Annals of Work Exposures and Health, 2017, Vol. 61, No. 6, 611–620
doi: 10.1093/annweh/wxx034
Advance Access publication 8 June 2017
Original Article

Original Article

Ability to Discriminate Between Sustainable and Unsustainable Heat Stress Exposures—Part 1: WBGT Exposure Limits

Ximena P. Garzón-Villalba, Yougui Wu, Candi D. Ashley and Thomas E. Bernard*

College of Public Health, University of South Florida, 13201 Bruce B Downs Boulevard, Tampa, FL 33612, USA

*Author to whom correspondence should be addressed. Tel: 813-974-6629; e-mail: tbernard@health.usf.edu

Submitted 22 September 2016; revised 28 March, 2017; editiorial decision 28 March 2017; revised version accepted 11 May 2017.

Abstract

Objectives: Heat stress exposure limits based on wet-bulb globe temperature (WBGT) were designed to limit exposures to those that could be sustained for an 8-h day using limited data from Lind in the 1960s. In general, Sustainable exposures are heat stress levels at which thermal equilibrium can be achieved, and Unsustainable exposures occur when there is a steady increase in core temperature. This paper addresses the ability of the ACGIH® Threshold Limit Value (TLV®) to differentiate between Sustainable and Unsustainable heat exposures, to propose alternative occupational exposure limits, and ask whether an adjustment for body surface area improves the exposure decision.
Methods: Two progressive heat stress studies provided data on 176 trials with 352 pairs of Sustainable and Unsustainable exposures over a range of relative humidities and metabolic rates using 29 participants wearing woven cotton clothing. To assess the discrimination ability of the TLV, the exposure metric was the difference between the observed WBGT and the TLV adjusted for metabolic rate. Conditional logistic regression models and receiver operating characteristic curves (ROC) along with ROC's area under the curve (AUC) were used. Four alternative models for an occupational exposure limit were also developed and compared to the TLV.
Results: For the TLV, the odds ratio (OR) for Unsustainable was 2.5 per 1°C-WBGT [confidence interval (CI) 2.12–2.88]. The AUC for the TLV was 0.85 (CI 0.81–0.89). For the alternative models, the ORs were also about 2.5/°C-WBGT, with AUCs between 0.84 and 0.88, which were significantly different from the TLV's AUC but have little practical difference.
Conclusions: This study (1) confirmed that the TLV is appropriate for heat stress screening; (2) demonstrated the TLV's discrimination accuracy with an ROC AUC of 0.85; and (3) established the OR of 2.5/°C-WBGT for unsustainable exposures. The TLV has high sensitivity, but its specificity is very low, which is protective. There were no important improvements with alternative exposure limits, and there was weak evidence to support metabolic rate normalized to body surface area. In sum, the TLV

is protective with an appropriate margin of safety for relatively constant occupational exposures to heat stress.

Keywords: heat stress; occupational exposure limit; ROC; sustainable; threshold limit values; WBGT; wet-bulb globe temperature; unsustainable

Introduction

The assessment of heat stress considers the effects of environmental conditions, metabolic rate, and clothing. The wet-bulb globe temperature (WBGT) index considers the combined environmental effects of air temperature, radiant heat load, air movement, and humidity (NIOSH, 2016). Despite its limitations (d'Ambrosio Alfano *et al.*, 2014; Budd, 2008), WBGT is a widely used heat stress index because it is convenient and easy to interpret (Budd, 2008; NIOSH, 2016). In 1974, the ACGIH® adopted WBGT as the index for expressing environmental contributions to heat stress Threshold Limit Values (TLVs®) (ACGIH, 2017). Because the ACGIH TLV was the first widely used WBGT-based occupational exposure limit for heat stress, it is the stand-in for the same occupational exposure limits as the NIOSH Recommended Exposure Limit (REL) (NIOSH, 2016) and the draft ISO 7243 (ISO, 1989).

Based on ordinary work clothes, the TLV for heat stress has its origin from the upper limit of the prescriptive zone (ULPZ) proposed by Lind (1963a) using three different metabolic rates. The combinations of effective temperature and metabolic rate where core temperature maintained consistent values independent of the environment were called the prescriptive zone, preferred exposure range was marked by the ULPZ. For exposures above the ULPZ, there was a trend of rising core temperatures with increasing levels of environmental heat. In three additional studies, Lind (1963b, 1970); Lind *et al.* (1970) confirmed that subjects as a group could not maintain thermal equilibrium above the ULPZ. In our paper, an inability to maintain thermal equilibrium was called Unsustainable. Conversely, thermal equilibrium could be sustained for a long period at and below the ULPZ, and we called these exposures Sustainable. Dukes-Dobos and Henschel (1973) used Lind's data as the primary rationale to support the TLV and the 1986 revision of the NIOSH REL (NIOSH 1986).

The TLV was later examined by Kuhlemeier et al. (1977) in acclimatized workers (working in hot industries in the summer) and unacclimatized workers (from cold and neutral industries such as meat packing, janitorial service, and hospital orderlies, working in the winter). The investigators exposed workers randomly to environments either below or above the TLV at three levels of metabolic rate. They confirmed that as a group, exposures below the TLV yielded consistent core temperatures and heart rates (HRs), and above the TLV the data showed an upward trend with increasing WBGT. For their unacclimatized population, they used a lower threshold for the ULPZ.

In one study, Lind (1970) reported outcomes for individuals for four environments at one metabolic rate (350 W) as able to complete 3 h (Sustainable) and failing to complete 3 h due to heat stress (Unsustainable). With respect to the TLV of 27.4°C-WBGT at 350 W, two environmental conditions were clearly above (30.7 and 33.3°C-WBGT); one was well below (22.8°C-WBGT); and another one was closer to, but higher than, the TLV (28.9°C-WBGT). Although the subjects were seminude and unacclimatized, these two effects were considered by Dukes-Dobos and Henschel (1973) as equivalent to being acclimatized and wearing work clothes. Using Lind's report of individual data, we reanalyzed the data noting that there were 22 Unsustainable cases (i.e. individuals not able to complete the 3-h exposure due to heat stress), all of them above the TLV, and 73 Sustainable cases with 47 of them above the TLV and 26 below. It is worth noting the sensitivity, which is the ability to detect true Unsustainable exposures, and specificity, which is the ability to detect true Sustainable exposures. The sensitivity was 1.00, which means all the Unsustainable exposures were above the TLV and a specificity of 0.35, which means that there were also many Sustainable exposures above the TLV. An ideal decision tool would have both high sensitivity and specificity. A screening tool should have at least high sensitivity, always taking into consideration the possibility of having many false positive decisions.

Another way to characterize an exposure assessment tool is to consider the odds that an exposure is Unsustainable. The odds are defined as the probability that the exposure to heat stress is Unsustainable divided by the probability that it is Sustainable. The odds range from zero to positive infinity. The odds are usually related through a log-linear model to an exposure metric. That is, ln(odds) = $a + bx$, which is the functional equivalent of performing a logistic regression. The odds ratio (OR) is the ratio of the odds for a given exposure divided by the odds at a reference point such as the occupational exposure limit. The OR is related to the slope of the log-linear relationship such that OR = e^b. The probability that an exposure is Unsustainable based on an exposure metric is nonlinear. In our continuing reanalysis of the data, the individual data were assigned as the difference between the WBGT of the exposure minus the value for the TLV (27.4°C-WBGT) at 350 W. Based on a logistic regression model using this difference to assess its association with the outcome of Unsustainable, we found a significant OR = 2.00 [confidence interval (CI) 1.44–2.81]/°C-WBGT. That is, the odds of an Unsustainable exposure increased 2-fold for each 1°C of difference between the TLV and the observed WBGT.

Kuhlemeier *et al.* (1977) provided individual data graphically (not in tables). Their experimental design purposely had exposures above and below the TLV. In a similar reanalysis of the data, we classified all individual exposures below the TLV as Sustainable, and these exposures were usually well below the TLV. Above the TLV, the individual's classification was Sustainable if the reported rectal temperature (T_{re}) was at or below the average T_{re} for the observations below the TLV. If the individual was above the average T_{re}, the exposure was classified as Unsustainable. There were 67 Unsustainable cases above the TLV and 88 Sustainable conditions, 71 of them under the TLV. The resulting sensitivity was 1.00 and specificity was 0.81. The high sensitivity and specificity were expected because there were no exposure conditions near the TLV. The OR from the logistic regression model was 2.10 (1.56–2.83)/°C-WBGT, which was similar to our analysis of Lind's data in the previous paragraph. In a similar fashion for the unacclimatized trials against the ACGIH action limit (AL) (ACGIH, 2017), we examined the sensitivity, specificity, and OR. There were 86 Unsustainable conditions above the AL, and none below it, for a sensitivity of 1.00. There were 62 Sustainable below the AL and four above it, resulting in a specificity of 0.94. The OR from the logistic regression model was 2.45/°C-WBGT (CI 1.63–3.71).

While the sensitivity and specificity described above from the Lind (1970) and Kuhlemeier *et al.* (1977) studies appeared strong, they were based on categories of metabolic rate and not individual values; and for the Kuhlemeier data, there was a potential for classification bias. Our paper used a progressive heat stress protocol data set that was designed to find the individual critical conditions where thermal equilibrium was no longer supported. The critical condition was approximately equivalent to an individual's ULPZ. Thus, individually known metabolic rates and WBGTs were available to assess the ability of the current and candidate WBGT-based occupational exposure limits to discriminate between Sustainable and Unsustainable heat exposures.

The dissipation rate of metabolic heat to the ambient environment depends on body surface area (BSA) (Havenith, 2001). One way to account for the BSA effects is to express the metabolic heat generation normalized to BSA in the exposure limits (i.e. W m^{-2}). This can be examined through alternative exposure limits (AELs).

