Test-Retest Reliability of Thermal Temporal Summation Using an Individualized Protocol

Jiang-Ti Kong[1], **Kevin A. Johnson**[1], **Raymond R. Balise**[2], and **Sean Mackey**[1]

[1]Stanford University, Department of Anesthesia, Division of Pain Medicine, Stanford Neuroscience and Pain Laboratory

[2]Stanford University, Department of Health Research and Policy

Abstract

Temporal summation (TS) refers to the increased perception of pain with repetitive noxious stimuli. It is a behavioral correlate of wind-up, the spinal facilitation of recurring C fiber stimulation. In order to utilize TS in clinical pain research, it is important to characterize TS in a wide range of individuals and to establish its test-retest reliability. Building on a fixed-parameter protocol, we developed an individually adjusted protocol to broadly capture thermally-generated TS. We then examined the test-retest reliability of TS within-day (inter-trial intervals ranging from 2 minutes to 30 minutes) and between-days (inter-session interval of 7 days). We generated TS-like effects in 19 of the 21 participants. Strong correlations were observed across all trials over both days [ICC (A, 10) = 0.97, 95% CL = 0.94 to 0.99] and across the initial trials between days [ICC (A, 1) = 0.83, 95% CL = 0.58 to 0.93]. Repeated measures mixed effects modeling demonstrated no significant within day variation, and only a small (5 out of 100 point) between-day variation. Finally, a Bland-Altman analysis suggested that TS is reliable across the range of observed scores. Without intervention, thermally-generated TS is generally stable within-day and between days.

Perspective: Our study introduces a new strategy to generate thermal TS in a high proportion of individuals. This study confirms the test-retest reliability of thermal TS, supporting its use as a consistent behavioral correlate of central nociceptive facilitation.

Keywords

temporal summation; second pain; thermal; individualized; test-retest reliability

INTRODUCTION

Temporal summation (TS) refers to the increase in perceived pain from repetitive, noxious stimuli delivered at frequencies higher than 0.33Hz.[25, 3] TS is also known as temporal

Corresponding Author: Sean Mackey, MD/PhD, Associate Professor, Stanford Systems Neuroscience & Pain Laboratory, 780 Welch Road #208, Palo Alto, CA 94304, Office: 650-725-9636, Fax: 650-725-9642, smackey@stanford.edu.

Disclosure:
No conflicts of interest exist for any of the authors. Nor is there commercial interest to be disclosed in the execution or the experiment or in the preparation of the manuscript.

summation of second pain (TSSP), because it is thought to represent summation of C fiber-mediated second pain.[24] In the distal extremities, second pain is often perceived after first pain, which is mediated by A-δ fibers.[24, 37] TS represents the behavioral correlate of wind-up, a physiological phenomenon of central facilitation where increased firing of spinal secondary neurons results from repetitive C fiber stimulation.[25, 11]

TS can be generated by a variety of noxious stimuli, such as heat, pressure, and pin-prick.[11] In the present study we focus on TS induced by phasic thermal stimuli (heat pulses). Phasic thermal TS paradigms have gained popularity because of their tolerability, ease of administration and standardization.[15, 33] As with the generation of TS, a great variety of methods exist to quantify TS. Currently there is no consensus on the best way to quantify thermal TS.[9] We quantified TS by subtracting the pain rating of the first heat pulse from that of the most painful heat pulse. We chose this method because it has been used by many other researchers[9, 10, 6, 16, 28] and best captures the pain increase due to sequential heat pulse administration.[10] We also compared TS calculated by this method to two other commonly used approaches (see Materials and Methods).

The concept of TS has evolved from a laboratory observation to a behavioral measure of central sensitization.[30, 4] For example, compared to healthy controls, investigators discovered augmented TS in several chronic pain conditions, such as fibromyalgia (FM),[35] chronic pelvic pain,[23] and temporomandibular joint dysfunction.[31] We propose that TS may further be utilized as a longitudinal behavioral marker to quantify therapeutic response to treatment. Prerequisite to achieve this objective are the ability to: (1) elicit TS in a broad range of individuals, and (2) characterize the stability of TS over time. However, neither prerequisite has been fully achieved due to the following limitations.

Regarding the first prerequisite, most current thermal TS protocols apply the same baseline and peak pulse temperatures for all individuals, resulting in up to 50% of subjects failing to exhibit any TS-like perception.[26, 15, 27] Earlier researchers generated heat pulses by intermittently contacting the palmar skin of the subject via a preheated thermode.[37, 35] Using this method, it was unclear how much the intermittent physical contact by the thermode contributed to alternations in pain perception. Subsequently, with the advent of the CHEPS thermode (Medoc Ltd, Israel), which is capable of rapidly cycling between a wide range of temperatures, more researchers have switched to using this continuous-contact thermode to generate TS.[33] While the use of the CHEPS thermode eliminates the need for intermittent contact, it is even more difficult to generate TS with this thermode.[37, 9, 33] To increase the capture of TS, researchers have recruited a large number of participants and/or performed several trials with multiple different temperature settings.[9, 29, 16, 32, 36] The resulting group-averaged TS is small (<10 on a 100-point numerical rating scale) or at best moderate (< 20).[9, 29, 16, 32, 36] In the current study, we propose an individualized method to optimize TS via systematic adjustment of both the baseline and peak pulse temperatures. Both these temperatures have been shown to critically influence the magnitude of TS.[37, 21]

Regarding the second prerequisite, the test-retest reliability of phasic thermal TS over time has only been partially established. Two groups investigated the subject and reported high within-session reliability (one group)[2] but limited between-session reliability (both groups).[15, 2] However, the degree of variation in TS between sessions has not been sufficiently characterized. In the current study, we characterized both the within-session and the between-session reliability of thermal TS using the appropriate intra-class correlations.[38] We also quantified the variation of TS over time using a mixed-effect model.[13]

The primary aims of the this psychometric study in healthy subjects were thus two-fold: (1) to develop an individualized protocol to maximally capture TS with the continuous contact

thermode, and (2) to characterize both short-term (within-day) and long-term (between-day) test-retest reliability of thermal TS generated with the above individually optimized protocol.

MATERIALS AND METHODS

Participants

We recruited healthy adult volunteers from the general community via an Internet advertisement. The inclusion criteria were: (1) between 18 and 50 years of age, (2) healthy. The exclusion criteria were: (1) presence of any acute or chronic pain condition, (2) current use of antihypertensive medications, (3) daily use of allergy medications, (4) current use of antidepressants, (5) any major systemic or psychiatric illnesses (including but not limited to depression, anxiety and PTSD), and (6) inability to understand conversational English or follow instructions. The experimental protocols were approved by the Stanford University Institutional Review Board. All participants provided informed consent prior to participation in the study.

Heat-Pulse Administration

We administered heat pulses and acquired real-time pain ratings using a Pathway Contact Heat-Evoked Potential Stimulator (CHEPS) system (Medoc LTD, Israel). A CHEPS 2.7 cm-diameter thermode delivered heat pulses to the palmar skin over the thenar eminence of the non-dominant hand in the first trial of each session. Subsequently, we asked participants to alternate hands between each trial to minimize skin irritation. Adapting from Staud et al 2006,[33] we applied ten, identical, 0.5-second-long, heat pulses with an end-to-end inter-stimulus interval (ISI) of 2 seconds, at a ramp-up and ramp-down rate of 40 °C/second. The baseline temperatures ranged from 38 to 42°C, and the stimul us temperature ranged from 47 to 51°C (individually optimized as described in the next section).

The participants used a COVAS (Medoc LTD, Israel) box with a visual analog scale (VAS) to rate the intensity of the pain of each pulse continuously. We read a standard script to each participant (see supplemental materials), in which we defined the VAS scale, instructed each participant to rate only the slow, burning sensation from each pulse (second pain), and not the prickly sensation immediately felt at the delivery of each pulse (first pain).[24, 25, 5] We warned participants to expect a delay between the delivery of each heat pulse and the perception of the second pain. Each participant underwent approximately 15-20 minutes of training to be familiarized with: the VAS scale, the CHEPS thermode, use of the COVAS, and rating second pain in a mock TS trial on Day 1. Data from the training trial were not included in analysis. The participants then took a 15-min break before starting the optimization trials as described below.

Individualized Optimization of TS Parameters

We optimized the magnitude of TS by systematically adjusting the baseline temperature ($Base_{temp}$) and peak heat-pulse temperature ($Peak_{temp}$). During optimization, we defined the estimated TS magnitude, TS_E, as the difference in pain ratings between the first and the most painful heat pulse. The magnitude of these pain ratings were visually estimated from a real-time VAS vs time plot generated by the COVAS program immediately after each trial. (Supplemental Material)

First, we started all participants at the same initial TS trial, T1, where $Base_{temp}$ = 40 °C and $Peak_{temp}$ = 49 °C. We chose these two temperatures because e ach represented an average of what had been used in the literature ($Base_{temp}$ ranged between 38 and 42°C and $Peak_{temp}$ between 47 and 51°C). [25, 15, 33, 7, 29]

Next, we systematically adjusted $Base_{temp}$ and $Peak_{temp}$ to achieve TS_E between 30 and 70 VAS. We chose this intermediate range of TS_E (50 +/− 20) both to avoid floor and ceiling effects and to allow sensitive detection of changes in TS over time. Although our goal was to optimize TS_E between 30 and 70, many of our participants initially demonstrated low (<30) rather than high (>70) TS_E. Consequently we ended up maximizing TS_E for the majority of our participants.

Figure 1 summarizes the details of our optimization protocol. Each participant may have up to 5 optimization trials with 15 min between each trial to minimize carry over effects. If TS_E was less than 30 VAS, we increased $Peak_{temp}$ by 1°C to 50°C in Trial 2 (T2). If TS_E was still less than 30, we increased $Peak_{temp}$ again by 1°C to 51°C in Trial 3 (T3). If TS_E was still less than 30 VAS, we increased $Base_{temp}$ to 41 (T4) then to 42°C (T5). If, after all five t rials, TS_E was still less than 30 VAS pain points, we took the combination of $Peak_{temp}$ and $Base_{temp}$ that resulted in the highest TS_E as the final $Base_{temp}$ and $Peak_{temp}$ to be used in all subsequent trials. We followed a similar procedure for cases where TS_E was greater than 70 VAS points, except that we adjusted the $Base_{temp}$ before the $Peak_{temp}$ in this scenario. We inferred from Mauderli et al[21] and from our pilot data that $Base_{temp}$ exerts a stronger influence on TS magnitude than $Peak_{temp}$. Therefore we manipulated the former first, in cases of high TS. However, we increased $Peak_{temp}$ first, in cases of low TS in order to minimize inadvertent burn injury.

In rare cases, when a full degree Celsius change resulted in over-correcting TS_E in the opposite direction, we used a half a degree change in the following trial to achieve the target TS_E. For example, if TS_E was 20 during Trial 1 ($Peak_{temp}$ = 49°C), but became 90 in Trial 2 where $Peak_{temp}$ = 50°C, we would decrease $Peak_{temp}$ by 0.5°C to 49.5°C in Trial 3.

Lastly, almost half the participants rated the first heat pulse as very painful (>70 VAS), which could significantly limit the magnitude of TS due to a ceiling effect. Therefore, before moving on to the optimization algorithms for TS_E, we decreased the $Base_{temp}$ and/or $Peak_{temp}$ by 1-2°C in a manner similar to "Scenario C" in Figure 1 b), to reduce the estimated pain rating of the first pulse, $P1_E$, to be   50. Then we moved on to optimize TS_E. However, each participant could undergo no more than a total of 5 optimization trials, including the trials to bring $P1_E$ 50 and those to bring TS_E between 30 and 70.

Timeline for Assessing Test-retest Reliability

As shown in Figure 2, on Day 1, participants underwent up to five adjustment trials to optimize $Base_{temp}$ and $Peak_{temp}$. On Day 4, using the temperatures optimized on Day 1, the participant underwent five, consecutive TS trials with inter-trial intervals in the following order: 30 min, 15 min, 5 min, and 2 min. On Day 11, Day 4's trials were repeated. We intentionally tested unequal inter-trial intervals because of our a priori hypothesis that TS would be more stable with longer inter-trial intervals. We based this hypothesis on a small pilot study and the knowledge that certain physiological processes such as habituation[17, 18] and sensitization[18] tend to take place immediately after the heat stimulus.

Data Analysis

Instead of performing visual estimation as in Day 1, we calculated the magnitude of TS (TS_{mag}) for all trials on Days 4 and 11 using an algorithm described below. First, due to the transmission time of action potentials in the peripheral and central nervous system, the participants did not record the response to the first heat pulse until well into the latter part of the second pulse. Based on the conduction speed of C-fibers ($\approx$1m/s),[22] the brain's processing time (including perception, decision, and execution, approximately 500-800ms),[1, 19] and the minimal amount of time needed to activate the motor fibers,[22] we

estimated that the rating of the second pain to each heat pulse occurred roughly 1.5-2 seconds after it was delivered. This observation was consistent with findings by other researchers.[37, 21] Second, we examined all of the TS curves in this experiment and found that although the lapse time varied between delivery of heat pulse and rating of second pain on the COVAS, a stable increase in pain recording in response to the first heat pulse consistently occurred between 3.9 to 4 second, values which fell within the range of our theoretical estimate and the literature.[37, 21] Thus, the pain response to the first pulse (P1) was calculated as the average pain rating between 3.9 and 4 seconds. Third and finally, the peak pain score (P_{max}), occurring between 0 to 22 seconds, was identified. Even though the trial lasted only 20 seconds, given the delayed rating of the second pain, we decided to include pain ratings between 20 –22 seconds to capture the pain response to the 10th pulse. The magnitude of temporal summation (TS_{mag}) was thus calculated as:

$$TS_{mag} = P_{max} - P1.$$

To assess if our results would apply to TS calculated by other methods, we calculated the magnitude of TS with two commonly used alternative approaches: one which subtracts the pain rating of the first heat pulse from that of the last one[15, 34] (in our case, 10th); and another using linear regression to calculate the slope of pain ratings in response to the first 5 heat pulses as a function of time.[6, 15, 2, 16] We then calculated the correlations between TS_{mag} calculated by our method and those by the other methods.

We used three, successive methods to assess the test-retest reliability of TS_{mag}. First, we estimated the agreement amongst TS trials from various time points via intra-class correlations (ICC).[38] Specifically, the correlation amongst all 10 trials on Day 4 and Day 11 was calculated as ICC (A, 10). Next, the correlation between the first TS trials on Day 4 and Day 11 was calculated as ICC (A, 1). We picked the first trials from Day 4 and Day 11 because in clinical settings, we expected that an intervention's effect would be assessed by measuring a parameter once before and then again after the intervention. Second, we characterized the variability in TS_{mag} due to the day effect and trial (within-day) effect by using repeated measures mixed-effect modeling with fixed effects for day and trial.[13] Temporal/spatial, unstructured, and compound symmetry covariance structures were fitted. Third, we constructed a Bland-Altman plot to identify any bias between the variation and the size of TS_{mag}.[13] Given that the Bland-Altman plot cannot be applied to more than 1 pair of comparisons, as with ICC (A, 1), we picked the first trials from Day 4 and Day 11 to mimic clinical scenarios. Finally, we calculated the contrast between these TS values via paired t-test to provide an alternative estimate of the day effect.

We then performed the ICC and mixed effects analysis for test-retest reliability on TS_{mag1} and TS_{mag2}. All analyses were conducted with SAS 9.3 (SAS Institute, Cary, NC) except intra-class correlations, which were conducted with R 2.14.0 using the IRR package (version 0.83). For all tests, two sided P-values < .05 were considered statistically significant.

RESULTS

Participant Characteristics

A total of 27 healthy volunteers were recruited for the study. Four individuals could not continue due to scheduling incompatibility. One person deferred the study due to significant anxiety, whereas another individual had trouble following directions due to language issues. Thus a total of 21 participants completed the study. Out of these, only 2 were not able to achieve any TS (TS_E 0) on Day 1 and thus did not continue in the study. Specifically, both subjects rated the first pulse >70 VAS, and had negative temporal summation scores (the

first pulse was the most painful and pain decreased with each successive pulse) consistently across all temperature settings trialed on Day 1. The remaining 19 participants completed all three sessions. Of these 19 individuals, 10 were male (average age = 29.7, SD = 10.9) and 9 were female (average age = 28.2, SD = 10.9).

Results from Individualized Optimization of TS Parameters (Day 1)

Using our method for individual parameter optimization, we achieved positive TS_E in 19/21, or 90% of healthy participants in this study. As shown in Table 1: nine needed to go through all five optimization trials; it took on average 4 trials to reach the final parameters. The average $Base_{temp}$ was 39.9°C and average $Peak_{temp}$ was 49.3°C. The average of the optimized TS_E on Day 1 was 29.1 (range 7 - 64, SD=19.5). The time-to-peak-pain ranged from the second to the 10th pulse, and the average time-to-peak-pain was roughly 6 pulses.

Calculated TS$_{mag}$ from Day 4 and 11

Figure 3 summarizes TS_{mag} (peak pain - first pulse pain) in the form of a box plot.[39] It includes 5 trials from Day 4 and 5 trials from Day 11 by all 19 participants. For each participant, all ten trials were identical: 10 pulses at 0.5Hz with 0.5 second pulse duration, using baseline and stimulus temperatures optimized on Day 1. The mean and standard deviation (SD) of TS_{mag} from these 2 days are summarized in Table 2. For comparison, Table 2 also includes the estimated TS (TS_E) from the best trial on Day 1 for each participant.

We noted in Table 2 an 11- and 7- VAS point difference between the across-subject average of TS_E and that of TS_{mag} on Day 4 and 11, respectively. The TS_E was higher. However, at least three reasons may explain the different TS magnitudes: 1) a "startle" effect may have been present given that the participants encountered thermal TS protocols for the first time on Day 1. This "startle effect" has been demonstrated in a different QST study by Bishop et al and may take up to 2 sessions to disappear.[8] 2) TS_E was measured only once while TS_{mag} was measured 5 times on Day 4 and 11. 3) TS_E was visually estimated immediately after each trial run while TS_{mag} was computed by SASS program according to strict protocols. Given these concerns, we did not feel the testing conditions on Day 1 were identical to those on Day 4 and 11. Thus, a direct comparison of the magnitude of TS_E on Day 1 to TS_{mag} on Day 4 and 11 would be neither possible nor meaningful.

Alternative Ways to Quantify TS

We also calculated TS using 2 additional methods: subtracting pain of the first pulse from that of the last (10th) pulse (designated as TS_{mag1}), and calculating the rate of pain increase during the first five pulses (TS_{mag2}) by ordinary least squares regression. We summarized these results in Supplemental Materials under items III and IV. The correlations between TS_{mag}, TS_{mag1} and TS_{mag2} are listed in Table 3. They ranged from 0.72 (between TS_{mag1} and TS_{mag2}) to 0.85 (between TS_{mag} and TS_{mag1}) and 0.91 (between TS_{mag} and TS_{mag2}), demonstrating a high degree of consistency between all three methods to quantify TS.

Test-retest reliability of TS$_{mag}$

Three methods demonstrated excellent test-retest reliability of TS_{mag} measured on Day 4 and 11, with higher reliability within session than between sessions.

First, we calculated two types of intra-class correlations (ICC) to assess global consistency between trials. Both values were high: a) ICC across all 10 trials on both Day 4 and Day 11, calculated as ICC (A, 10), was 0.97 [95% CL = 0.94 to 0.99]; b) ICC of the first trial from both days, ICC (A, 1), was found to be 0.83 [95% CL = 0.58 to 0.93]. The decrease in ICC

from (A, 10) to (A, 1) suggested stronger correlation within-day than between-days, a result supported by the models below.

Second, we used repeated measures mixed effects models to quantify the variability of TS. Using a compound symmetry (CS) covariance structure (refer to Supplemental Material, item V, on how we arrived at this covariance structure), we discovered no statistically significant variation between trials within the same day, but a significant (*P<0.0005*), small (5/100 point), difference in TS_{mag} between days. Furthermore, not only did the trial effect not reach significance (*P* value between 0.2 and 0.8), but the between-trial differences of TS_{mag} (within the same day) were also small, ranging from –3 to 1 (out of 100 VAS points). These numbers confirmed that regardless of the length of the inter-trial interval, TS was stable within-session, thus negating our previous hypothesis of more stable TS with longer inter-trial intervals. Finally, the models were repeated with log-transformed TS scores to account for outliers, which did not affect the results.

Third, to identify systematic biases as a function of the size of the TS_{mag}, we produced a Bland-Altman plot, which showed a systematic difference across days but no bias as a function of the magnitude of the TS_{mag} (Figure 4). The 5 point increase in TS_{mag} from the first trial on Day 4 to that of Day 11 seen in Fig 4 was demonstrated to be statistically significant via paired t-test (*p < .05*). The 95% confidence interval of this difference was between 0.03 and 10.

Test-retest reliability of TS_{mag1} and TS_{mag2}

As an exploratory exercise, we performed the same analysis (mixed effects models and ICC) of reliability for TS_{mag1} and TS_{mag2}. The mixed effects models of TS_{mag1} and TS_{mag2} showed comparable results to that of TS_{mag}: no trial effect existed but a slight day effect did where TS increased by a small amount (4-5/100) from Day 4 to 11. Calculations of the ICC's showed high degrees of correlation (ICC[A,10]>0.9) for all 10 trials from both days 4 and 11. However, the correlation between the first trial from Day 4 and 11 was less with TS_{mag1} (ICC[A,1]=0.58) and even less with TS_{mag2} (ICC[A,1]=0.25) compared to the same correlations computed for TS_{mag}.

DISCUSSION

Using an individualized optimization protocol, we generated thermal TS in most (19/21) of the participants. We also demonstrated a high level of test-retest reliability of thermal TS both within an hour testing session and between sessions over one week.

Comparison between our individualized protocol and fixed temperature protocols

The magnitude of TS generated by our protocol was comparable to those of the traditional fixed temperature protocol.[9, 26, 16] We obtained TS with an average between 17.7 (Day 4) and 22.3 VAS points (Day 11) while traditional fixed-parameter methods generate TS ranging between 10-20.[9, 26, 16] However, we were able to achieve some degree of increase of pain rating in 19 out of the 21 individual tested while a success rate of 50-60% is reported in the literature,[9, 26, 16] where "success rate" is often not defined. We also heard similar numbers upon communication with other investigators (including Drs. Joel Riley and Roger Fillingim). