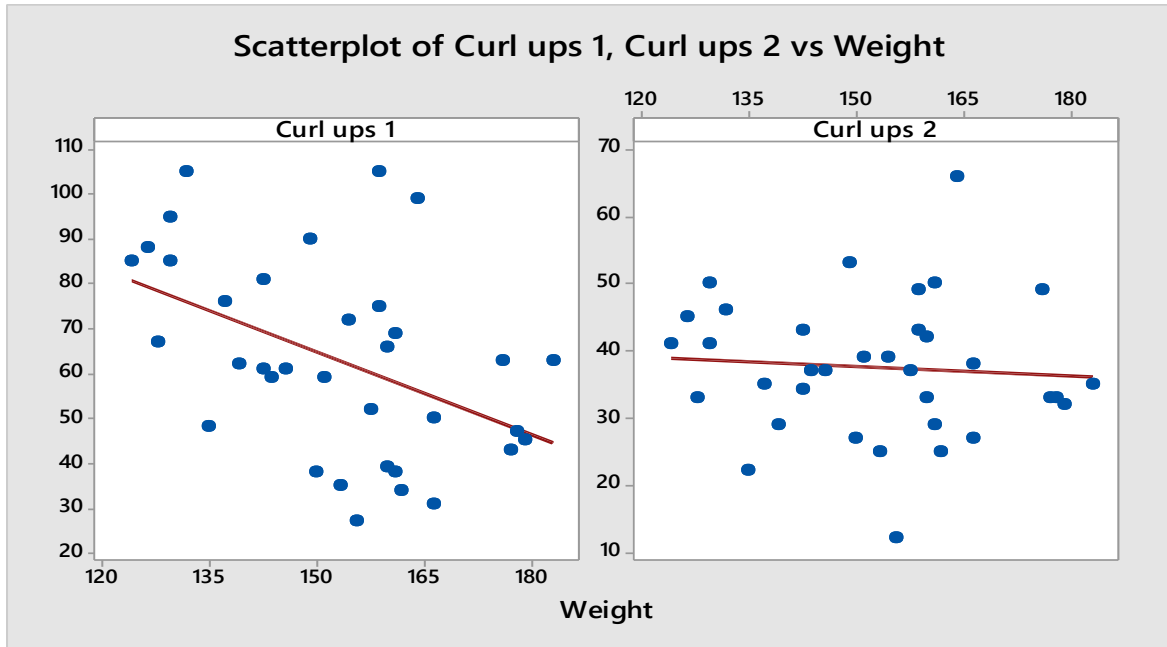


Figure 30: Scatterplot of Curl-Ups vs Weight Correlation Unloaded and Loaded

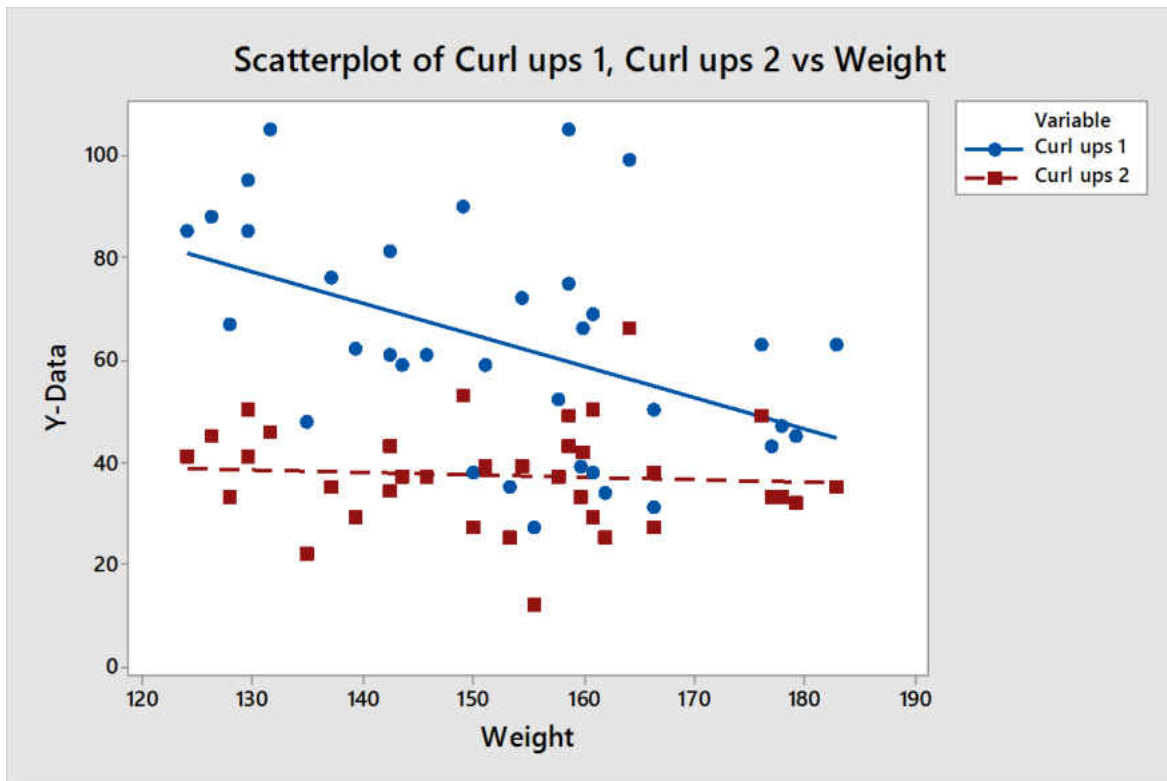


Figure 31: Scatterplot of Curl –Ups vs Weight Correlation Unloaded and Loaded (Overlaid Graph)

4.5.3 Correlation Results for Timed Run Unloaded versus Loaded

Correlation analysis has been completed for both the unloaded and loaded timed-run tests of the PRT performed by each research participants. The correlation was observed between the participant's weight and the number of seconds taken to complete the 1.5 mile run before and after the use of the 20 lb weighted vest. There was a strong positive correlation coefficient of $r = .611$, $P = .000$ which indicates as the participant's weight increases their run time also increases. The correlation coefficient changes on observation of the loaded timed distance run test with $r = .095$, $P = .587$. This indicates there is still a very weak positive correlation, but it is reduced by 84%. Figure 32 and 33 depict both the unloaded and loaded correlation between weight and time (in seconds) to complete the timed run. It can be seen that the overall mean time for the run increases as well.



Figure 32: Scatterplot of timed-Run vs Weight Correlation Unloaded and Loaded

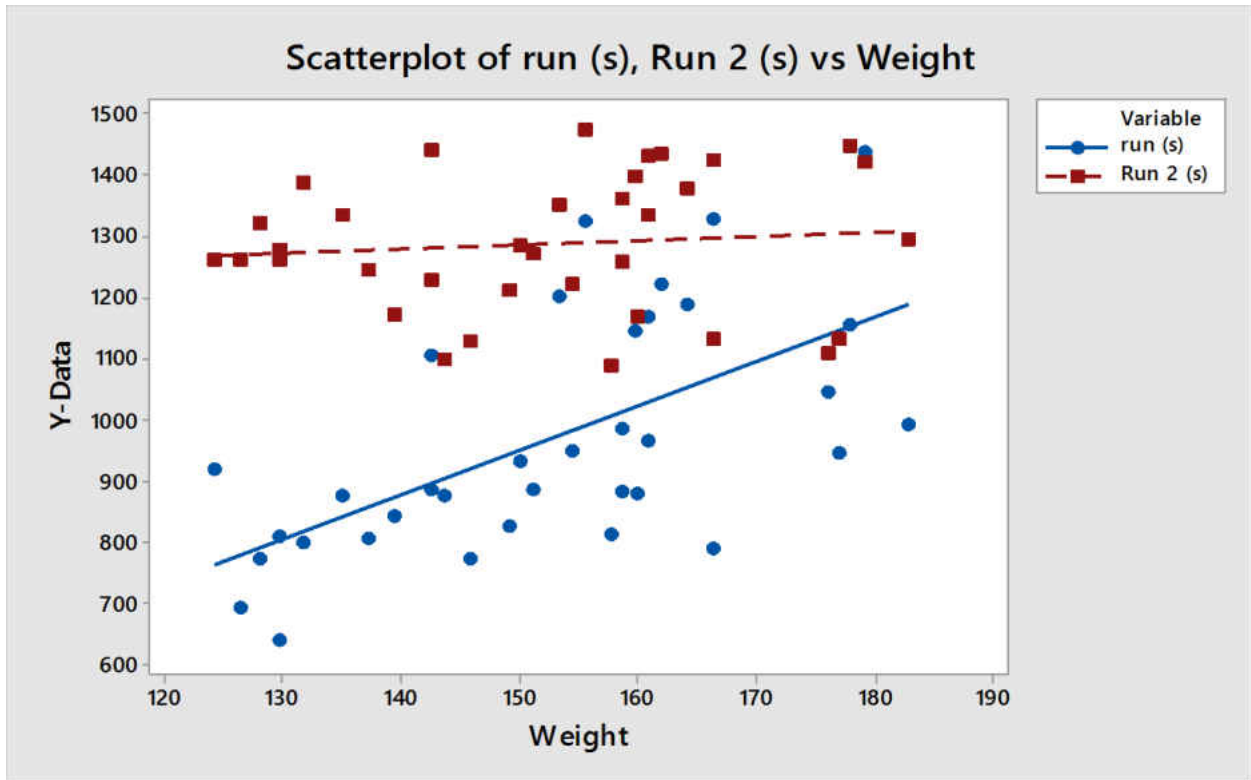


Figure 33: Scatterplot of timed Run vs Weight Correlation Unloaded and Loaded (Overlaid Graph)

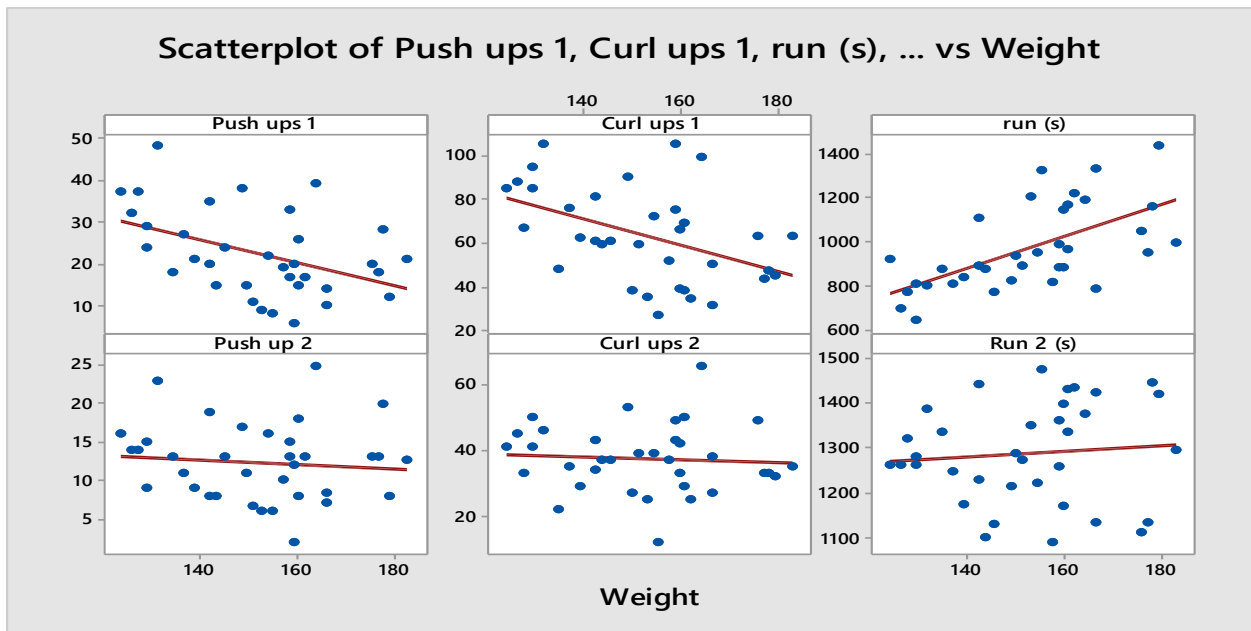


Figure 34 : Scatterplot Overview of All Exercises Push-Ups, Curl-Ups, & Timed Run performance vs Weight

4.5.4 Research Question Hypothesis Test

In order to answer the research question and determine if the weight vests significantly reduced the mass bias found in the PRT exercises we conduct the following hypothesis tests. Each test has been rewritten to be more specific to each PRT exercise being tested.

H_{01} : The correlation coefficient IS NOT significantly different from 0 for the PRT push-ups.

H_{a1} : The correlation coefficient IS significantly different from 0 for the PRT curl-ups.

The Pearson correlation of participant weight and Loaded Push-ups $r = -0.096$ with P-Value = 0.583 which is greater than $>$ our .05 significance level for this test and we accept the null hypothesis that there IS NOT a significant difference from 0 for the PRT push-ups. The weight vests successfully mitigated the bias in the Push-ups exercise of the PRT.

H_{02} : The correlation coefficient IS NOT significantly different from 0 for the PRT curl-ups.

H_{a2} : The correlation coefficient IS significantly different from 0 for the PRT push-ups.

The Pearson correlation of participant weight and loaded Curl ups $r = -0.070$ with P-Value = 0.688 which is greater than $>$ our .05 significance level for this test and we accept the null hypothesis that there IS NOT a significant difference from 0 for the PRT push-ups. The weight vests successfully mitigated the bias in the Curl-Ups exercise of the PRT.

H₀₃: The correlation coefficient IS NOT significantly different from 0 for the PRT timed run.

H_{a3}: The correlation coefficient IS significantly different from 0 for the PRT timed run.

The Pearson correlation of participant weight and loaded Timed Run is $r = 0.095$ with P-Value = 0.587 which is greater than $>$ our .05 significance level for this test and we accept the null hypothesis that there IS NOT a significant difference from 0 for the PRT push-ups. The weight vests successfully mitigated the bias in the Timed Run exercise of the PRT.

