March 2019

The Ability of the U.S. Military's WBGT-based Flag System to Recommend Safe Heat Stress Exposures

David R. Almario
University of South Florida, david.almario@gmail.com

The Ability of the U.S. Military's WBGT-based Flag System

to Recommend Safe Heat Stress Exposures

by

David R. Almario

A thesis submitted in partial fulfillment
of the requirements for the degree
Master of Science in Public Health
College of Public Health
University of South Florida

Major Professor: Thomas E. Bernard, Ph.D.
Rachel Williams, M.D.
Alfred Mbah, Ph.D.

Date of Approval:
March 20, 2019

Keywords: sensitivity, specificity, occupational, prevention

ACKNOWLEDGMENTS

The NIOSH Sunshine Education and Research Center, as well as the U.S. Department of Veterans Affairs, have provided me with the extraordinary opportunity to pursue both a Master of Science in Public Health degree and a residency in Occupational Medicine at the University of South Florida, College of Public Health. Throughout my time at the University of South Florida, I have been fortunate to be mentored by consummate experts in their respective fields, including Drs. Rachel Williams, Alfred Mbah, and Thomas Bernard. Dr. Bernard was absolutely vital in the completion of my thesis project, and I am forever grateful for his support during this challenging but rewarding endeavor. There are many other professors and classmates who have been instrumental in my education and personal growth that I must acknowledge by other means. Spending the past two years in such an environment with these wonderful individuals has truly been a privilege. There was support from CDC/NIOSH under the Sunshine ERC (T42-OH008438), and the data came from a research grant (R01-OH03983).

TABLE OF CONTENTS

List of Tables ii

List of Figures iii

Abstract iv

Chapter One: Introduction 1
    History of the WBGT-based Flag System 1
    The Current WBGT-based Flag System 8

Chapter Two: Methods 11
    Statistical Analyses 13

Chapter Three: Results 16

Chapter Four: Discussion 20

Chapter Five: Conclusion 24

References 25

Appendix A: Permission Statement 28

LIST OF TABLES

Table 1: Summary of Observations from Progressive Protocol 17

Table 2: Summary of Observations from Time-Limited Protocol 17

Table 3: 2x2 Table for Progressive Protocol 18

Table 4: 2x2 Table for Time-Limited Protocol 18

Table 5: 2x2 Table for Combined Protocols 18

Table 6: Summary of Sensitivities, Specificities, and Kappa Scores 18

Table 7: Summary of Exposures and Durations from Time-Limited Protocol 19

LIST OF FIGURES

Figure 1: Characteristics of participants from the progressive protocol 14

Figure 2: Characteristics of participants from the time-limited protocol................................... 14

Figure 3: Time course of a progressive protocol trial. 15

Figure 4: Modified metabolic rate ranges for the flag system 15

ABSTRACT

The U.S. military currently uses a flag system based on wet-bulb globe temperature (WBGT) and metabolic rate to recommend heat stress exposure limits. This paper addresses the ability of the flag system to recommend safe heat exposures in a non-military population.

Two progressive heat stress studies provided data on 528 observations of safe or unsafe exposures of 4 hours over a range of WBGT conditions and metabolic rates using 29 participants wearing woven cotton clothing. For the two studies, range of WBGT conditions was 25 to 42°C, and the range of metabolic rates was 100 to 650 watts. These exposures were compared with the flag system's recommendations of safe exposure to determine the sensitivity and specificity of the flag system. A separate study provided 62 observations with participants undergoing a time-limited protocol at constant WBGT conditions. Observed durations of safe exposure time were compared to the flag system's recommended safe limits to determine sensitivity and specificity.

Based on the progressive protocol, sensitivity and specificity of the flag system for five ranges of WBGT and three categories of metabolic rate were 0.98 and 0.25, respectively. For the time-limited protocol, which applied only to the highest range of WBGT and light and moderate metabolic rate, both sensitivity and specificity were zero.

This study suggests that the flag system has high sensitivity but low specificity for long duration exposures, along with low sensitivity and specificity for time-limited exposures. However, the WBGT exposures in the time-limited trials were substantially higher than the threshold for the highest WBGT range in the flag system, which may account for the system's unexpected performance in the time-limited protocol.

CHAPTER ONE:

INTRODUCTION

History of the WBGT-based Flag System

The origin of the United States military's current wet-bulb globe temperature (WBGT) based flag system can be traced back to the 1950s, when heat-related illness regularly affected hundreds of military recruits. In particular, recruits at the Marine Corps Recruit Depot (MCRD) in Parris Island, South Carolina, had a weekly heat casualty incidence rate of 53 per 10,000 during the summer, and represented two-thirds of all Navy and Marine Corps heat casualties in 1952 (Minard, 1961). In 1954, MCRD instituted a flag policy to limit strenuous activity during conditions of high heat. This flag system was based on the temperature-humidity index, with different colored flags such as red and yellow reflecting specific combinations of air temperature and relative humidity. These flags would be raised in the vicinity of a training area, visible to supervisory staff. Based on the flag's color, active training was modified or suspended completely, regardless of previous acclimatization.

However, unlike the temperature-humidity index, WBGT incorporates not only air temperature and humidity, but also thermal radiation and air movement (NIOSH, 2016). Other indices such as the effective temperature including radiation (ETR) also combined these four variables (Minard, 1961). With ETR, direct measurements of air velocity with specialized instruments and interpretation by trained technicians were required. Also, complicated charts were needed to determine the radiation component of the ETR value. In contrast, the WBGT index did not require direct air velocity measurements, although wind is still considered indirectly. Personnel determining WBGT simply needed to be able to read multiple thermometers and apply the readings to a basic formula. A WBGT value was calculated by combining measurements from three types of thermometers: dry bulb, natural wet bulb, and globe (Minard, 1961). A dry bulb thermometer indicates air temperature. A natural wet bulb thermometer is affected by evaporative cooling as well as convection, and thus conveys a temperature reading that incorporates relative humidity and air movement. The globe thermometer contains a copper sphere painted black that absorbs radiation from the sun as well as from the surrounding environment. Thus, the WBGT index combines the effects of the four environmental variables of air temperature, humidity, thermal radiation, and air movement. To account for these variables, there are two equations to calculate WBGT, depending on the presence or absence of direct exposure to the sun (NIOSH, 2016). WBGT for indoors is equivalent to $0.7t_{nwb} + 0.3t_g$, while WBGT in an outdoor setting with sunshine is $0.7t_{nwb} + 0.2t_g + 0.1t_a$.

In the summer of 1954, shortly after MCRD instituted the flag system based on the temperature-humidity index, Yaglou and Minard first conducted studies among Marine Corps recruits comparing the effectiveness of both the ETR and WBGT indices in predicting heat casualties. Mean evaporative sweat rate, mean change in core body temperature, and mean change in heart rate were chosen as the physiological heat stress criteria with which to correlate the ETR and WBGT indices (Yaglou & Minard, 1956). Evaporative sweat loss was determined by first assuming 1 kg/hr of water evaporated corresponded to the removal of approximately 580 calories/hr of body heat. Recruits were then weighed before and after various training exercises. During these exercises, each recruit wore a full fatigue uniform, boots, and a helmet, along with carrying a bayonet and marching pack. Core body temperature was measured with rectal probes, while heart rate was measured manually. Additionally, daily incidence of heat stress events during the summer was recorded, and then compared with the weather conditions at the time. These conditions included air temperature, solar radiation, air movement, and humidity. Both ETR and WBGT were found to have higher correlation with physiological heat stress criteria and incidence than the previously used temperature-humidity index. Although ETR had slightly higher correlation than WBGT, WBGT was determined to be a more practical index, given the relatively simple measurement protocols. Also, the temperature-humidity index was found to unnecessarily restrict training during cloudy and windy days, since it did not consider radiation or air movement. This would be the equivalent of a false positive prediction of heat stress. Furthermore, training during sunny and windless days, when most of the heat casualties occurred, was unsafely permitted by the temperature-humidity index. This scenario would serve as a false negative prediction of heat stress.

Ultimately, Yaglou and Minard (1956) recommended training duration limits based on both WBGT and level of acclimatization as follows. For WBGT values between 80-84°F, new unacclimated recruits were restricted to three hours per day of training during the first week, four to five hours the second week, and no more than six hours for the rest of the recruitment program. WBGT values between 85-87.9°F indicated a yellow flag in which strenuous exercises were suspended for new recruits during the first two weeks of training. Values above 88°F indicated a red flag, in which any physical training for all trainees, regardless of acclimatization, was halted. However, for seasoned troops who had already endured a summer of acclimatization, limited physical activity was allowed for WBGT values between 88-90°F. Unless specified, these limits assumed recruits performed a moderate level of physical activity such as marching at a standard military pace during the day. Yaglou and Minard suggested that more research would be needed to determine duration limits for training intensities other than moderate. Additionally, Yaglou and Minard were not able to perform direct measurements of metabolic rate. Instead, they assumed that an evaporative sweat loss of 1 kg/hr translated to a total heat load of 580 calories/hr. This heat load rate was deemed the maximum value that could be tolerated by healthy, acclimatized, young men for four to six hours per day. Also assumed was a 10-minute rest period every hour. Higher evaporative sweat rates translated to higher total heat loads, with the degree of heat load corresponding to the intensity of a given training exercise. Despite this indirect method of gauging exercise intensity, Yaglou and Minard recommended further research to determine work limits based on both WBGT and varying levels of work intensity.

The WBGT-based flag system introduced by Yaglou and Minard in 1954 at MCRD continued to develop during the period of 1956-1960. A green flag was added to the system, indicating WBGT values from 82-84.9°F (Minard, 1961). Training was not restricted during a green flag, but instructor staff was alerted to monitor for signs of heat strain in unacclimated recruits. Yellow flag actions were slightly modified, applying to the first three weeks of training, instead of only the first two weeks as previously required. Over the course of 1956-1960, the summer weekly incidence of heat casualties significantly decreased to 4.34 per 10,000, compared to 39.5 per 10,000 in 1952-1953, prior to the implementation of the WBGT-based flag system at MCRD (Minard, 1961). Despite this nine-fold reduction of heat casualties, Minard also had to demonstrate that prevention of heat casualties did not excessively interfere with training hours. Military training commands still had to achieve their mission of adequately preparing recruits for deployments to hot and humid environments. Thus, heat casualty prevention had to be balanced with building seasoned and productive troops. There would be no flag system in the middle of a battle in the Pacific. Subsequently, Minard calculated the percentage of total working hours in which there was a yellow or red flag raised. Yellow flag conditions only applied to recruits in their first three weeks of the 3-month training program to accommodate for their assumed unacclimated state, while red flag conditions affected all trainees regardless of acclimatization. At MCRD, approximately 20% of the training hours from 1956-1960 were under a yellow flag, while only 5%

were under a red flag (Minard, 1961). Since a yellow flag only applied to new recruits in the first three weeks of the 3-month program, only a small portion of trainees had training interrupted during these conditions. Thus, the WBGT-based flag system appropriately focused its restrictions on unacclimated recruits, while also permitting training in most conditions for lower risk, acclimatized troops. Given its minimal interference with training hours, along with a successful reduction of heat casualties, the WBGT-based flag system developed by Yaglou and Minard gained widespread acceptance in the United States military.

