Annals of Work Exposures and Health, 2023, **67**, 345–353
https://doi.org/10.1093/annweh/wxac083
Advance access publication 1 December 2022
Original Article

The Chartered Society for Worker Health Protection

OXFORD

Group Outcomes for Time-Weighted Averaging in WBGT-Based Heat Stress Exposure Assessment

Thomas E. Bernard[1,*], John W. Flach[1,†] and Candi D. Ashley[2]

[1]College of Public Health, University of South Florida, Tampa, FL 33612, USA;
[2]Exercise Science Program, College of Education, University of South Florida, Tampa, FL 33620, USA

*Author to whom correspondence should be addressed. Tel: 1-813-974-6629; e-mail: tbernar2@usf.edu
†LCDR Flach is currently serving the USCG at HSWL SC (se-fo) Det Cleveland, 1240 E 9th Street, STE 2101, Cleveland, OH 44199, USA.

Abstract

The wet bulb globe temperature (WBGT)-based occupational exposure limits (OELs) were developed from steady exposures to heat stress at constant WBGT and metabolic rate (M). The exposure limits were based on compensable heat stress exposures at the upper limit of the prescriptive zone for most healthy people. Professional practice allows for using time-weighted averages (TWAs) of WBGT and M to account for heterogeneous heat stress exposures. The purpose of the current paper was to report on the effectiveness of time-weighted averaging to assess occupational heat stress using published studies. Our hypothesis was using TWA-WBGT and TWA-M was as protective as the recommended OELs for steady exposures. The current paper reports on 62 observations of work that alternate between at least two heat stress conditions (usually work and recovery) reported in 16 papers. The TWA-WBGT and TWA-M were determined for all observations. ΔLimit was the observed TWA-WBGT minus the exposure limit at the TWA-M based on acclimatization state. The observations were then classified as above or below ΔLimit = 0. Each observation was also classified as uncompensable if the mean core temperature for the group was greater than 38°C or a less tolerant individual was above 38.5°C. When comparing exposure classifications to outcome classifications using 2 × 2 tables, the sensitivity and specificity for all observations were 0.72 and 0.73, respectively. The sensitivity was much less than the expected value near 1.0, and the large difference called into question the ability of TWAs to represent actual heat stress. There was some suspicion that there were differences between acclimatized and unacclimatized observations. Before any of these findings are embedded in policy or practice, a more careful evaluation of TWAs is required. In conclusion, we believe that the use of TWAs for heat stress analysis was not fully evaluated, and we proposed a framework for evaluation.

Keywords: heat stress indices; TWA; work rest cycles

What's Important About This Paper

In the use of WBGT-based occupational exposure limits, it is often necessary to use time-weighted averaging (TWA) for both assessment and for the prescription of work/rest cycles. While a common practice of long standing, using TWAs have never been systematically evaluated to determine if the process represents the actual heat stress. This analysis of existing studies suggests that TWAs may not be adequately protective, and this study describes an approach to evaluation.

Introduction

Heat stress assessment often includes environmental conditions, metabolic rate, and clothing as well as acclimatization state. A wet bulb globe temperature (WBGT)-based occupational exposure limit (OEL) for acclimatized, healthy workers is described by the ACGIH® TLV® (ACGIH®, 2022), the NIOSH recommended exposure limit (REL) (NIOSH, 2016) and the ISO 7243 exposure limit (ISO, 2017). (Nomenclature is described in Table 1.) The threshold WBGT is reduced with increasing

Received: April 7, 2022. Accepted: November 2, 2022.

Table 1. Nomenclature. This table provides the crosswalk from abbreviation to description for classifications used in the analysis and for abbreviations used throughout the text.

Classifications	Description
Accl state	Acclimatization state
A	Acclimatized
UA	Unacclimatized
Clothing	Ensemble worn during trials
SN	Semi-nude (e.g. shorts and shoes)
SC	Semi-clothed (e.g. shorts, tee-shirt/sports bra, shoes)
WC	Work/woven clothing (e.g. long sleeve shirt and trousers, modesty clothing underneath, and shoes)
VB	Vapor-barrier coveralls
MWU	Modified work uniform: test garment in one study with modified wicking capability
WBGT	Classification of exposures as same or different WBGT
M	Classification of exposures as same or different metabolic rate
ΔLimit	The elevation above the OEL for acclimatized participants or above the OAL for unacclimatized participants in °C-WBGT (continuous variable)
Above	Exposure classification when ΔLimit > 0
Below	Exposure classification when ΔLimit ≤ 0
Uncompensable	Decision based on final core temperature (may be projected to 3 h)
Yes	Core temperature for the mean and/or 95th percentile exceeds 38 or 38.5°C, respectively.
No	Core temperature for the mean and 95th percentile does not exceed 38 or 38.5°C, respectively.
Abbreviations	
ACGIH	ACGIH (formerly American Conference of Governmental Industrial Hygienists)
AL	Action limit of ACGIH for unacclimatized workers
CAV	Clothing adjustment value (°C) from the ACGIH TLV for heat stress and strain (2022, Notice of Intended Change)
ISO	International Organization for Standardization
M	Metabolic rate in Watts
NIOSH	National Institute for Occupational Safety and Health
OAL	Occupational alert/action limit defined by the limits for unacclimatized workers (e.g. ACGIH AL, NIOSH RAL)
OEL	Occupational exposure limit defined by the limits for acclimatized workers (e.g. ACGIH TLV, NIOSH REL, ISO)
RAL	NIOSH recommended alert limit
REL	NIOSH recommended exposure limit
TLV	Threshold limit value of ACGIH for acclimatized workers
TWA	Time-weighted average
TWA-M	Time-weighted average of metabolic rate
TWA-WBGT	Time-weighted average of WBGT
ULPZ	Upper limit of prescriptive zone
WBGT	Wet bulb globe temperature index in °C

metabolic rate. To account for lower heat tolerance among unacclimatized workers, the OEL is reduced for the action limit (AL) in ACGIH, recommended alert limit (RAL) in NIOSH, and unacclimatized in ISO. We will refer to the limit for unacclimatized as the occupational alert limit (OAL). As a note, there are no differences among the OELs and the OALs for ACGIH, NIOSH, and ISO. Because clothing can alter heat stress, both ACGIH® and ISO use a clothing adjustment value (CAV) to adjust the observed WBGT by adding the appropriate CAV (ISO, 2017; ACGIH®, 2022).

The heat stress OEL was developed and validated using laboratory trials at constant environmental conditions and metabolic rates (M) and designed to be protective for more than 95% of exposures at the upper limit of the prescriptive zone (ULPZ) (Lind, 1963a, 1970; Dukes-Dobos and Henschel, 1973). Often heat stress exposures are not comprised of a single, constant WBGT and M. In the special case where exposure conditions consist of work and rest in the same environment, ACGIH provides a table of screening criteria (see Table 2 in TLV) (ACGIH®, 2022) and NIOSH provides four curves for both the REL and the RAL based on the proportion of work in an hour with the balance being rest in the same environment (see NIOSH criteria document Figures 8-1 for RAL and 8-2 for REL) (NIOSH, 2016). (Note: There are unintentional discrepancies between the equations used to describe the REL and RAL and the figures; and the equations are taken to be the intended limit.) The ACGIH and intended NIOSH exposure limits were based on a time-weighted average (TWA) of M assuming the same WBGT. The purpose was to provide a simple, semi-quantitative method to assess heat stress exposures. Quantitative exposure assessment was based on the OEL and OAL equations and the associated curves as described by ACGIH, ISO, and NIOSH documentation (NIOSH, 2016; ISO, 2017; ACGIH®, 2022).

For exposures that include different WBGTs, different Ms, or both, TWA of WBGT (TWA-WBGT) and M (TWA-M) over a period of about 1 h was recommended (Dukes-Dobos and Henschel, 1973) and adopted by ACGIH and NIOSH. This recommendation was based on one study by Lind (1963b) that looked at work and rest in a constant environment in two participants. In effect, the Lind study was a suggestion that TWA may work; it does not represent a validation.

In general, TWA scenarios fit into four combinations of two categories of environment (same or different WBGTs) and two categories of metabolic rate (same or different Ms). For a common scenario, work and rest can occur in the same environment for a combination of same WBGT and different Ms. Another scenario for work and rest is work in a hot environment and rest in a cool environment for a combination of different WBGTs and different Ms. Another reasonable pair would be different WBGTs but the same M, which would be the same metabolic rate in different work locations with different WBGTs. The fourth combination of same WBGT and same M is the steady condition.

University of Ottawa investigators have raised concerns that the exposure limits for work and recovery in the same environment based on TWA-M are not protective (Wright Beatty *et al.*, 2015; Meade *et al.*, 2016; Lamarche *et al.*, 2017; Seo *et al.*, 2019; Kaltsatou *et al.*, 2020). While investigators have examined some

Table 2. 2 × 2 tables with number of observation (%) in each cell for all observations; acclimatized only; unacclimatized only; for same WBGT; and for all without projections. Included are Pearson χ^2 outcome, sensitivity, and specificity. See footnote for comparisons between selected groups.[a]

ΔLimit = 0	Uncompensable	Compensable	Sensitivity	Specificity
All (χ^2: $P < 0.001$)				
Above	23 (37)	8 (13)	0.72	0.73
Below	9 (15)	22 (35)		
All without projections (χ^2: $P < 0.05$)				
Above	11 (18)	20 (32)	0.79	0.58
Below	3 (5)	28 (45)		
Acclimatized (χ^2: $P < 0.01$)				
Above	7 (37)	4 (21)	1.0	0.67
Below	0	8 (42)		
Unacclimatized (χ^2: $P < 0.01$)				
Above	16 (37)	4 (9)	0.64	0.78
Below	9 (21)	14 (33)		
Same WBGT (χ^2: $P < 0.01$)				
Above	20 (43)	7 (15)	0.77	0.67
Below	6 (13)	14 (30)		

[a]Selected comparisons using https://www.statology.org/chi-square-test-of-independence-calculator/ to compare two groups over four categories:
All versus all without projections: χ^2: $P < 0.01$.
Acclimatized versus unacclimatized: χ^2: $P = 0.13$.
All versus same WBGT: χ^2: $P = 0.90$.

aspects of work and rest cycles (same WBGT/different Ms and different WBGTs/different Ms), we are not aware of a systematic evaluation of TWA for all combinations of WBGT at two (or more) levels and M at two (or more) levels to assess heat stress exposures. The purpose of the current paper was to test the effectiveness of time-weighted averaging to assess occupational heat stress using a sample of published studies. Our working hypothesis was that using TWA-WBGT and TWA-M for exposure assessment will lead to compensable heat stress for most people below the appropriate exposure limit.

Methods

The data for this study were from a sample of peer-reviewed papers in which the study design fit one of three pairs of WBGT and M categories: (i) same WBGT and different Ms; (ii) different WBGTs and different Ms; and (iii) different WBGTs and same M. Other criteria for inclusion were experimental conditions lasting at least 90 min and the time for a complete cycle was less than an hour, reported exposure conditions with sufficient information to assign a TWA-WBGT (°C) and a TWA-M (Watts), and reporting of core temperature (most often rectal temperature) to determine the exposure condition (compensable or uncompensable). In all cases, the reported WBGTs and Ms were for the group; that is, individual exposure data were not available. Sixteen papers were identified and are referenced in Supplementary Table S (available at *Annals of Work Exposures and Health* online).

For each paper, the number and acclimatization state of participants, and clothing worn were noted. For acclimatization state, if the paper did not explicitly state that the participants were acclimatized, they were considered unacclimatized; a common occupational hygiene practice to be protective. Clothing worn was categorized as semi-nude (shorts and socks with shoes), semi-clothed (shorts, tee-shirt/halter top with socks and shoes), or single layer of woven clothing. One study reported on a vapor-barrier coverall (Seo *et al*., 2019) and another reported on a modified work uniform (Stapleton *et al*., 2012). A CAV was assigned to each clothing configuration as follows. A CAV of −1.6°C was used for semi-nude based on our analysis of critical conditions between semi-nude and work clothing (Belding and Kamon, 1973). For semi-clothed, CAV was −1.0°C as described in the 2022 TLV Notice of Intended Change (ACGIH®, 2022). For vapor-barrier coveralls, CAV of 11°C was assigned following the TLV (ACGIH®, 2022). For the modified work uniform, a CAV of −0.5°C was estimated based on the data in the paper (Stapleton *et al*., 2012) and our professional judgment. While the CAV is usually added to the observed value of WBGT to represent the effective WBGT (ACGIH®, 2022), we subtracted it from the exposure limit, which was effectively equivalent.

The study design within a paper often included more than one experimental condition which included different work and rest periods, different resting environments, or different Ms during work. Each exposure condition within a paper was treated as a separate observation. For a given observation, the work times and rest times, sufficient environmental data to determine WBGT, and Ms were noted for each period within the cycle. If water vapor pressure or relative humidity was reported in lieu of a psychrometric wet bulb temperature, psychrometric wet bulb was calculated, then the natural wet bulb was estimated following the method described in ISO 7243, Annex D (ISO, 2017). (An Excel® workbook is available on request, which has the data and macros used to estimate environmental parameters.) In all the papers, the M for work was reported as representative of the group. Because M for rest was often not reported, 100 W was used. The resting metabolic rate was estimated as an average value based on the US adult population of about 1800 kcal day^{-1} (Bernard *et al*., 2018). The 1800 kcal day^{-1} represents an average resting metabolic rate of 90 W. Allowing for some activity beyond sitting, 100 W was chosen.