Table 11: Summary Results of Correlation Testing

PRT Exercise	Unloaded Results r value with P value	Loaded Results r value with P value	% Decrease in r value
Push-Ups	-.442, .008	-.096, 0.583	78
Curl- Ups	-.458, .006	-.070, 0.688	84
Timed Run (1.5 Mile)	.611, 0.00	.095, 0.587	85

CHAPTER FIVE: CONCLUSION

This research study investigated the use of load carriage through weighted vest as a method to mitigate mass bias found in military fitness testing. Mass bias is an inherent problem with the types of body weight only exercises utilized in all of the U.S. Military branches. We have seen through reviewing literature the current military fitness testing has changed little throughout the past several decades although there has been steady evolutions that should have provided the impetus such as a drastically changing demographic. The current test methods of performing push-ups, curl-ups, and a timed run is logistically sound and simple in nature providing a roadblock to change. Although we have examined several different methods make the fitness testing more accurate and fair, some methods would be difficult to implement. The need remains for a viable way that commanders can easily eliminate mass bias making the current test method fare for all service members. Especially since the resulting test outcomes can greatly affected their career paths. Prior research has shown loaded carriage to be effective when dealing with male cadets and mass bias. This research sought out to determine if an improved load carriage method (weighted vests versus loaded back packs) would work for female service members to which there is little to no data. A sample research participant group was organized based on the current U.S. Military female demographics. Gathering the target participants included evaluations of current fitness levels and physical characteristics prior to testing. The group was acclimated to exercising with weight vests and participant performance was gathered through multiple evolutions of the Physical Readiness Test. The data sets gathered were statistically analyzed to determine if the mass bias seen prior to using the weight vests was eliminated during the second test with the weighted vests. Hypothesis testing using Pearson

Product Moment Correlation (r value) yielded results showing the weight vests effectively lowered these correlation values 78%, 85%, and 84% for the push-ups, curl-ups, and timed run events respectively. The resulting r values were effectively zero through implementing the changes. Although these findings cannot be directly related to prior studies on male members, they are in alignment with the prior successful use of load carriage to mitigate mass bias. It is my hope that the resulting findings from this study can be used as a basis to implement load carriage and make military fitness testing fairer for all members.

5.1 Contributions to Body of Knowledge

There is a wealth of information in prior research pertaining to exercise fitness and testing. This body of knowledge is narrowed down when relating these studies to military fitness testing and even further filtered when discussing bias within the current military fitness tests. There is a substantial gap of knowledge pertaining to required fitness tests and females members which should be addressed. The lack of data on female members is probably due to the fact the military has traditionally been a male dominated organization.

Prior studies have shown that mass bias is a quantifiable issue in the current fitness testing and there are limited studies demonstrating methods on how to mitigate it. Various authors recommend different techniques including mathematical corrections to the scoring systems and changing exercises completely, but none of them have been implemented. This is likely due to specific arguments against such methods combined with the simplistic nature of the current test method. Commanders require a solution that can be deployed to remote locations, ships at sea, and home bases alike. Utilizing loaded carriage vests may provide that solution in

that they are not over encumbering to be deployed to needed locations and can be easily utilized during testing. Limited prior research involving males has shown the load carriage method to work in eliminating mass bias for the push-ups and timed run events. The researchers used back packs which did not allow them to gather data on the curl-ups component of testing and could be awkward in maneuvering during fitness testing. The acclimation period for members to get used to back packs could very well be higher than vests which are designed for fitness purposes. Although the study showed loaded carriage could be used for males on the push-ups and timed run, there was no validation that load carriage would work for female members.

This research study contributed to the body of knowledge in that an improved load carriage method employing weighted vests can now be used for female members that will successfully mitigate mass bias on all three components of the military fitness tests. This method could not have been implemented in units which consist of mixed genders. There was no validation of success or usefulness for females in prior work conducted before this study. Additionally, data gathered from this study can be extended to research pertaining to other military fitness tests including occupational relevance and combat fitness. The addition of fitness data on females for military testing helps fill the lack of data that exists for this gender.

5.2 Future Research

The military will most certainly be a required asset well into the future of the U.S. and the global community as well. As the world continually advances there will be an unending need for the evolution of Armed Forces along with it. Failure to do so will result in losses to efficiency, operating effectiveness, and global security. We have seen an example of an

unchanged fitness test method which can effect service men and women's careers and doesn't provide an accurate to commander of actual fitness levels and readiness. Continued research will help close gaps in knowledge related to the military and the realm of fitness is no exception. As discussed, there remains a large lack of data for female service members. Recent topics within the news headlines have brought some attention back to the topic of military fitness testing such as the debate of whether women should be allowed in combat roles. The tests required for military occupational specialties which require combat training and fitness differ from traditional military fitness tests involving push-ups, curl-ups, and timed runs but are very important in bringing equality for all members. A larger pool of highly qualified personnel who can perform the required duties will only benefit the Armed Forces. Further studies are required in occupational relevance, combat fitness requirements, and the advancement of overall wellbeing for all members.