Methods

The data used for this paper were taken from two previous studies at University of South Florida (USF) (Bernard *et al.*, 2005, 2008), which were approved by the USF institutional review board. The progressive heat stress protocol used in these studies began with a comfortable environment that was easily sustainable. After thermal equilibrium was established, the temperature and humidity were slowly increased in 5-min intervals and the steps were designed to establish a quasi–steady-state physiological response for each step increase in heat load. T_{re}, HR, skin temperature (T_{sk}), and ambient conditions were monitored continuously and recorded every 5 min. Metabolic rate was estimated from the assessment of oxygen consumption via expired gases sampled every 30 minutes in a trial. The transition from a steady value for T_{re} to values that were steadily increasing were marked as the critical condition (see Fig. 1). A compensable point where the individual clearly could maintain thermal equilibrium was selected as 15 minutes before the critical condition; and an uncompensable point where the individual clearly could not maintain thermal equilibrium was selected as 15 minutes after the critical condition. The 15-minute period before and after the critical condition was selected to be near the critical condition to minimize the difference in WBGT but with high confidence that the classifications of compensable and uncompensable were correct.

Bernard *et al.* (2005) used the progressive heat stress protocol to find the effect of relative humidity on critical WBGT ($WBGT_{crit}$) for five clothing ensembles that included work clothes (140 g/m^2 cotton shirt and 270 g/m^2 cotton pants) and cotton coveralls (310 g/m^2). The target metabolic rate was 160 W m^{-2} to approximate moderate work. The subjects were exposed to three environments: warm, humid at 70% RH; hot, dry at 20% RH; and a moderate at 50% RH. In the other study, Bernard *et al.* (2008) were interested in the effects of varying metabolic rates at a relative humidity of 50%. The three target metabolic rates were 115, 175, and 250 W m^{-2} to approximate light, moderate, and heavy work. There were 176 trials for woven cotton clothing (work clothes and coveralls) over the two studies. The characteristics of the 29 participants who took part in these trials are summarized in Table 1. All participants were acclimatized by 2-h exposures over five successive days to dry heat (50°C and 20% relative humidity) at 160 W m^{-2} while wearing shorts and tee shirt.

To explore if the metabolic rates should be adjusted for populations with systematically lower BSA, a subset of the data was created for those with a BSA below 1.65 m^2. For this group, the metabolic rate was adjusted by multiplying the observed metabolic rate by a nominal BSA of 1.8 m^2 and divided by the individual's BSA following the discussion draft of ISO 7243 (ISO, 1989). This effectively increased the metabolic rate for those participants with low BSA and thus reduced the effective TLV for comparison purposes.

Statistical analyses

For each trial illustrated in Fig. 1, the compensable point was classified as Sustainable and the uncompensable point was classified as Unsustainable. The data at the critical condition were classified as Sustainable if T_{re} was less than 38°C or if the change in T_{re} was less than or equal to 0.1°C over the preceding 20 min. The critical condition observation was classified as Unsustainable if T_{re} was greater than or equal to 38°C and T_{re} increased by more than 0.1°C over the preceding 20 min. This method of classification resulted in two pairs of Sustainable–Unsustainable observations for each trial. Due to the fact that each individual served as their own control, it was crossover study design and the observations were dependent. Because there were 176 trials included in this study, there were 356 pairs of Sustainable/Unsustainable observations. Each subject completed three different trials; that is, either three levels of humidity or three levels of metabolic rate. The three trials represent repeated measures.

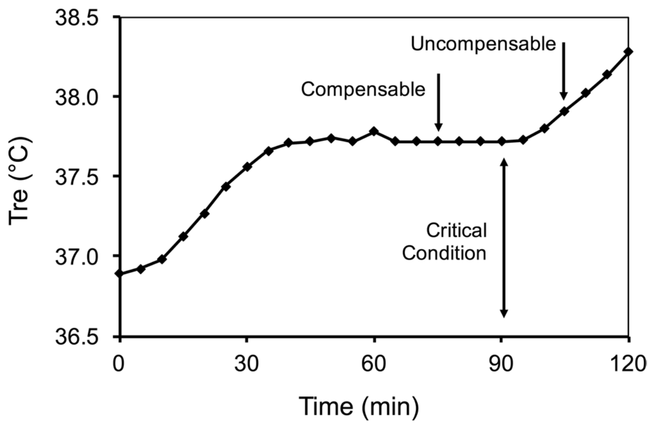

Figure 1. The time course of T_{re} for an example trial with arrows to indicate the critical condition, the compensable condition established 15 min before the critical condition and uncompensable 15 min after it.

Table 1. Physical characteristics (mean ± standard deviation) of participants.

	N	Age (years)	Height (cm)	Weight (kg)	Body surface area (m^2)
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

To assess the TLV, the exposure metric was the difference between the observed WBGT and the TLV value of WBGT adjusted for metabolic rate (M). That is, the TLV = 56.7–11.5 × $\log_{10}$ (M) (NIOSH, 2016); and the metric was ΔTLV = $\text{WBGT}_{\text{observed}}$ – TLV.

To characterize the data, descriptive statistics and bivariate distributions of the outcome by exposure were obtained using the SAS functions Proc Univariate and Proc Freq. The association between the exposure metric and the outcome was assessed with SAS Proc Logistic (SAS Institute Inc., 2013). Due to the dependent observations within the trials and the repeated observations on the participants, a conditional logistic regression model fitted with the binary outcome and a continuous predictor was used.