To classify an exposure as above or at/below the exposure limit, the following steps were taken. First, the TWA-WBGT and TWA-M were computed. Then the appropriate occupational limit was calculated from the TWA-M using equations (1) and (2).

$$\text{Acclimatized}: \text{OEL} = 56.7 - 11.5 \log_{10} \text{TWA-M} \tag{1}$$

$$\text{Unacclimatized}: \text{OAL} = 59.9 - 14.1 \log_{10} \text{TWA-M} \tag{2}$$

Next the elevation above the limit (ΔLimit) was computed as

$$\text{Acclimatized}: \Delta\text{Limit} = \text{TWA-WBGT} - (\text{OEL} - \text{CAV}) \tag{3}$$

$$\text{Unacclimatized}: \Delta\text{Limit} = \text{TWA-WBGT} - (\text{OAL} - \text{CAV}) \tag{4}$$

Finally, the exposures were classified as above for $\Delta\text{Limit} > 0$, and below for $\Delta\text{Limit} \le 0$.

The outcome variable (uncompensable or compensable) was judged by the final peak core temperature. Ashley *et al.* (2008) reported data at the ULPZ noting a difference due to metabolic rate but not clothing. The mean core temperature was 37.8°C with a standard deviation of about 0.3°C for a mix of men and women. We extracted the core temperature data from Lind's study on individual variability near the ULPZ (Lind, 1970) and found a mean and

Table 3. Distributions of compensable and uncompensable observations based on acclimatization status, clothing, WBGT (same or different), M (same or different), and pairs of WBGT and M (same or different). Also reported are the mean and range of TWA-WBGT, and TWA-M by outcome.

	Compensable	Uncompensable
Number of observations	30	32
Acclimatization status		
Acclimatized	12	7
Unacclimatized	18	25
Clothing		
Semi-nude	16	10
Semi-clothed	2	1
Woven clothing	9	20
Modified work uniform	2	0
Vapor-barrier coveralls	1	1
WBGT		
Same	21	26
Different	9	6
M		
Same	2	2
Different	28	30
Pairs of WBGT and M		
Same WBGT and different M	21	26
Different WBGT and different M	7	4
Different WBGT and same M	2	2
ΔLimit		
Mean	−0.9	1.8
Standard deviation	3.3	2.9
TWA-WBGT (°C)		
Mean	27.3	29.1
Standard deviation	3.4	3.2
TWA-M (W)		
Mean	270	260
Standard deviation	97	99

standard deviation of 38.2 and 0.5°C. From these observations, a value of 38°C was selected to represent the mean response for sustainable heat stress exposures and 38.5°C for a high individual response. For outcome classification purposes, if the reported mean core temperature of the participants was greater than 38°C; if the core temperature of any individual (if reported) was greater than 38.5°C; or if 95th percentile was greater than 38.5°C, the outcome was classified as uncompensable. Otherwise, the outcome was classified as compensable.

If the core temperature did not appear to be steady at the end of the trial (increasing values of peak core temperature at the end of a work period) and the classification was compensable at the end of the trial, a projected core temperature into the following hour was used to override the decision. This was accomplished by extending a straight line from the last two peak temperatures out to 180 min. Eighteen observations were classified as uncompensable based on projections [2 for Gagnon (Gagnon and Kenny, 2011); 2 for Kaltsatou (Kaltsatou *et al.*, 2020); 2 for Maiariaux (Mairiaux *et al.*, 1986); 1 for Meade (Meade *et al.*, 2016); 3 for Seo (Seo *et al.*, 2019); 2 for Stapleton (Stapleton *et al.*, 2012); 2 for Wright Beatty (Wright Beatty *et al.*, 2015); and 4 for Wright (Wright *et al.*, 2014)].

Tabulation of data were performed using JMP Pro 16 Software (SAS, Cary NC). 2 × 2 tables were constructed to report the exposure classification (above or at/below ΔLimit) versus the outcome (compensable or uncompensable). A Pearson χ^2 was computed to determine the statistical significance of the distribution of values in the 2 × 2 tables. A χ^2 test for independence was used to compare two 2 × 2 tables to see if the distributions were similar. An ANOVA was used to see if there was a difference in ΔLimit for acclimatization state and outcome including the interaction.

Results

Supplementary Table S (available at *Annals of Work Exposures and Health* online) provides details about each observation for 16 reported studies for a total of 62 observations. All studies with different Ms used sitting as one of the exposure conditions with an assigned resting metabolic rate of 100 W.

Table 3 reports distributions of the observations classified as compensable or uncompensable. About half (30) of the observations were associated with compensable levels of heat stress and the other half (32) were classified as uncompensable. Most of the observations employed the same environment for work and rest (47 or 75%), had unacclimatized participants (43 or 70%), and used a work clothing ensemble (29 or 47%). Table 3 also reports the outcomes for all 62 observations based on the core temperatures at the end of the trial; that is, without a projection to 180 min. Fig. 1 illustrates the observations as a plot of ΔLimit versus last recorded peak core temperature during the trial (i.e. no projections). The uncompensable observations where the classification was based on a projection are those located below or at 38°C.