This research was limited to a small group of participants and could be expanded upon. Some limitations included issues with the current scoring system. When individuals complete military fitness testing their raw scores are translated to another scale. It was found that utilizing the weighted vests lowered scores on this scale so that essentially all members would receive a failing grade of unsatisfactory. Research in developing a new scale in conjunction with loaded carriage would require an expanded study to determine what scores are considered satisfactory, good, excellent, and maximum. It was also discovered some individuals were uncomfortable with the weighted vests and they could not properly acclimate to using them for the test. Further research could be conducted in better assessing the average time and methods best used for making members comfortable. The study conducted went through great lengths to mimic the current U.S. Military female demographic, but the U.S. Military would obviously have the best

means of hitting the target demographics with their own personnel. These studies could be extended to current active duty members both male and female to gain further insight and make possible improvements. There is an inherent risk with any physical activity and increased mass has been shown to be a possible contributor to injury during distance evolutions. Prior studies pointed to long marches rather than 1.5 -2.0 mile runs, but prolonged use of vests that increase mass could possibly contribute to complications among members. Further research could be conducted to determine if utilizing load carriage 2 times per year would be a safety concern over a longer period of time. There were no injuries in this study, but safety should always be of the highest concern. The 20 lb load chosen was based off of ratios taken from prior studies on males. A lighter load may indeed yield acceptable results in mitigating the mass bias, but additional testing would need to be performed. Future research could include a study for various loaded conditions including weights at 20, 15, and 10 lbs. to determine which load is the most efficient at mitigating the mass bias.

These are some recommendations for future research in hopes that a continued effort will be made to fill the growing gap of knowledge related to military fitness and the female gender. This study found that loaded carriage could be implemented as a means for commanders to eliminate bias in current fitness testing. This method has now been validated for both males and females. Although different vehicles for establishing the load were utilized, it should be viable to perform the current fitness tests for combined gender units.

APPENDIX A: RISK ASSESSMENT QUESTIONNAIRE

Physical Readiness Test Risk Assessment Questionnaire

(1) Do you have a current PHA? If no, you may not participate today.

(2) Do you have chest pain (with or without exertion), bone or joint pain, high blood pressure or high cholesterol? If yes, have you been cleared, by your medical provider, to participate in PT?

(3) Have you had a change in your medical status since the last time you were asked these questions?

(4) Are you ill today or know of any medical condition that may prevent you from participating in physical activity today?

(5) (For PRT Only) Did you answer yes to any questions? If yes, do you have a PFA medical waiver or clearance form on file? If no you may not participate today?

Signature: _____ Date: _____

APPENDIX B: PARTICPANT AVAILABILITY QUESTIONNAIRE



**BODY MASS BIAS MITIGATION FOR WOMEN IN MILITARY PHYSICAL
 READINESS
 TESTING THROUGH LOAD CARRIAGE IMPLEMENTATION
 PARTICIPANT AVAILABILITY QUESTIONNAIRE**

Principal Investigator(s): Aaron Yeaton

Faculty Supervisor: Pamela McCauley, PhD

This research study requires several meetings over the period of 4 weeks. Each meeting is specifically designed to gather vital information used for the final analysis of the project. It is very important that test participants take part in each meeting throughout the study. We realize that participating in this study is a commitment outside your normal obligations and may be difficult at times. This form will be used to determine your availability for the study (and the groups as a whole) to make the most acceptable and efficient schedule.

Estimated time required per each week of the study

Week	Number of Meetings	Total time of all meetings for week
1	2	1 hour 45 minutes
2	2	2 hours
3	1	2 hours
4	1	2 hours

Your weekly availability - please indicate by (x) the day of the week you are most available

Time of day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
a.m.							
p.m.							

Participant Number: _____

**APPENDIX C: PARTICPANT PRETEST PHYSICAL ASSESSMENT
DATA SHEET**



**BODY MASS BIAS MITIGATION FOR WOMEN IN MILITARY PHYSICAL
 READINESS
 TESTING THROUGH LOAD CARRIAGE IMPLEMENTATION
 PARTICIPANT PRETEST PHYSICAL ASSESSMENT DATA COLLECTION SHEET**

Principal Investigator(s): Aaron Yeaton

Faculty Supervisor: Pamela McCauley, PhD

Participant General/Contact Information

Name	Phone number	Email address	Times workout per week?

Participant Measurements taken prior to arm cycle test:

Measurement	Reading
Age	
Height	
Weight	
Ethnicity	
Omron HBF-516 Body Composition	
Natural Waist	
Hip	

Neck	
CV= (Waist + Hips –Neck)	
Max HR = 220 – Age	
70% Max HR for arm cycle test = (220-Age)*.70	

The persons asked to perform the PRT should be in a similar fitness condition of an active duty sailor. A short physical fitness screening will be given to estimate current levels of overall fitness. Participants will be asked to stretch prior to exercises and a summary of objectives given. Resting heart rate will be taken. Each person will be asked to arm cycle using a portable body cycle for a short period of time in order to raise their heart rate to 70% of max (220-age) and then given 80 seconds to rest. Heart rates will be monitored with the Polar Electro E600 heart rate monitors equipped with chest straps to discover whose heart rates return to near resting over 80 seconds. This test will be used as a rough indicator of current fitness.