The ability to accurately discriminate between Sustainable and Unsustainable was assessed using a receiver operating characteristic curve (ROC curve). The ROC curve graphically displays the predictive accuracy of the logistic regression model, and the area under the curve (AUC) is a measure of the overall ability to predict Unsustainable exposures (Gallop, 2001; Gönen, 2006). A SAS ODS statement along with a request to produce the ROC plotted the fitted logistic regression models (SAS Institute Inc., 2013). Working from the premise that a WBGT-based occupational exposure limit functions as a screening method, a cut point of 0.95 for sensitivity was chosen. The AUCs, their correspondent CI, and the *P*-values from the comparison between the alternative models' AUC against the TLV's AUC were obtained using R (R Development Core Team, 2015) and following the method described by Wu and Wang (2011).

Alternative WBGT-based limits were also addressed using conditional logistic regression models using the same binary outcome (Sustainable versus Unsustainable). The alternatives used both WBGT and metabolic rate as the independent variables. First, WBGT was verified as the main predictor ($P < 0.001$), then metabolic rate was added to assess confounding and effect modification. The metabolic rate variables used were *M* in watts (W); $\log_{10}$(M); *M* divided by BSA (MSA) (W m^{-2}); and $\log_{10}$(MSA). Logistic regression using Proc log (SAS Institute Inc., 2013) was used to find the alternative models' estimates to compare their equivalent threshold limit lines expressed as WBGT = $f(M)$. A three-step process was followed as described in the Appendix.

Results

Descriptive statistics

Table 2 provides the mean, standard deviation, and 5%–95% quantiles for the metabolic rate (*M*), T_{re}, and HR for Sustainable and Unsustainable observations.

The frequency distribution of Sustainable heat exposure as a function of ΔTLV began with conditions under the TLV. It was notable that most of Sustainable conditions were above the TLV. Of the 255 Unsustainable observations, all of them were above the TLV. There were 259 Sustainable observations above the TLV and 14 under it. Therefore, sensitivity = 1.00 and specificity = 0.05.

For the population with a BSA under 1.65 m^2 and adjusting the metabolic rate by 1.8 m^2/BSA, there were 49 Unsustainable cases above the TLV, there were no Unsustainable cases under it. There were 39 Sustainable observations above and 8 Sustainable under the TLV, with a sensitivity = 1.00 and a specificity = 0.17.

Multivariate analysis and ROC curves

For the difference between the observed WBGT and the TLV (ΔTLV), the conditional logistic regression model showed that the odds of being Unsustainable increased 2.5 times per 1°C-WBGT of difference above the TLV (OR = 2.47, CI 2.12–2.88). The ROC curve had an AUC of 0.85 (CI 0.81–0.89), see Fig. 2.

There were four AEL developed from the data based on the treatment of metabolic rate. Table 3 summarizes the TLV and the four AELs. The dose–response curve for ΔTLV using the critical data is shown in Fig. 3. The TLV has a predicted probability of Unsustainable = 0.01. Three limit lines based on M (not normalized to BSA) are illustrated in Fig. 4.

The ORs for the TLV and the four alternative exposure metrics were obtained from the conditional logistic regression model and presented in Table 4. The AUCs are provided in the next column. The *P*-values from

Table 2. Summary statistics for metabolic rate, rectal temperature (T_{re}), and heart rate (HR) by outcome classification.

		Metabolic rate (W)			Rectal temperature (°C)			Heart rate (bpm)		
Outcome	*N*	Mean	SD	Quantiles 5%–95%	Mean	SD	Quantiles 5%–95%	Mean	SD	Quantiles 5%–95%
Sustainable	273	326	104	166–506	37.6	0.3	37.1–38.1	108	17	82–136
Unsustainable	255				37.9	0.3	37.4–38.5	130	20	100–164

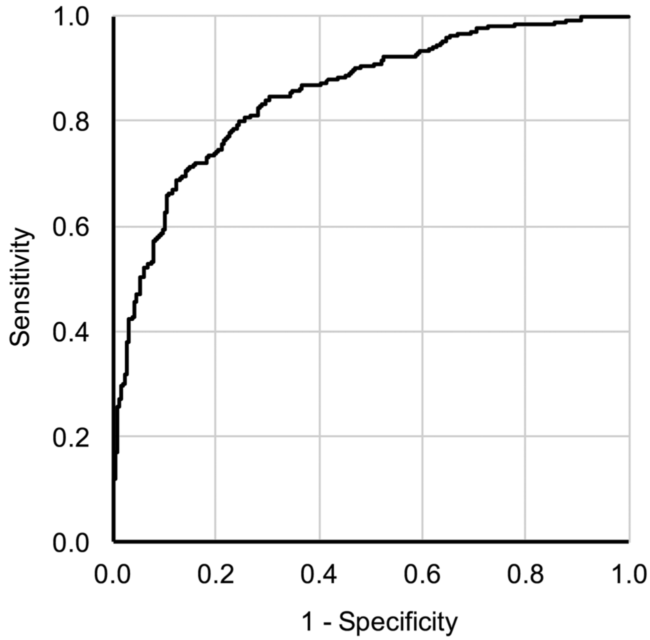

Figure 2. ROC curve for ΔTLV (AUC = 0.85).