By 1980, WBGT had progressed to become the primary heat stress index used by the United States Army, Air Force, and Navy. As described in the Department of the Army's *Technical Bulletin MED 507* (1980), WBGT-based heat stress prevention measures had significantly evolved from the 1950s, with consideration of the metabolic rate associated with work, as well as requiring the addition of 10°F to the measured WBGT if military personnel wore body armor or nuclear, biological, chemical (NBC) protective attire. Three different categories of metabolic rate were defined: light, moderate, and heavy work. Metabolic rates for light work ranged from 63-164 kcal/hr (73-191 watts), corresponding to activities such as sleeping, performing desk work, or driving a vehicle. Moderate work activities ranged from 164-353 kcal/hr (191-411 watts), and included standing while performing work at a machine, or walking while performing moderate lifting. Heavy work ranged from 378-605 kcal/hr (440-704 watts), and was associated with intermittent heavy lifting or pushing such as pick and shovel work.

Given the variation of work rates during a period of time, a time-weighted-mean metabolic rate (t_{wm} MR) was calculated to determine mean work load. The t_{wm} MR's of 177, 223, and 270 watts were then associated with light, moderate, or heavy work load, respectively (Department of the Army, 1980). Subsequently, WBGT threshold values, in combination with a given mean work load category, defined maximum limits that would require additional heat stress preventive measures if exceeded. It should be noted that these WBGT threshold values applied only to industrial-type settings such as shipyards or machine shops, and did not apply to recruit training programs which still used the WBGT-based flag system. Threshold WBGT values were 86, 82, and 77°F for light, moderate, and heavy workloads, respectively. The WBGT threshold values pertained to the hottest two-hour period during a given work shift. For example, if a given shift involved a light work load, and the WBGT exceeded the 86°F threshold, then additional hot weather practices were required. These measures included minimum water requirements for a given shift based on mean work load and WBGT, modified work schedules during the acclimatization period, as well as a work-rest cycle of 25 minutes of work and 5 minutes of rest during moderately hot conditions. Thus, WBGT-based preventive measures had advanced considerably, incorporating multiple variables such as metabolic rate, attire, and water intake, while applying to other military work settings besides recruit training programs.

In 1991, the U.S. Army Research Institute of Environmental Medicine (USARIEM) made substantial contributions to the WBGT-based preventive measures used in the military. These recommendations coincided with the deployment of U.S. military personnel to the hot environment of the Middle East. The most notable advancement involved specifying maximum safe work times for a given WBGT, work level, and attire (USARIEM, 1991). These limits were depicted in tables and described the maximum number of continuous work minutes that could be sustained while minimizing the risk of heat casualties to 5%. One work limit table was designated for daytime operations, while the second table applied to nighttime operations. The work limits assumed that troops were fully hydrated, rested, and acclimatized, prior to performing activities. Additionally, work categories were modified from the *TB MED 507*, with the inclusion of very light work, and different corresponding work rate ranges, with very light, light, moderate, and heavy work levels associated with ranges of 105-175W, 172-325W, 325-500W, and 500+ W, respectively. Specific examples of work activities were identified for each category. Very light activities including lying on the ground or driving a truck. Light activities included cleaning weapons or walking on a hard surface at approximately 2.25 mph without a load. Moderate activities included calisthenics or walking on a hard surface at 3.5 mph without a load. Heavy activities included digging an emplacement or walking on hard surface at 3.5 mph with a 30kg load. Attire categories included the wearing of the regular battle dress uniform (BDU), mission oriented protective posture (MOPP) gear for protection against chemical and biological agents, or MOPP combined with BDU. Furthermore, more detailed water requirements than those in the *TB MED 507* were provided, pertaining to specific combinations of WBGT, work level, attire, and presence of sunshine, while being denoted by quarts per hour instead of the previously used quarts per work shift. From 1991 onward, USARIEM's extensive recommendations would serve as the primary guidance for heat stress prevention in the U.S. military, as well as provide the foundation for the current WBGT-based flag system.

The rationale for USARIEM's maximum work limits is derived from predictive equations developed by physiologists Givoni and Goldman in the 1970s. One equation predicts a final equilibrium rectal temperature based on metabolic and environmental heat loads, and evaporative heat exchange (Givoni & Goldman, 1972). Metabolic load was derived from multiple variables including metabolic rate, external work, body mass, clothing and equipment mass, and walking velocity. Environmental heat load was derived from dry bulb temperature, skin temperature, body surface area, and thermal resistance of a given article of clothing. Evaporative capacity was based on multiple factors including relative humidity and permeability index of the clothing. With this general formula, time patterns of core temperature could be predicted based on a given work rate and environmental conditions. A separate equation, similar to that for equilibrium rectal temperature, predicts an equilibrium heart rate based on metabolic rate, environmental conditions, and type of clothing worn, while also estimating heart rate patterns during work or recovery (Givoni & Goldman, 1973). In the 1980s, the Military Ergonomics Division of USARIEM devised a software program based on the aforementioned predictive equations for a portable hand-held calculator that could be used in the field. This program calculated maximum work time, optimal work-rest cycles, and water requirements based on multiple inputs such as clothing, metabolic rate, state of acclimation, ambient air temperature, wet bulb temperature, wind speed category such as calm or windy, solar heat load categories such as cloudy or clear, and acceptable heat casualty levels (Pandolf et al., 1986). The resulting collection of predictive formulas became known as USARIEM's Heat Strain Algorithm, which then evolved into the Heat Strain Decision Aid (HSDA) program after converting the formulas into more advanced computer programming language (SAIC, 1993).