1. Calculate 70% Max Heart Rate for start time~ t0
 $(220 - \text{Participants Age}) \times .70 = \text{HRo}$

2. Calculate decrease % in heart rate
 $(\text{HRo} - \text{HR80})/\text{HRo} = \% \text{ decrease}$

**Time will be considered = 0 at the 70% of heart rate mark and the clock will start concurrently with exercise stopping. The participants hart rate will be recorded at time equals 80seconds (HR80) of resting. The percent decrease in heart rate will be evaluated and recorded.

Measurements taken during arm cycle test:

Measurement	Reading
Time start cycle	
70% MHR Start time (To)	
Actual Heart Rate at To (HRo)	
HR after 80 seconds (RHR)	
% Decrease = $(\text{HRo} - \text{HR80})/\text{HRo}$	

APPENDIX D: PARTICIPANT SELF FITNESS ASSESSMENT



**BODY MASS BIAS MITIGATION FOR WOMEN IN MILITARY PHYSICAL READINESS
TESTING THROUGH LOAD CARRIAGE IMPLEMENTATION
SELF FITNESS ASSESSMENT QUESTIONNAIRE**

Principal Investigator(s): Aaron Yeaton Faculty Supervisor: Pamela McCauley, PhD

PARTICIPANT #: Your participant number found on the race bib assigned during first meeting.

1. I am or have been affiliated with the U.S. Armed Forces including status of any of the following: veteran, retired, active duty, Reserve Guard, ROTC.

Yes / No ** If you answered yes, please note your affiliation/experience and if you have ever failed a Military Fitness Test and for which branch of service. **

2. How many times a week do you exercise and usually how long?

3. What exercises do you mostly perform? Please circle any that apply.

Weight Lifting

Running

Cycling

Aerobics/Dance...

Curl-Ups/Push-Ups

Walking

Organized Sport

Stationary Equipment (Bike/Elliptical)

Other (Please Specify) _____

4. Have you ever used a weighted vest for working out? Yes / No

** If yes, please list how often you use/used weighted vests, what weight, and what activities or exercises di you perform.

APPENDIX E: PHYSICAL READINESS TEST DATA COLLECTION SHEETS



**BODY MASS BIAS MITIGATION FOR WOMEN IN MILITARY PHYSICAL
 READINESS
 TESTING THROUGH LOAD CARRIAGE IMPLEMENTATION**

PHYSICAL READINESS TEST DATA COLLECTION SHEET

Principal Investigator(s): Aaron Yeaton

Faculty Supervisor: Pamela McCauley, PhD

Participant Contact/General Info:

Participant #	Phone number	Email	Health Questionnaire Complete (x)

PRT Unloaded Test # 1 Results

Measurement / Test	Result / #
Bib number (1-40)	
Curl – Ups	
Push- Ups	
Time start push- ups / time start run	
Timed Run (min:sec)	

PRT Loaded Test # 2 Results

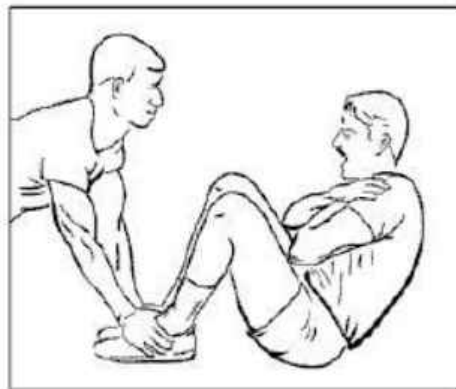
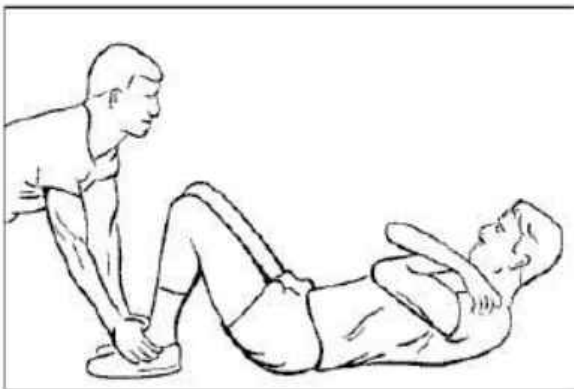
Measurement / Test	Result / #
Bib number (1-40)	
Curl- Ups	
Push- Ups	
Time start push-ups/ time start run	
Timed run (min:sec)	

PRT events shall be completed on the same day, at least 2 minutes, but no more than 15 minutes between each event and in the following sequence:

- a. Curl-ups
- b. Push-ups
- c. Cardio-respiratory fitness

(1) The member will begin by lying flat on back with knees bent, heels about 10 inches from buttocks. Arms shall be folded across and touching chest with palms of hands touching upper chest and shoulders (thumbs touching clavicle (collarbone)).