Table 3. TLV and the four alternative exposure limits (AELs).

OEL	Exposure limit equation
TLV	TLV (WBGT) = 56.7 – 11.5 $\log_{10} M$
AEL: M	AEL_M (WBGT) = 34.34 – 0.0201 M
AEL: $\log_{10} M$	$AEL_{\log10M}$ (WBGT) = 64.67 – 14.9 $\log_{10} M$
AEL: MSA	AEL_{MSA} (WBGT) = 35.0 – 0.0367 MSA
AEL: $\log_{10}$MSA	$AEL_{\log10MSA}$ (WBGT) = 61.56 – 14.9 $\log_{10}$MSA

the comparison between each of the alternative model AUCs against TLV AUC are also reported. For all five exposure metrics, the odds to become Unsustainable were about 2.5/°C-WBGT, with AUCs between 0.84 and 0.88. A sensitivity of 0.95 was considered as an optimal operating point (OOP) (Gallop, 2001), believing that a higher probability for the OEL's prediction of true Unsustainable was preferred over the probability to accurately determine false positives. The specificities ranged from 0.36 to 0.50.

Discussion

Table 2 shows that the mean metabolic rate found among the participants of the present study was 325 W, with data divided among three group averages of 115, 175, and 250 W m^{-2} and a range from 170 to 500 W (Bernard *et al.*, 2008). In the Kuhlemeier *et al.* (1977) study, mean metabolic rate was 394 W over three groups (280, 380, and 540 W). While Lind (1970) reported only one group metabolic rate at 350 W, the present and Kuhlemeier studies reflected a range of metabolic rates that support validity over a wide range of heat stress exposures.

Using the critical condition as the approximation of the threshold was reasonable on two counts. First, it was analogous to Lind's ULPZ in that it was the upper limit for Sustainable exposures. Second, the critical condition represented a borderline condition for the population where about half the trials were classified as Sustainable and the other half Unsustainable. The other two data points in a trial were selected to be 15 minutes before and after the critical condition to assure an accurate classification. For our study, the compensable condition (Sustainable) was 2.9 ± 1.2°C-WBGT less that the critical condition and the uncompensable condition (Unsustainable) was 2.5 ± 1.1°C-WBGT above the critical condition. This was a relatively tight range of data compared to Lind's trials described in the Introduction section at 4.6°C below and 1.5°C, 3.3°C, and 5.9°C-WBGT above the TLV.

The TLV, based on the rationale offered by Dukes-Dobos and Henschel (1973), was assumed to protect more than 95% of those exposed based on Lind (1970) finding of 24 of 25 participants at the ULPZ demonstrated sustainable exposures. While not statistically rigorous, it has been the accepted premise of protection.

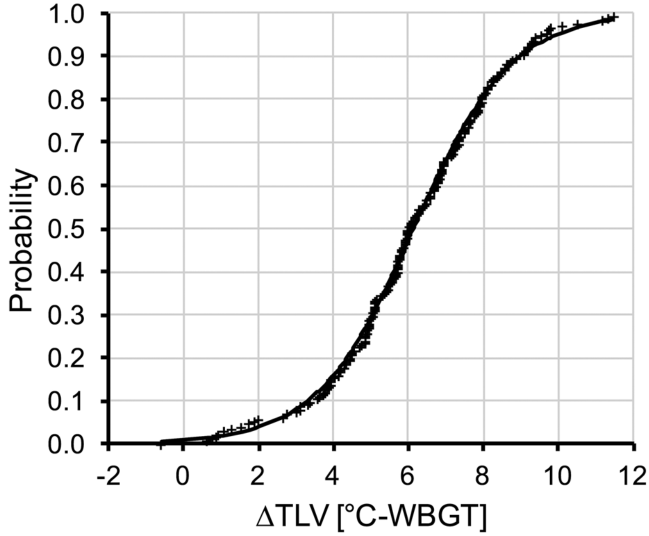

Figure 3. Relationship between the probability of being Unsustainable and ΔTLV (°C – WBGT) (= WBGT – TLV).

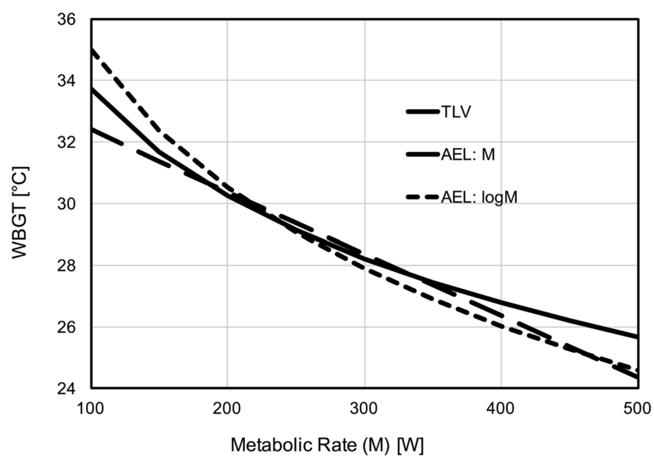

Figure 4. Comparison of the TLV to two alternative models using M.