The Current WBGT-based Flag System

The HSDA program serves as the basis for the WBGT-based flag system used currently by U.S. military forces. This flag system, as depicted in the Department of the Army's *TRADOC Regulation 350-29*, recommends sustainable continuous work limits based on different combinations of WBGT measurements and metabolic rates, with a corresponding flag color for 5 different heat categories (Department of the Army, 2016). The five colors, in order from lowest to highest WBGT range, are white, green, yellow, red, and black. The three categories of metabolic rate are easy, moderate, and hard work. There are also specific recommended water intake rates for each combination. There are several other assumptions with this flag system. Individuals are assumed to be wearing combat uniforms, and to have already been acclimatized after having performed two weeks of training. Also, individuals are presumed to be free of heat stress and dehydration prior to the designated work session, and to have a significant rest period after the activity. Of note, the limits recommended by this flag system assume no rest or other breaks taken during the work session.

The WBGT-based flag system may be compared to the Heat Hazard Assessment described in the Occupational Safety and Health Administration's (OSHA) Technical Manual section for heat stress (OSHA, 2017). The assessment entails a sequence of actions that will determine if the current environmental and work conditions present a significant heat stress risk. Like the military's WBGT-based flag system, WBGT and metabolic are both taken into account. However, there are several notable differences. First, the Heat Hazard Assessment applies ACGIH TLV screening criteria, which recommend WBGT limits based on workload levels, as well as work and rest percentages. These recommendations vary based on the acclimatization status of the worker, with a separate table of action limits for the unacclimated worker. Also, workload definitions differ from those in the military flag system. For example, moderate work for the military system is approximately 425 watts, whereas moderate work as defined by ACGIH is approximately 300 watts. Furthermore, the ACGIH TLV for heat stress adjusts for different types of clothing ensembles, utilizing a clothing adjustment factor to determine an effective WBGT.

Another heat stress prevention tool is the phone application developed by OSHA and NIOSH (NIOSH, 2018). This app lists real-time heat index measurements based on one's current location, and subsequently provides precaution recommendations corresponding to the given heat stress risk level. The app also provides detailed guidance on heat stress symptoms, as well as first aid recommendations. With its phone accessibility and user friendly interface, the app provides a convenient tool for gauging potential heat stress conditions in the field. However, there may be several disadvantages to the app. Unlike both the military flag system and the ACGIH TLV for heat stress, the app is unable to factor in varying levels of workload. Also, the app utilizes heat index rather than WBGT, which may lend itself to a more incomplete assessment of the outdoor environmental conditions. The app does not consider acclimatization status, nor does it have any recommendations for water intake rates like the military flag system.

The military's WBGT-based flag system for continuous work may offer several advantages to other heat stress prevention tools. Use of flag colors may more readily convey heat stress risk to workers, as opposed to using effective WBGT numbers with the ACGIH TLV. Also, the military flag system uses three workload levels, as opposed to five levels of the ACGIH TLV, which may simplify categorization of different work tasks for supervisors. The flag-based system also recommends sustainable work duration limits, which may be beneficial for supervisors who need to schedule work tasks or for those concerned about impact on productivity. Despite these potential advantages, there may be concerns about the flag system oversimplifying workload categories. Additionally, unlike the ACGIH TLV, the flag system assumes only one work attire and does not consider various clothing ensembles. Furthermore, the military-based flag system is based primarily on data from military personnel since the 1950s, rather than the general population. The applicability of such a system to non-military populations, in which individuals may have chronic medical conditions or not be as physically fit, is a legitimate concern. Nevertheless, if appropriately sensitive and specific, the flag system may serve as a practical tool for industries in which workers are regularly outdoors, such as agriculture and construction. The construction and agriculture industries comprise the highest proportions of heat-related deaths among all industries, at 36.8% and 21.0%, respectively (Gubernot, Anderson, & Hunting, 2015). The present paper analyzes the ability of the military's WBGT- based flag system to recommend safe heat stress exposures in a non-military population.

CHAPTER TWO:

METHODS

One portion of data for the present paper originated from two prior studies using a progressive heat stress protocol at the University of South Florida (USF) (Bernard et al., 2005, 2008). The other data were taken from trials using a time-limited protocol performed at USF (Bernard & Ashley, 2009). All three studies were approved by the USF institutional review board. The two studies using a progressive heat stress protocol involved a total of 29 non-military participants, each performing multiple trials. These studies resulted in 528 observations of compensable and uncompensable heat stress exposures. The progressive heat protocol entailed participants walking on a motorized treadmill in a controlled climatic chamber. Each trial began in comfortable conditions that were easily sustainable. Temperature and humidity were then slowly increased in 5-min intervals, with ambient conditions monitored continuously and recorded every 5 minutes. The speed and grade of the treadmill were set to the desired workload. Actual metabolic rate was estimated by analyzing oxygen consumption via expired gases, which was sampled every 30 minutes during a trial. The average of these values was designated as the metabolic rate for the entire trial. Core temperature of participants was measured with a rectal thermistor, which was continuously monitored and recorded every 5 minutes. Each trial was designated to last 120 minutes unless one of the following criteria were met: (1) a clear rise in T_{re} associated with a loss of thermal equilibrium; (2) T_{re} reached 39° C; (3) a sustained heart rate greater than 90% of the age-predicted maximum heart rate; (4) participant wished to stop (Bernard et al., 2005).

Participants were acclimatized over the course of 5 consecutive days via 2-h exposures to dry heat at a metabolic rate of 160 W m^{-2} while wearing shorts, a tee-shirt, socks, and shoes. On the day of a trial, participants refrained from consuming caffeinated beverages and from performing vigorous exercises within 3 hours before the start of a trial. Each participant was healthy without any chronic medical issues, and cleared by a physician prior to performing in the trials. Characteristics of the participants from the progressive and time-limited protocols are summarized in Figures 1 and 2, respectively.

For trials using the progressive heat stress protocol, compensable and uncompensable points were marked from the recordings of core temperature (Figure 3). The compensable point marked conditions at which thermal equilibrium could be maintained. The uncompensable point marked conditions for which a physiological steady state could not be maintained as evidenced by increasing core temperature. The critical point was denoted as the point during the trial 5 minutes prior to the steady increase in the core temperature. Compensable/Sustainable conditions are those in which thermal equilibrium could maintained for at least 4 hours, assuming adequate water intake.

The time-limited protocol involved 62 trials performed by 12 participants. For each trial, the participant was subjected to a constant WBGT for the entirety of the session. The participant would continue activity on the treadmill until one of the following conditions was met: (1) T_{re} reached 38.5°C; (2) a sustained heart rate greater than 85% of the age-predicted maximum heart rate; or (3) participant wished to stop (Bernard & Ashley, 2009). The duration of the trial from onset to the fulfillment of one of the aforementioned criteria was considered the safe exposure time for the trial. Specific WBGT conditions were chosen with the intent of participants experiencing a loss of thermal equilibrium throughout the trial and of satisfying one of the three stop criteria prior to reaching the 120-min mark.

Statistical Analyses

Observations from the progressive protocol trials were compared to recommendations by the military's flag-based system in the following manner. Stated average metabolic rates for easy, moderate, and hard work are 250, 425, and 600 W, respectively (Department of the Army, 2016). Thus, each work category is separated by 175 W. To determine a range for each work category, the difference between levels was divided by two to approximate a midpoint. Midpoints functioned as the upper limit for the preceding level, and as the lower limit for the higher level, resulting in three ranges of work for the flag system (Figure 4).

Conditions marked as 'NL' in the flag system table were classified as Safe, indicating that work can be sustained at that particular WBGT and metabolic rate for at least 4 hours at a physiological steady state. In contrast, the conditions of the flag system not designated as 'NL' all have recommended limits of 180 minutes or less. These other conditions are not expected to permit maintenance of thermal equilibrium. Thus, all conditions not marked as 'NL' would be recommended as Unsafe in the setting of a long duration exposure in which thermal equilibrium could not be maintained. Data from the progressive protocol included an average metabolic rate and the WBGT values at the compensable and uncompensable points for each trial. A classification of Safe was designated for each combination of the trial's average metabolic rate and the WBGT at the compensable point. The combination of a trial's metabolic rate and the WBGT observed at the uncompensable point was classified as Unsafe. Each combination of metabolic rate and WBGT was mapped to the corresponding combination in the military flag table. Flag recommendations were then compared to observed outcomes to generate a 2x2 table, allowing for calculation of specificity and sensitivity, as well as a Kappa statistic.

In the time-limited protocol, the flag table recommended a maximum safe exposure time for a given combination of WBGT and metabolic rate, which was then compared to the observed duration of activity in the following manner. A recommendation of Unsafe was designated for trial durations that exceeded the flag's maximum limit, while a recommendation of Safe was assigned to trial durations that were shorter than the flag's maximum limit. An outcome of Unsafe was designated for trials in which the observed safe exposure time was less than the recommended limit, while Safe outcomes were assigned to trials in which the observed safe exposure time was greater than the recommended limit. Flag recommendations were then plotted against trial outcomes to generate a 2x2 table. Further statistical analysis of observed durations was performed using SPSS, comparing recommended exposure times to observed durations.

	N	Age (years)	Height (cm)	Weight (kg)	Body surface area (m²)
Relative humidity study (Bernard *et al.*, 2005)					
Men	9	29 ± 6.8	183 ± 6	97.2 ± 18.5	2.18 ± 0.20
Women	5	31.8 ± 9.1	161 ± 7	63.5 ± 17.2	1.66 ± 0.23
Metabolic rate study (Bernard *et al.*, 2008)					
Men	11	28 ± 10	176 ± 11	81.7 ± 12.0	1.98 ± 0.47
Women	4	23 ± 5	165 ± 6	64.2 ± 18.0	1.70 ± 0.22
Pooled					
Men	20	29 ± 9	179 ± 34	88.7 ± 23.2	2.07 ± 0.41
Women	9	28 ± 8	163 ± 7	63.7 ± 16.6	1.74 ± 0.29

Figure 1. Characteristics of participants from the progressive protocol. Adapted with permission from "Ability to Discriminate Between Sustainable and Unsustainable Heat Stress Exposures – Part 1: WBGT Exposure Limits," by X.P. Garzón-Villalba et al., 2017, *Annals of Work Exposures and Health*.

	Number	**Age (yr)**	**Height (cm)**	**Weight (kg)**	**Body Surface Area (m^2)**
Men	8	33 ± 10	181 ± 4	95 ± 10	2.15 ± 0.09
Women	4	28 ± 9	160 ± 7	66 ± 27	1.67 ± 0.33
All	12	32 ± 10	174 ± 11	85 ± 22	1.99 ± 0.30

Figure 2. Characteristics of participants from the time-limited protocol. Adapted with permission from "Short-term heat stress exposure limits based on wet bulb globe temperature adjusted for clothing and metabolic rate," by T.E. Bernard and C.D. Ashley, 2009, *Journal of Occupational and Environmental Hygiene*.

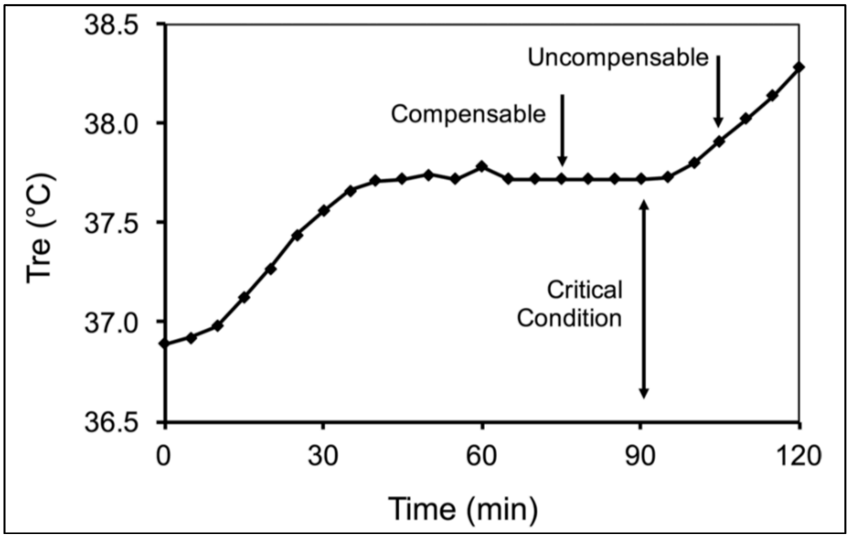

Figure 3. Time course of a progressive protocol trial. Adapted with permission from “Ability to Discriminate Between Sustainable and Unsustainable Heat Stress Exposures – Part 1: WBGT Exposure Limits,” by X.P. Garzón-Villalba et al., 2017, *Annals of Work Exposures and Health*.

WORK CATEGORY	RANGE
Easy	< 337.5 W
Moderate	337.6-512.5 W
Hard	> 512.6 W

Figure 4. Modified metabolic rate ranges for the flag system.

CHAPTER THREE:

RESULTS

Table 1 provides a summary of the number of matching combinations of WBGT and metabolic rate between the flag system and the 528 observations during the progressive protocol trials. The majority of observed conditions from the progressive protocol trials were classified in the flag system's black WBGT category, and mostly in the easy and moderate work levels. Table 2 summarizes the observations from the time-limited protocols, with all trials falling under the black WBGT level, and divided among easy and moderate work categories. Table 3 illustrates an overall 2x2 table comparing the flag system's recommendations of safe conditions to the observed conditions of uncompensable heat stress from the progressive protocol trials. Sensitivity and specificity from this table were 0.98 and 0.25, respectively. Calculated Kappa score was 0.22, indicating fair agreement.

Table 4 illustrates an overall 2x2 table comparing the flag system's recommendations of safe exposure time to the observed durations from the time-limited trials. The majority of observed durations were shorter than the flag system's recommended limit, with the remainder of observed durations exceeding the flag system's recommended limit. Sensitivity and specificity for the time-limited protocol data were both zero, with a Kappa score of zero. Table 5 illustrates a 2x2 table that combines the observations from the progressive and time-limited protocols. Sensitivity and specificity for this data was 0.81 and 0.24, respectively, indicating that inclusion of the time-limited protocol data decreased the sensitivity of flag system recommendations. Kappa score for the combined data was 0.05, indicating poor agreement. Table 6 provides an overview of the sensitivities, specificities, and Kappa scores for all the protocols. Table 7 summarizes the mean observed WBGT exposures, as well as the recommended and observed trial durations from the time-limited protocol dataset. Observed durations were significantly shorter than those recommended by the flag system. 5th percentile durations are shown to indicate the lower limit of 95% of the observed durations for each condition.

Table 1. Summary of Observations from Progressive Protocol.

	EASY	MODERATE	HARD	**TOTAL**
WHITE	3	3	2	**8**
GREEN	9	21	4	**34**
YELLOW	25	22	2	**49**
RED	34	26	3	**63**
BLACK	232	138	4	**374**
TOTAL	**303**	**210**	**15**	**528**

Table 2. Summary of Observations from Time-Limited Protocol.

	EASY	MODERATE	HARD	**TOTAL**
WHITE	0	0	0	**0**
GREEN	0	0	0	**0**
YELLOW	0	0	0	**0**
RED	0	0	0	**0**
BLACK	17	45	0	**62**
TOTAL	**17**	**45**	**0**	**62**

Table 3. 2x2 Table for Progressive Protocol.

PROGRESSIVE PROTOCOL		OBSERVED	
		Unsafe	Safe
FLAG	Unsafe	249	205
RECOMMENDATION	Safe	6	68

Table 4. 2x2 Table for Time-Limited Protocol.

TIME-LIMITED PROTOCOL		OBSERVED	
		Unsafe	Safe
FLAG	Unsafe	0	10
RECOMMENDATION	Safe	52	0

Table 5. 2x2 Table for Combined Protocols.

COMBINED		OBSERVED	
		Unsafe	Safe
FLAG	Unsafe	249	215
RECOMMENDATION	Safe	58	68

Table 6. Summary of Sensitivities, Specificities, and Kappa Scores.

	Progressive	Time-limited	Combined
Sensitivity	0.98	0	0.81
Specificity	0.25	0	0.24
Kappa	0.22	0	0.05

Table 7. Summary of Exposures and Durations from Time-Limited Protocol.

Condition	Mean Observed WBGT (°C)	Recommended Duration (min)	Mean Observed Duration (min)	SD	Mean Difference in Duration (min)	P value	5th Percentile Duration (min)
Black/Easy	38.9	180	43	18	-137	<0.0005	15
Black/Mod	39.1	70	54	23	-17	<0.0005	22

CHAPTER FOUR:

DISCUSSION

Analysis of the observations from the progressive protocol trials suggests that the flag system has high sensitivity but low specificity when recommending safe heat stress exposures in a non-military population. There were a high number of false positives, indicating conditions recommended as unsafe but actually observed to be safe. There were few false negative observations, signifying conditions recommended as safe but actually observed to be unsafe. For heat stress recommendations, both high sensitivity and high specificity would be ideal. However, high sensitivity at the cost of lower specificity, as was the case for the progressive protocol, is suitable as this would result in a protective system that errs on the side of safety. In other words, a false positive recommendation in which an individual does not experience heat stress despite being expected to, may be preferred to a false negative recommendation in which an individual suffers from heat stress after being told conditions are safe.

In contrast, for a time-limited protocol, the flag system appears to have both low sensitivity and specificity when recommending safe exposures. There were a high number of false negatives, which entailed the flag system overestimating the duration of time a participant could safely perform activity at a given WBGT and metabolic rate. All participants exposed to easy work at a black WBGT level had observed durations significantly shorter than the recommended limit of 180 minutes. The majority of participants exposed to moderate work at a black WBGT level also had safe exposure times significantly shorter than the maximum recommended limit of 70 minutes. Thus, participants largely experienced heat stress at durations they were not expected to do so.

However, analysis of the observed WBGT exposures may explain the unexpectedly low sensitivity and specificity of the flag system for the time-limited protocol. Mean observed WBGT exposure for both easy and moderate work at the black flag level was approximately 39.0°C (102.2°F). This is 7°C more than the threshold WBGT for a black flag level exposure of 32.0°C (90°F). Accordingly, participants were largely exposed to substantially higher WBGT's than the threshold for a black flag exposure. Currently, there is no upper limit for the black WBGT category in the flag system. The other four flag categories lower than black are separated by only 1-2°C. With this degree of stratification, a WBGT of 39°C would be approximately two to three categories beyond the black flag threshold. Recommended safe exposure times for a 3-fold higher flag category may be significantly lower than the current limits for the black flag level. Thus, the extreme WBGT exposures observed in the time-limited trials may have accounted for the unexpected performance of the flag system.

Analysis of the observed durations from the time-limited protocol showed significantly lower durations than recommended, with a mean difference of 137 min for easy work at a black level, and 17 min for moderate work at a black level. The 5^{th} percentile duration for easy work at a black level was shorter than that for moderate work at a black level by 7 minutes. A longer 5^{th} percentile duration would have been expected for easy work compared to moderate, but the smaller sample size of 17 trials for easy work compared to 45 for moderate may have contributed to an underestimation of duration. To increase sensitivity of the flag system and theoretically protect 95% of participants from heat stress, the recommended maximum durations by the flag system could be adjusted to the 5^{th} percentile durations. Thus, instead of recommended limits of 180 min and 70 min for easy and moderate work, respectively, the 5^{th} percentile durations of 15 min for easy work and 22 min for moderate work would serve as the maximum limits. Given the likely underestimation of the 5^{th} percentile for easy work, a more appropriate limit would be greater than

22 min for moderate work, although determining a more precise threshold warrants further investigation.

Applying these recommended durations to non-military industries may prove problematic. Working for 15-22 minutes, and then resting for the remainder of the hour, may have a negative impact on productivity. Less protective threshold limits, such as the 50th percentile duration, may be more desirable by employers, or even workers themselves if incentive pay is present. These higher percentile durations equate to longer recommended durations of sustainable work, but may unnecessarily expose workers to unsustainable heat stress conditions. The 5th percentile durations may be more appropriate as a short-term exposure limit or as the work duration in a work-rest cycle per hour. Additionally, workers may have a myriad of medical conditions or engage in substance abuse that may be detrimental to their capacity for heat stress. An obese diabetic or an alcoholic may be chronically dehydrated, increasing their risk of heat stress, particularly if they are seasonal workers beginning work in a hot and humid environment in an unacclimated state. A flag system compared to a non-military study population that was generally healthy may lack the sensitivity for such high-risk workers. More research may be needed to more precisely determine safe limits for a general population inclusive of those with chronic medical conditions.

One potential limitation of the present study involves the different acclimatization periods. Study participants were acclimatized over the course of five days, whereas the flag system presumes individuals have acclimated over the course of two weeks. Study participants may not have been fully acclimatized to hot and humid conditions prior to undergoing trials. This may have led to observed durations of safe exposure time that were shorter than could have been achieved if the participant acclimated over a greater period of time. Nevertheless, the study benefitted from having a generally healthy sample of participants, likely comparable to military recruits and active duty personnel on whom the flag system recommendations are based. Additionally, the protocols used in the study were meticulous in the monitoring and recording of core temperature, metabolic rate, temperature, and humidity. Such direct measurements, rather than depending on subjective symptom reporting, optimized the determination of changes in thermal equilibrium.

CHAPTER FIVE:

CONCLUSION

In its current form, the military's WBGT-based flag system has high sensitivity for long duration heat exposures, albeit with low specificity. Such a configuration can be overprotective, but this may be desired in order to minimize false negative recommendations and err on the side of safety. False negatives represent a worst case scenario, in which individuals are recommended to work in conditions that are actually unsafe. A high number of false negatives were observed when the flag system was applied to trials from the time-limited protocol. However, a plausible explanation for this unexpected result involves the substantially higher WBGT exposures experienced by participants than are currently classified by the flag system. Such a finding warrants further investigation into potential classifications of heat categories beyond the black flag level. Despite the mixed results from this study, the military's WBGT-based flag system can be an effective heat stress preventive tool for a non-military workforce. The flag system may be particularly helpful in situations where workers are subjected to long durations of heat exposure at intermediate WBGT ranges. These environmental conditions may already be commonplace in industries such as agriculture and construction, where more comprehensive heat stress preventive measures are needed to reduce heat-related fatalities.

REFERENCES

Bernard, T.E., & Ashley, C.D. (2009). Short-term heat stress exposure limits based on wet bulb globe temperature adjusted for clothing and metabolic rate. *Journal of Occupational and Environmental Hygiene, 6*(10), 632-638. https://doi.org/10.1080/15459620903133642

Bernard, T.E., Caravello, V., Schwartz, S.W., & Ashley, C.D. (2008). WBGT clothing adjustment factors for four clothing ensembles and the effects of metabolic demands. *Journal of Occupational and Environmental Hygiene, 5*(1), 1-5. https://doi.org/10.1080/15459620701732355

Bernard, T.E., Luecke, C.L., Schwartz, S.W., Kirkland, K.S., & Ashley, C.D. (2005). WBGT clothing adjustments for four clothing ensembles under three relative humidity levels. *Journal of Occupational and Environmental Hygiene*, *2*(5), 251-256. https://doi.org/10.1080/15459620590934224

Department of the Army. (2016). *TRADPC 350-29: Prevention of Heat and Cold Casualties* [PDF file]. Retrieved from https://adminpubs.tradoc.army.mil/regulations/TR350-29.pdf

Department of the Army. (1980). *TB Med 507: Prevention, treatment, and control of heat injury* [PDF file]. Retrieved from https://www.med.navy.mil/directives/Pub/5052-5.pdf

Garzón-Villalba, X.P., Wu, Y., Ashley, C.D., & Bernard, T.E. (2017). Ability to discriminate between sustainable and unsustainable heat stress exposures – Part 1: WBGT exposure limits. *Annals of Work Exposures and Health, 61*(6), 611-620. https://doi.org/10.1093/annweh/wxx034

Givoni, B., & Goldman, R.F. (1973). Predicting heart rate response to work, environment, and clothing. *Journal of Applied Physiology, 34*(2), 201-204. https://doi.org/10.1152/jappl.1973.34.2.201

Givoni, B. & Goldman, R.F. (1972). Predicting rectal temperature response to work, environment, and clothing. *Journal of Applied Physiology, 32*(6), 812-822. https://doi.org/10.1152/jappl.1972.32.6.812

Gubernot, D.M., Anderson, G.B., & Hunting, K.L. (2015). Characterizing occupational heat-related mortality in the United States, 2000-2010: An analysis using the Census of Fatal Occupational Injuries database. *American Journal of Industrial Medicine, 58*(2), 203-211.

Minard, D. (1961). Prevention of heat casualties in Marine Corps recruits: Period of 1955-60, with comparative incidence rates and climatic heat stresses in other training categories. *Military Medicine, 126*, 261-272. https://doi.org/10.1093/milmed/126.4.261

National Institute for Occupational Safety and Health. (2018, Jun 6). OSHA-NIOSH Heat Safety Tool App. Retrieved from https://www.cdc.gov/niosh/topics/heatstress/heatapp.html

National Institute for Occupational Safety and Health. (2016). *Criteria for a Recommended Standard, Occupational Exposure to Heat and Hot Environments* [PDF file]. Retrieved from https://www.cdc.gov/niosh/docs/2016-106/pdfs/2016-106.pdf

Occupational Safety and Health Administration. (2017, Sep 15). Section III: Chapter 4, Heat Stress. Retrieved from https://www.osha.gov/dts/osta/otm/otm_iii/otm_iii_4.html

Pandolf, K.B., Stroschein, L.A., Drolet, L.L, Gonzalez, R.R, and Sawka, M.N. (1986). Prediction modeling of physiological responses and human performance in the heat [PDF file]. *Computers in Biology and Medicine,* (6), 319-329. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a160513.pdf

Science Applications International Corporation (SAIC). (1993). *P^2NBC^2 Heat Strain Decision Aid Users Guide, Version 2.1* [PDF file]. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a392286.pdf

United States Army Research Institute of Environmental Medicine (USARIEM). (1991). *Heat illness: A handbook for medical officers* [PDF file]. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a238974.pdf

Yaglou, C., & Minard, D. (1956). *Prevention of Heat Casualties at Marine Corps Training Centers* [PDF file]. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/099920.pdf

APPENDIX A:

PERMISSION STATEMENT

Permissions Statement

Bernard, Thomas
Mon 4/8/2019 2:55 PM
To: Almario, David Sean <dalmario@health.usf.edu>

To Whom It May Concern via Dr. David Almario

This email is intended to serve as granting of permission to MSPH candidate, David Almario, to use the figures previously published in articles for which I am a co-author. These articles are cited below. If there are any questions, I can be reached by email at tbernard@health.usf.edu.

Garzón-Villalba, X.P., Wu, Y., Ashley, C.D., & Bernard, T.E. (2017). Ability to discriminate between sustainable and unsustainable heat stress exposures – Part 1: WBGT exposure limits. *Annals of Work Exposures and Health, 61*(6), 611-620. https://doi.org/10.1093/annweh/wxx034

Bernard, T.E., & Ashley, C.D. (2009). Short-term heat stress exposure limits based on wet bulb globe temperature adjusted for clothing and metabolic rate. *Journal of Occupational and Environmental Hygiene, 6*(10), 632-638. https://doi.org/10.1080/15459620903133642

Thank you for the consideration

tb
--
Thomas E. Bernard
University of South Florida, College of Public Health
13201 Bruce B. Downs Blvd, Tampa FL 33612-3805
813-974-6629 / http://personal.health.usf.edu/tbernard/
"our practice is our passion"