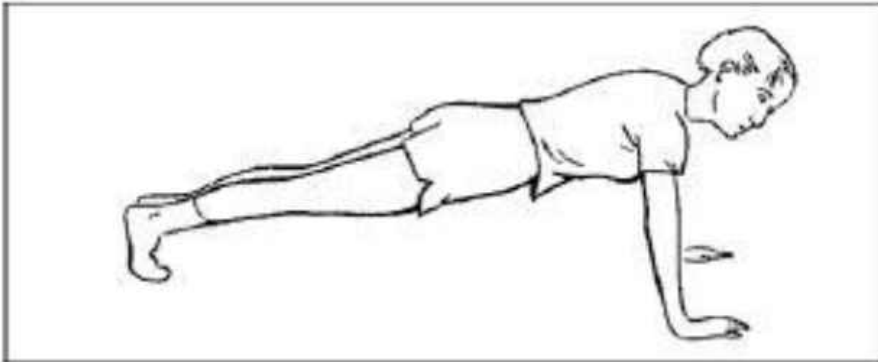
(2) Feet shall be flat on the deck and held by partner's hands. If preferred the partner may use their knees (on the side of the feet only) in addition to their hands to secure member's feet. Any other means of securing the member's feet is not authorized.



a. Push-ups procedures:

(1) Member will begin in leaning rest position on the deck so that body forms a straight line through the shoulders, back, buttocks, and legs. Weight is supported only with the toes and palm of the hands. Feet shall not be in contact with the bulkhead or other vertical support surface.

(2) Arms are to be straight with palms flat on the deck, directly under the shoulders or slightly wider than shoulder width.



3. 1.5-mile Run and/or Walk

Event consists of running or walking 1.5 miles as quickly as possible. Any combination of running or walking is allowed to complete the event.

a. Event Procedures:

- (1) Conducted on a flat and solid surface track or outdoor course.
- (2) Member will stand at start line.
- (3) Timer will signal start and call out time intervals until completion of test.
- (4) Time is recorded with stopwatch to nearest second.

b. **1.5 Mile Run and/or Walk Event is ended when:**

- (1) The member completes the 1.5 mile. CFL will record the time.
- (2) Takes a short cut or does not complete the entire 1.5 mile course. If not complete (other than for injury), the cardio portion of the PRT will be scored a **59:59** in PRIMS and scored a "fail." The member will not be eligible for a retake under the Bad Day policy.

APPENDIX F: IRB OUTCOME APPROVAL LETTER



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Aaron Yeaton**

Date: **May 28, 2015**

Dear Researcher:

On 05/28/2015, the IRB approved the following human participant research until 05/27/2016 inclusive:

Type of Review: UCF Initial Review Submission Form
Project Title: BODY MASS BIAS MITIGATION FOR WOMEN IN
MILITARY PHYSICAL READINESSTESTING THROUGH
LOAD CARRIAGE IMPLEMENTATION
Investigator: Aaron Yeaton, PhD Industrial Engineering
IRB Number: SBE-15-11205
Funding Agency:
Grant Title:
Research ID: N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30 days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at <https://iris.research.ucf.edu>.


If continuing review approval is not granted before the expiration date of 05/27/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the [Investigator Manual](#).

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

A handwritten signature in black ink that reads "Joanne Muratori". The signature is written in a cursive style with a small dot above the letter 'i' in "Muratori".

Signature applied by Joanne Muratori on 05/28/2015 03:10:40 PM EDT

IRB manager

LIST OF REFERENCES

- Armed Forces Health Surveillance, C. (2011). Diagnoses of overweight/obesity, active component, U.S. Armed Forces, 1998-2010. *MSMR*, *18*(1), 7-11.
- Barry, B. b. i. o. (2013). Women in Combat. *Survival (00396338)*, *55*(2), 19-30. doi: 10.1080/00396338.2013.784461
- Bell, N. S., Mangione, T. W., Hemenway, D., Amoroso, P. J., & Jones, B. H. (2000). High injury rates among female army trainees: A function of gender? *American Journal of Preventive Medicine*, *18*(Suppl. 3), 141-146.
- Bilzon, J. L. J., Allsopp, A. J., & Tipton, M. J. (2001). *Assessment of physical fitness for occupations encompassing load-carriage tasks*.
- Bohnker, B. K., Sack, D. M., Wedierhold, L., & Malakooti, M. (2005). Navy physical readiness test scores and body mass index (spring 2002 cycle). *Military Medicine*, *170*(10), 851-854.
- Burland, D., & Lundquist, J. H. (2013). The Dynamic Lives and Static Institutions of the “Two Armies”: Data from the Survey of Active Duty Personnel. *Armed Forces & Society*, *39*(1), 78-101. doi: 10.1177/0095327x11410858
- Clark, H. D. (2013). *Military load carriage during prolonged marches on lower extremity mechanics: Influence of gender*. (1539808 M.S.), State University of New York at Buffalo, Ann Arbor. Retrieved from
- Conway, T. L., & Naval Health Research Center, S. D. C. A. (1987). Behavioral, Psychological, and Demographic Predictors of Physical Fitness.

- Do, J. J., Samuels, S. M., Adkins, D. J., Clinard, M. E., & Koveleskie, A. J. (2013). Gender bias and pluralistic ignorance in perceptions of fitness assessments. *Military Psychology*, 25(1), 23-35. doi: 10.1037/h0094754
- Flanagan, S. (2008). *Fitness and performance classified according to body mass index and waist circumference for Canadian Forces personnel aged 50--59 years*. (MR48600 M.Sc.), University of Ottawa (Canada), Ann Arbor. Retrieved from
- Gooch, R. H. (2001). *Women and men in the new Navy: Life since Tailhook*. (3005672 Ph.D.), Salve Regina University, Ann Arbor. Retrieved from
- Hodgdon, J. A. (1999). A History of the U.S. Navy Physical Readiness Program From 1976 to 1999: 1999-08-18.
- Jackson, C. A., & Wilson, D. (2013). The Gender-Neutral Timed Obstacle Course: a valid test of police fitness? *Occupational Medicine*, 63(7), 479-484.
- Johnson, D., Lynch, J., Nash, K., Cygan, J., & Mayhew, J. L. *Relationship of Lat-Pull Repetitions and Pull-Ups to Maximal Lat-Pull and Pull-up Strength in Men and Women*.
- Johnson, V. V. T. (2009). *The association between stress and physical fitness testing in the 2005 Department of Defense population survey*. (1467326 M.P.H.), The University of Texas School of Public Health, Ann Arbor. Retrieved from
- Kennedy-Armbruster, C. (2012). *Effects of a supervised exercise program on fitness level, functional-movement ability, sedentary time, and fatigue for active-duty military personnel*. (3550826 Ph.D.), Indiana University, Ann Arbor.

- Kyröläinen, H., Häkkinen, K., Kautiainen, H., Santtila, M., Pihlainen, K., & Häkkinen, A. (2008). Physical fitness, BMI and sickness absence in male military personnel. *Occupational Medicine (Oxford, England)*, 58(4), 251-256. doi: 10.1093/occmed/kqn010
- Moschella, K. N. (2013). *Predicting chest press strength from a 4RM triceps brachii exercise in trained women*. (1543279 M.S.), University of Nevada, Las Vegas, Ann Arbor.
- Nolte, R., Franckowiak, S. C., Crespo, C. J., & Andersen, R. E. (2002). U.S. military weight standards: What percentage of U.S. young adults meet the current standards? *American Journal of Medicine*, 113(6), 486-490.
- Paul, V. (2007). Correction Factors for Body Mass Bias in Military Physical Fitness Tests. *Military Medicine*, 172(7), 738-742.
- Peterson, D. D. (2013). Proposed Performance Standards for the Plank for Inclusion Consideration Into the Navy's Physical Readiness Test. *Strength & Conditioning Journal (Lippincott Williams & Wilkins)*, 35(5), 22-26.
- Reyes-Guzman, C. M., R. M. Bray, V. L. Forman-Hoffman, and J. Williams. 2015. "Overweight and obesity trends among active duty military personnel: a 13-year perspective." *Am J Prev Med* 48 (2):145-53. doi: 10.1016/j.amepre.2014.08.033.
- Savonen, K., Krachler, B., Hassinen, M., Komulainen, P., Kiviniemi, V., Lakka, T. A., & Rauramaa, R. (2012). The current standard measure of cardiorespiratory fitness introduces confounding by body mass: the DR's EXTRA study. *International Journal of Obesity*, 36(8), 1135-1140.
- Sell, K. M. (2006). *Development of minimal physical fitness test standards for firefighters*. (3226100 Ph.D.), The University of Utah, Ann Arbor.

- U.S. Census Bureau, S. A. o. t. U. S. (2012). Table 510. Department of Defense Personnel: 1960 to 2010. *National Security and Veterans Affairs*.
- Vanderburgh, P. M. (2008). Occupational relevance and body mass bias in military physical fitness tests. *Medicine & Science in Sports & Exercise*, 40(8), 1538-1545.
- Vanderburgh, P. M., & Crowder, T. A. (2006). Body mass penalties in the physical fitness tests of the Army, Air Force, and Navy. *Military Medicine*, 171(8), 753-756.
- Vanderburgh, P. M., & Katch, F. I. (1996). Ratio scaling of VO₂max penalizes women with larger percent body fat, not lean body mass. *Medicine & Science in Sports & Exercise*, 28(9), 1204-1208.
- Vanderburgh, P. M., Mickley, N. S., & Anloague, P. A. (2011). Load-carriage distance run and push-ups tests: no body mass bias and occupationally relevant. *Military Medicine*, 176(9), 1032-1036.
- Weiglein, L., Herrick, J., Kirk, S., & Kirk, E. P. (2011). The 1-mile walk test is a valid predictor of VO₂max and is a reliable alternative fitness test to the 1.5-mile run in U.S. Air Force males. *Military Medicine*, 176(6), 669-673.
- Weina, S. U. (2006). Effects of Pregnancy on the Army Physical Fitness Test. *Military Medicine*, 171(6), 534-537.