Table 4. ORs from the conditional logistic models; AUCs, and *P*-values from the comparison of the TLV to the alternative models; and their specificity at an OOP sensitivity of 0.95.

Model	OR (CI)	AUC (CI)	*P*-value referenced to TLV	Specificity at sensitivity = 0.95
AEL_M	2.55 (2.17–2.99)	0.85 (0.81–0.8)	0.139	0.41
AEL_{log10M}	2.53 (2.16–2.96)	0.84 (0.79–0.88)	0.004	0.40
AEL_{MSA}	2.57 (2.19–3.01)	0.88 (0.85–0.91)	0.02	0.50
$AEL_{log10MSA}$	2.53 (2.16–2.96)	0.88 (0.85–0.91)	0.02	0.49

The present study found that the model probability of an Unsustainable exposure was closer to 1% rather than 4%. It also assessed the possibility of changing the cut point to improve the TLV's specificity with minimal tradeoff of sensitivity. Gallop (2001) called this threshold the OOP. Being conscious that any occupational exposure limit must be designed to protect the majority of the exposed population, an OOP of 0.95 for sensitivity was chosen to increase the accuracy with a small loss of sensitivity. Reducing the sensitivity to 0.95, the specificity increased from 0.05 to 0.36 (see Table 4) and the value for ΔTLV was 2.0. Based on sensitivity alone, this paper provided an argument in favor or raising the TLV by 2°C-WBGT.

Before recommending an increase in the TLV, several other factors should be considered. First, the ability to know the ambient WBGT is difficult in practice due to normal fluctuations and instrument accuracy. Second, the ability to know the metabolic rate for an individual is not easy and most heat stress assessments are based on population data; that is, what is the metabolic rate for the workers performing a given job. This range can be 20% of the mean, which would be about 60 W for a moderate metabolic rate of 300 W. The equivalent precision in terms of WBGT around 300 W would be about 1°C-WBGT. Thus a 2°C-WBGT is a reasonable margin of protection. Further, the probability of Unsustainable increased to 9% with another 1°C-WBGT step after a 2°C-WBGT increase. Thus, the increment in the risk associated with the higher threshold plus the uncertainty of the assessment is an important practical constraint related to increasing the cut point to improve specificity, despite the small loss of sensitivity. In this context, a similar increase in risk for exertional heat illness was reported by Garzon-Villalba *et al.* (2016) and Wyndham and Heyns (1973) for heat stroke. In addition to the narrow margin of safety related to exposure assessment, it is also clear that exertional heat illness may occur at exposures less than the TLV [see discussion in (Garzon-Villalba *et al.*, 2016)].

Ours is the first time that a paper examined heat stress exposure metrics using logistic regressions to characterize the ORs. The advantage of the AELs was they were data-driven rather than based on a pre-existing exposure limit. Among all five exposure limits, the odds to become Unsustainable were about 2.5/°C-WBGT, all with significant and small CIs (see Table 4). This was greater than the estimated odds from Lind (1970) data and higher than that found for Kuhlemeier *et al.* (1977) described in the Introduction section. Lind's data were based on one group metabolic rate (a somewhat narrow range of individual rates) and four environments, while Kuhlemeier were over a three groups of metabolic rates. The current data were over a greater range of environments (relative humidities at 20%, 50%, and 70%) and individual metabolic rates. While not estimated previously, the OR of 2.5/°C-WBGT for the WBGT-based exposure limits was well supported.

The assessment of predictive accuracy is a critical aspect of evaluating and comparing models that produce predictions, and the ROC curves are commonly used for such purposes when probabilities change with a decision threshold (Gönen, 2006). The AUC of the ROC was used in the current study as a summary statistic representing the overall performance. For reference, Shin and Coulter (2009) and Gallop (2001) pointed out that a metric with no predictive value would have an AUC of 0.5, while a metric with perfect ability to predict outcome would have an AUC of 1.00. The TLV's AUC was 0.85; and the AUCs of the four alternative limits were very close, with values between 0.84 and 0.88. While most of the alternative limits were statistically different than the TLV, all of the metrics had similar capabilities to discriminate between Sustainable and Unsustainable.

As seen in Fig. 4, the differences among the TLV and the two alternatives based on metabolic rate (not adjusted for BSA) were similar, especially between 200 and 300 W. The alternative models had lower threshold WBGTs at higher metabolic rates. There was also a spread at the lower metabolic rates where the linear model had lower thresholds and the nonlinear model had higher thresholds compared to the TLV. The tentative OOP chosen for the alternative occupational exposure limits followed the same criteria used for the TLV; that is, the cut point was set at a sensitivity of 0.95, with specificities from 0.36 to 0.50. All of the proposed models kept a high sensitivity, decreasing the proportion of false positives above the occupational exposure limit from 18 to approximately 8%. This performance is still relatively poor and not enough to justify a change in the threshold levels.

The widely accepted premise that a high BSA is beneficial in the heat (Havenith, 2001) was examined in the current study. Compared to the TLV, the AUCs for the two BSA alternatives were significantly different, with a slight improvement from 0.85 to 0.88. These AUC's accuracies, however, were not practically different. That is, the inclusion of BSA in an exposure assessment does not have strong support.

A strength of the study was the crossover design. It allowed for better estimation of the ORs and the relative contribution of metabolic rate through a conditional logistic regression. That is, the conditional logistic regression allowed us to establish the shape of the AELs. To establish an intercept, a logistic regression on just the critical data was required. Of the 176 observations at the critical point, the number of Sustainable and Unsustainable observations was about the same; thus, the estimation of the intercept value was probably not affected by the more limited data set.

One possible limitation of the present study was that the participants were relatively young with a mean age of about 30 years. Lind *et al*. (1970) reported on the average physiological responses of two age groups to exposure below and near the ULPZ and above the ULPZ, finding stable physiological responses and no differences due to age at and below the ULPZ. That would suggest that age was not a factor in establishing a Sustainable exposure.

Another limitation was that the data obtained in both USF studies were collected on laboratory trials, under controlled conditions with acclimatized participants and assuming constant, sustained exposures, which were not similar than those present in real work settings, and therefore generalization could be affected. Any changes to the current TLV (and similar occupational exposure limits) should wait for epidemiological data from actual workplaces with heat stress.

Conclusions

The original TLV was based on few subjects and limited qualitative approach. This study provided data-driven exposure limits. Compared with the current TLV, the four alternative OELs also had very high sensitivity and slightly improved specificities. The three most important findings from this study were (1) confirming that the TLV is appropriate for heat stress screening with a probability of being Unsustainable of 0.01; (2) establishing the accuracy to discriminate through an ROC curve (AUC = 0.85); and (3) establishing the OR for Unsustainable exposures at 2.5/°C-WBGT. For relatively constant occupational exposures to heat stress, the TLV and NIOSH REL are protective with an appropriate margin of safety. Of course, the current TLV has a high sensitivity, meaning that the exposed working population is protected. Because TLV's specificity is very low, it has a high percentage of false positives for the Unsustainable condition.

Any decision to decrease sensitivity in favor of improving the specificity of the TLV must be taken with caution because of the increasing probability to become Unsustainable with small increments of environmental temperatures. Further, there is some evidence of exertional heat illness below the TLV. Changes should be supported by epidemiological studies.

Considering BSA for AELs provided some marginal improvement in the ability to distinguish Unsustainable and Sustainable exposures.

Funding

Funding for this project was provided by the Republic of Ecuador and CDC/NIOSH (R01-OH03983 and T42-OH008438).

Acknowledgments

The data used in this study were collected under CDC/NIOSH R01-OH03983. Dr. Garzón was supported by the Republic of Ecuador through the Ecuadorian Institute of Human Talent Development. The authors recognize and thank the many laboratory assistants and trial participants who made this study possible. We would like to especially recognize Victor Caravello along with Mathew D. Dooris, Brian E. Grace, Patrick L. Rodriguez, and Brianna Clendenin Patton. Caravello was supported by the US Air Force, Dooris by the US Coast Guard, and Grace, Rodriguez and Patton were supported by CDC/NIOSH T42-OH008438.

Declaration

The authors declare no conflict of interest relating to the material presented in this article. 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. The contents, including any opinions and/or conclusions expressed, are solely those of the authors.

Appendix: Development of alternative exposure models

Alternative wet-bulb globe temperature (WBGT)-based limits were developed using both conditional logistic regression and logistic regression models. The three-step process was:

1. The weights associated with WBGT and each of the four functions of metabolic rate were determined from the conditional logistic equation based on the full data set of three conditions (compensable, transition, and uncompensable) in 176 progressive heat stress trials. In general, β_1 WBGT and $\beta_2\ f(M)$ represent the relative weightings. In effect, this step provided the weighting between WBGT and $f(M)$; and sets the shape of the curve when restated as WBGT and $(\beta_2/\beta_1)\ f(M)$.
2. The alternative exposure limit (AEL) model was framed as AEL[WBGT] = $\alpha + (\beta_2/\beta_1)\ f(M)$. A seed value of α was found by minimizing the sum of squares of the difference of AEL from the TLV. Then AEL = $\alpha + (\beta_2/\beta_1)\ f(M)$. From this, ΔAEL = $\text{WBGT}_{\text{observed}}$ − AEL = $\text{WBGT}_{\text{observed}} - [\alpha + (\beta_2/\beta_1)\ f(M)]$.
3. To obtain equations at a probability of 0.01, the data set was limited to 176 critical conditions which were used to fit a logit-p for ΔAEL. When the data were rank ordered by increasing ΔAEL, this approximated a dose-response curve. Logistic regression was used based on the progressive count of cases divided by

176 as the dependent variable. The result was ln [p/(1 − p)] = α′ + β’ ΔAEL. α from Step 2 was adjusted until α′ was −4.77, which gave a probability of Unsustainable equal to 0.01 at ΔAEL = 0, which was the value of α′ at the TLV (see Table A1).

Table A1. TLV and the four alternative models [AEL expressed as $\alpha - (\beta_2/\beta_1)\, f(\text{metabolic rate})$] at P=0.051 and the respective exposure metric based on observed WBGT and M and logit-p models.

	Alternative exposure limit development
TLV	$TLV_{@p=0.01}$ (WBGT) = 56.7 − 11.5 $\log_{10}M$ ΔTLV = WBGT − (56.7 − 11.5 $\log_{10}M$) $\ln[p/(1-p)] = -4.77 + 0.78\ \Delta TLV$
AEL: M	$AEL_{M@p=0.01}$ (WBGT) = 34.34 − 0.0201 M ΔAEL_M = WBGT − (34.34 − 0.0201 M) $\ln[p/(1-p)] = -4.77 + 0.74\ \Delta AEL_M$
AEL: $\log_{10}M$	$AEL_{\log10M@p=0.01}$ (WBGT) = 64.67 − 14.9 $\log_{10}M$ $\Delta AEL_{\log10M}$ = WBGT − (64.67 − 14.9 $\log_{10}M$) $\ln[p/(1-p)] = -4.77 + 0.71\ \Delta AEL_{\log10M}$
AEL: MSA	$AEL_{MSA@p=0.01}$ (WBGT) = 35.0 − 0.0367 MSA ΔAEL_{MSA} = WBGT − (35.0 − 0.0367 MSA) $\ln[p/(1-p)] = -4.77 + 0.89\ \Delta AEL_{MSA}$
AEL: $\log_{10}$MSA	$AEL_{\log10MSA@p=0.01}$ (WBGT) = 61.56 − 14.9 $\log_{10}$MSA $\Delta AEL_{\log10MSA}$ = WBGT − (61.56 − 14.9 $\log_{10}$MSA) $\ln[p/(1-p)] = -4.77 + 0.84\ \Delta AEL_{\log10MSA}$

References

ACGIH. (2017) *Heat stress, TLVs and BEIs: Threshold limit values for chemical substances and physical agents & biological exposure indices*. Cincinnati: ACGIH.

d’Ambrosio Alfano FR, Malchaire J, Palella BI *et al.* (2014) WBGT index revisited after 60 years of use. *Ann Occup Hyg*; **58**: 955–70.

Bernard TE, Luecke CL, Schwartz SW *et al.* (2005) WBGT clothing adjustments for four clothing ensembles under three relative humidity levels. *J Occup Environ Hyg*; **2**: 251–6.

Bernard TE, Caravello V, Schwartz SW *et al.* (2008) WBGT clothing adjustment factors for four clothing ensembles and the effects of metabolic demands. *J Occup Environ Hyg*; **5**: 1–5; quiz d21–3.

Budd GM. (2008) Wet-bulb globe temperature (WBGT)—its history and its limitations. *J Sci Med Sport*; **11**: 20–32.

Dukes-Dobos F, Henschel A. (1973) Development of permissible heat exposure limits for occupational work. *ASHRAE J* **9**: 57–62.

Gallop RJ. (2001) Determination and Interpretation of the OOP for ROC’s with PROC LOGISTIC. *Proceeding of NESUG*. Baltimore, MD; pp. 777–782.

Garzon-Villalba XP, Mbah A, Wu Y, Hiles M, Moore H, Schwartz SW, Bernard TE. (2016) Exertional heat illness and acute injury related to ambient wet bulb globe temperature. *Am J Ind Med*; **59**(12): 1169–1176.

Gönen M. (2006) Receiver operating characteristic (ROC) curves. SAS Users Group International (SUGI); **31**: 210–231.

Havenith G. (2001) Human surface to mass ratio and body core temperature in exercise heat stress—a concept revisited. *J Thermal Biol*; **26**(4): 387–393.

ISO. (1989) *ISO 7243, Hot environments—estimation of heat stress on working man, based on the WBGT-index (wet bulb globe temperature)*. Geneva: International Standards Organization.

Kuhlemeier KV, Miller JM, Dukes-Dobos FN *et al.* (1977) Determinants of the prescriptive zone of industrial workers. *J Appl Physiol Respir Environ Exerc Physiol*; **43**: 347–51.

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.

Lind AR, Humphreys PW, Collins KJ *et al.* (1970) Influence of age and daily duration of exposure on responses of men to work in heat. *J Appl Physiol*; **28**: 50–6.

NIOSH. (2016) *Criteria for a recommended standard: occupational exposure to heath and hot environments*. By Jacklitsch B, Williams WJ, Musolin K, Coca A, Kim J-H, Turner N. Cincinnati, OH: Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH). Publication 2016–106.

R Development Core Team. (2015) *R Core Team*. 2013. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

SAS Institute Inc. (2013) Cary, NC, USA.

Shin S, Coulter B. (2009) *ROC analysis for the evaluation of continuous biomarkers: existing tools and new features in SAS® 9.2*. Paper SP09-2009. San Diego: Beckman Coulter.

Wu Y, Wang X. (2011) Optimal weight in estimating and comparing areas under the receiver operating characteristic curve using longitudinal data. *Biom J*; **53**: 764–78.

Wyndham CH, Heyns AJ. (1973) The probability of heat stroke developing at different levels of heat stress. *Arch Sci Physiol (Paris)*; **27**: 545–62.