Table 2 provides 2 × 2 tables for selected observations, which compares the exposure classification

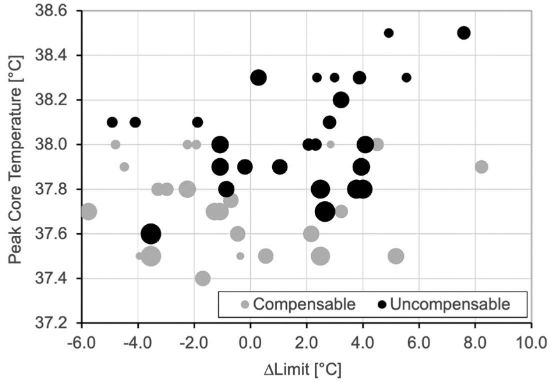

Figure 1. Observations plotted as ΔLimit by mean core temperature. The compensable observations are gray circles; the uncompensable observations are black circles; and the area of the circle is proportional to the number of participants in that observation. The core temperatures are the last recorded peak value in the trial period. The uncompensable observations (black circles) below 38°C are the ones that were reclassified based on a projection out to 3 h.

(above or below) to outcome state (compensable or uncompensable). The Pearson χ^2 was significant ($P < 0.05$) for the selected observations. Overall, the sensitivity and specificity were 0.72 and 0.73. Sensitivity for acclimatized observations was 1.0; and specificity was 0.67. In contrast, the sensitivity and specificity for unacclimatized were 0.64 and 0.78. Looking at observations associated with the same WBGT, sensitivity and specificity were 0.77 and 0.67, respectively. Using just the data contained within a trial (i.e. no extrapolation), the sensitivity and specificity were 0.79 and 0.58. Table 2 also reports χ^2 tests for three pairs of comparisons. There was a significant difference in the 2 × 2 distributions between all observations and those for all observations without projections to 3 h. Comparing the distributions of all to same WBGT and acclimatized to unacclimatized, there were no significant findings.

The two-way ANOVA looked for a change in ΔLimit due to the outcome (uncompensable versus compensable) and acclimatization state as the independent variables in addition to the interaction of outcome and acclimatization. Both outcome and acclimatization state were significant ($P < 0.05$) and the interaction was not. Rerunning the model without interaction found a difference in least square means due to outcome with 2.3°C for uncompensable and −0.7°C for compensable; and a difference in least square means for acclimatization state was −0.1°C for unacclimatized and 1.7°C for acclimatized. The overall mean for reference was 0.5°C.

Discussion

The goal of WBGT-based heat stress exposure limits was to limit exposures to the ULPZ (Lind, 1963a, 1970; Dukes-Dobos and Henschel, 1973; Garzón-Villalba *et al.*, 2017), or at least to assure that thermal equilibrium can be maintained (ACGIH®, 2022). Several steps were taken to classify the heterogeneous exposures as above the exposure limit or not. The first step was to determine TWA-WBGT and TWA-M based on the data for environment and M provided in each paper. The second step was to determine the difference between the exposure based on TWA-WBGT and TWA-M and the exposure limit appropriate to the acclimatization state (OEL or OAL) (ΔLimit). The last step was to classify the exposures (above or at/below the limit).

The outcome decision was based on reported core temperature (usually rectal temperature). An extrapolation past the exposure time to 3 h was considered if core temperature was continuing to increase and the outcome was compensable at that time. The observation was classified as uncompensable if the mean value of core temperature for the participants was >38°C, any individual's reported core temperature was >38.5°C, or the 95th percentile core temperature was >38.5°C. Using both the mean and the 95th percentile criteria provided two checks on whether the rectal temperature responses appeared to be consistent with the ULPZ (Ashley *et al.*, 2008) or demonstrated ability to sustain the exposure for 3 h (Lind, 1970). The classification criterion for the mean was for the population of participants in the study, and thus represented the group response. The 95th percentile would identify a less heat-tolerant individual. The outcomes (compensable or uncompensable) are illustrated in Fig. 1 on a plot of ΔLimit versus last recorded peak core temperature. It is not surprising to see the difficulty in separating the compensable from uncompensable data. The problem addressed below is the number of uncompensable observations less than ΔLimit = 0. These represent false-negative decisions.

Least square mean values of ΔLimit for compensable and uncompensable exposures were −0.7 and 2.3°C-WBGT, respectively. The higher average ΔLimit for uncompensable was expected, but there was considerable overlap associated with standard deviations of 3°C. We expected to see compensable levels of heat stress above the exposure limit (ΔLimit = 0) because of the protective nature of the exposure limits (Lind, 1970; Dukes-Dobos and Henschel, 1973; Garzón-Villalba *et al.*, 2017). Lind's study of 95 unacclimatized, semi-nude men divided among four WBGTs at 350 W (OAL adjusted for semi-nude is 25.6°C-WBGT) provided insight into how many participants could sustain a heat stress exposure for 3 h (Lind, 1970).

While 24 of 25 completed the trial at 27.4°C-WBGT (ΔLimit = 1.8°C-WBGT), Lind reported that 68% of participants (13 of 19) could complete the 3-h trial at 29.2°C-WBGT (ΔLimit = 3.6°C-WBGT) and 40% (10 of 25) at 30.8°C-WBGT (ΔLimit = 5.2°C-WBGT). At ΔLimit of 2°C and assuming the heat tolerance curve has similar shape for unacclimatized, we would expect that 95% of the exposures would be compensable (Garzón-Villalba *et al.*, 2017). Thus, exposures above the OEL or OAL were expected to result in false-positive outcomes; and that the probability of false-positive outcomes would decrease with increasing levels above the exposure limit.

For sensitivity and specificity, the expected sensitivity was near 1.0 with lesser specificities that were dependent on the design of the study (Garzón-Villalba *et al.*, 2017). Table 2 presents a much different picture. The overall sensitivity and specificity were weak at 0.72 and 0.73, respectively. The goal of protection of most workers [e.g. 95% (Dukes-Dobos and Henschel, 1973)] was elusive with the lower sensitivity; and low specificity meant it was also difficult to identify compensable exposures with any certainty. This remained true when the 18 cases where the decision was based on projected changes in core temperature were classified based on core temperature at the end of the trail (specificity of 0.79 and sensitivity of 0.58). There was a statistically significant change in the distribution from compensable to uncompensable due to the projection. In sum, the TWA appeared to underestimate the actual heat stress experienced by the observations leading to false-negative classifications and lower sensitivity.

TWA appeared to work for the observations from acclimatized participants. That is, the sensitivity was 1.0 and the specificity near what would be expected for widely different exposures (Garzón-Villalba *et al.*, 2017). This was not true for the unacclimatized observations where sensitivity was 0.64. The mean ΔLimit was not statistically significant at 1.32°C for acclimatized and 0.12°C for unacclimatized. There was not a statistically significant difference in the distribution of observations between the acclimatized and unacclimatized. The high sensitivity for acclimatized and lack of statistical significance may be due to a small sample size.

Most of the exposure scenarios were work and recovery in the same environment (75%), which basically evaluated only TWA-M. When observations based on the same environment were considered alone, there was no improvement in the ability to discriminate for uncompensable heat stress and the distribution of same WBGT was not statistically different from all the observations.

There were only 11 of 62 observations for different WBGTs and different Ms. The sensitivity was poor at 0.25 and the specificity was 0.86. Again, the TWAs did not provide the expected sensitivity.

The four observations for same M and different WBGTs were above the limit and split two and two for compensable status. There was no opportunity to explore the value of TWA-WBGT alone because there were no observations below the exposure limit to add contrast.

Criteria for the case definition were based on a linear extrapolation of core temperature to 3 h. This led to 18 more cases (12 true positives and 6 false negatives) with a decrease in sensitivity (0.79–0.72) and increase in specificity (0.58–0.73). A more appropriate extrapolation might have been a first order exponential increase based on all the preceding data. At best, the exponential extrapolation might have resulted in fewer than 18 additional cases with specificity still less than 0.79. The extrapolation to 3 h did not explain the unexpected results.

Not included among the data for evaluating TWAs were seven observations of continuous exposures (same WBGT and same M) for unacclimatized groups among the included papers (Mairiaux *et al.*, 1986; Graveling and Morris, 1995; Gagnon and Kenny, 2011; Meade *et al.*, 2016; Lamarche *et al.*, 2017; Kaltsatou *et al.*, 2020). The sensitivity was 0 (no true positives and four false negatives) and the specificity was 0.67 based on three compensable observations. Assuming the OAL is adequately protective, the case definition using core temperature thresholds may not be representative of compensable heat stress.

The WHO report on heat stress (WHO, 1969) recommended a maximum core temperature of 38°C; this goal was stated in the development of WBGT-based exposure limits (Dukes-Dobos and Henschel, 1973); and repeated in earlier versions of the TLV and REL. If 38°C limit for any individual was used for the case definition, there would be many more outcome classifications as uncompensable. While not tested, the result would be even lower sensitivity for the TWA observations.

The core temperature limits were selected as a surrogate to demonstrate thermal equilibrium. The WBGT-based limits were based on Lind's concept of the ULPZ, below which thermal equilibrium can be established. In practice, the exposure limits have the goal of thermal equilibrium or compensable heat stress, which can occur above the ULPZ. While the belief that exposures below the ULPZ can limit core temperature to 38°C (Dukes-Dobos and Henschel, 1973) or about 38.3°C (Ashley *et al.*, 2008) or 39°C (Lind, 1970) might be reasonable, the reverse is not necessarily true. A fixed core temperature cannot be used to demonstrate an exposure below the ULPZ or to demonstrate thermal equilibrium.

In addition to the problem of using core temperature limits to demonstrate thermal equilibrium, there were notable limitations to evaluating TWAs. The first limitation was the fact that there was no uniform reporting of population or individual heat stress exposure data. This left us to make our best judgment about the exposure and outcome data. Another was that by reporting population data, (i) we could not capture individual metabolic rates to better classify the heat stress level and (ii) we needed to make an informed guess about the response of the less tolerant individuals. Another limitation was the need to estimate a WBGT for trial conditions and a lack of specificity about the instrumentation used to report environmental data. When air velocity was not reported, 0.3 m s^{-1} was assumed, which allows for air exchanges to control environments with little sensible motion. For globe temperature, the air temperature was used as the surrogate when no radiant heat sources were reported (61 of 64 observations). Often natural wet bulb temperature was not reported and it was estimated using a nonlinear relationship among natural wet bulb temperature, psychrometric wet bulb temperature (or ambient water vapor pressure), globe temperature, and air speed. Resting metabolic rate was not reported for most observations and a value of 100 W was used. Thus, there were potential classification errors for both exposure and outcome.

The number of participants in each study ranged from 2 to 14 with an average across observations of 8. Each observation was treated with the same weight, which overstated the findings with few participants in comparison to the observations with higher numbers.

The search for papers to include was not exhaustive but there was no intentional bias in the search or selection of papers. While more papers would add to the database, they would not likely change the outcome that sensitivity was lower than would be expected due to the large number of false-negative outcomes and low sensitivity already reported.

A more rigorous test of TWAs with exposures above and below compensable levels of heat stress should include the four TWA pairs of same and different WBGT and M, where the exposure data and the outcomes are reported and analyzed on an individual basis. To demonstrate that TWAs are applicable to heterogeneous exposures, TWAs should be determined for the four pairs of conditions (i.e. different WBGTs/different Ms; different WBGTs/same M; and same WBGT/different Ms with same WBGT/same M as a control). Ideally, a study would use the same participants in a cross-over design to better control for individual differences where the TWA-WBGT and TWA-M are held constant for each of the four pairs. A study that changes the proportions of work and rest cycles is also necessary to better evaluate TWAs. In summary, it is necessary to explore all the exposure pairs with some variation in the proportion of work and rest (e.g. 75/25, 50/50, and 75/25; or 33/67 and 67/33) and conditions below, at, and above the exposure limit. Acclimatized participants help reduce the variability but an unacclimatized population should also be included due to the uncertainty found in this paper and a relevant population for heat stress management. Steady exposures below the ULPZ for an individual will establish the work-driven body core temperature (Saltin and Hermansen, 1966) for comparison with the TWA trials for that individual to see if the core temperatures are similar or different for TWAs below the exposure limits and how much elevation there may be during exposures where thermal equilibrium can be demonstrated.

The OELs were designed to allow most healthy, hydrated people to achieve a compensable level of heat stress based on constant exposures. The classified data presented in the 2 × 2 tables (Table 2) suggested that the TWAs did not reflect heat stress exposures consistent with the goals of the OELs. Using core temperature criteria that were higher than 38°C, the findings of the University of Ottawa investigators (Wright Beatty *et al.*, 2015; Meade *et al.*, 2016; Lamarche *et al.*, 2017; Seo *et al.*, 2019; Kaltsatou *et al.*, 2020) were confirmed. There was some suspicion that there were differences between acclimatized and unacclimatized observations. Before any of these findings are embedded in policy or practice, a more careful evaluation of TWAs is required.

Funding

John W. Flach was supported by the United States Coast Guard during his studies for the MSPH at USF. This paper was developed from his MSPH thesis. Thomas E. Bernard was the Director of the Sunshine ERC and received support from Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (T42-OH008438).

Acknowledgements

We would like to thank Amie M. Gimon, College of Nursing, University of South Florida, who contributed to this paper through her MPH special project. Ms. Gimon was supported by a Sunshine ERC training grant from Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health (T42-OH008438).

Conflict of interest

The authors declare no conflict of interest relating to the material presented in this article. Its contents, including any opinions and/or conclusions expressed, are solely those of the authors. Specifically, the opinions offered in this paper do not reflect opinions, positions, or policies of USCG or CDC/NIOSH. One of the authors (Bernard) has acted as an expert witness for both private companies and OSHA in litigation concerning heat stress exposures and may in future serve as an expert witness in court proceedings related to heat stress.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary data

Supplementary data are available at *Annals of Work Exposures and Health* online.

References

ACGIH®. (2022) *Heat stress and strain. Threshold limit values and biological exposure indices for chemical substances and physical agents*. Cincinnati, OH: ACGIH®.

Ashley CD, Luecke C, Schwartz SW *et al*. (2008) Heat strain at the critical WBGT and the roles of clothing, metabolic rate and gender. *Int J Ind Ergon*; **38**: 640–4.

Belding HS, Kamon E. (1973) Evaporative coefficients for prediction of safe limits in prolonged exposures to work under hot conditions. *Fed Proc*; **32**: 1598–601.

Bernard TE, Yantek DS, Thimons ED. (2018) Estimation of metabolic heat input for refuge alternative thermal testing and simulation. *Min Eng*; **70**: 50–4.

Dukes-Dobos FN, Henschel A. (1973) Development of permissible heat exposure limits for occupational work. *Am Soc Heat Refrig Air Cond Eng J*; **15**: 57–62.

Gagnon D, Kenny GP. (2011) Exercise-rest cycles do not alter local and whole body heat loss responses. *Am J Physiol Regul Integr Comp Physiol*; **300**: R958–68.

Garzón-Villalba XP, Wu Y, Ashley CD *et al*. (2017) Ability to discriminate between sustainable and unsustainable heat stress exposures—Part 1: WBGT exposure limits. *Ann Work Expo Health*; **61**: 611–20.

Graveling RA, Morris LA. (1995) Influence of intermittency and static components of work on heat stress. *Ergonomics*; **38**: 101–14.

ISO. (2017) *ISO7243: Ergonomics of the thermal environment—assessment of heat stress using the WBGT (wet bulb globe temperature) index*. Geneva: ISO.

Kaltsatou A, Flouris AD, Herry CL *et al*. (2020) Heart rate variability in older workers during work under the Threshold Limit Values for heat exposure. *Am J Ind Med*; **63**: 787–95.

Lamarche DT, Meade RD, D'Souza AW *et al*. (2017) The recommended Threshold Limit Values for heat exposure fail to maintain body core temperature within safe limits in older working adults. *J Occup Environ Hyg*; **14**: 703–11.

Lind AR. (1963a) A physiological criterion for setting thermal environmental limits for everyday work. *J Appl Physiol*; **18**: 51–6.

Lind AR. (1963b) Physiological effects of continuous or intermittent work in the heat. *J Appl Physiol*; **18**: 57–60.

Lind AR. (1970) Effect of individual variation on upper limit of prescriptive zone of climates. *J Appl Physiol*; **28**: 57–62.

Mairiaux P, Libert JP, Candas V *et al*. (1986) Prediction of strain for intermittent heat exposures. *Ergonomics*; **29**: 913–23.

Meade RD, Poirier MP, Flouris AD *et al*. (2016) Do the threshold limit values for work in hot conditions adequately protect workers? *Med Sci Sports Exerc*; **48**: 1187–96.

NIOSH. (2016) In Jacklitsch B, Williams WJ, Musolin K, Coca A, Kim J-H, Turner N, editors. *NIOSH criteria for a recommended standard: occupational exposure to heat and hot environments*. Cincinnati, OH: Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health (DHHS (NIOSH) Publication No. 2016-106).

Saltin B, Hermansen L. (1966) Esophageal, rectal, and muscle temperature during exercise. *J Appl Physiol*; **21**: 1757–62.

Seo Y, Powell J, Strauch A *et al*. (2019) Heat stress assessment during intermittent work under different environmental conditions and clothing combinations of effective wet bulb globe temperature (WBGT). *J Occup Environ Hyg*; **16**: 467–76.

Stapleton JM, Wright HE, Hardcastle SG *et al*. (2012) Body heat storage during intermittent work in hot-dry and warm-wet environments. *Appl Physiol Nutr Metab*; **37**: 840–9.

WHO. (1969) *Health factors involved in working under conditions of heat stress*. Geneva: World Health Organization Technical Report Series 412.

Wright HE, Larose J, McLellan TM *et al*. (2014) Moderate-intensity intermittent work in the heat results in similar low-level dehydration in young and older males. *J Occup Environ Hyg*; **11**: 144–53.

Wright Beatty HE, Hardcastle SG, Boulay P *et al*. (2015) Increased air velocity reduces thermal and cardiovascular strain in young and older males during humid exertional heat stress. *J Occup Environ Hyg*; **12**: 625–34.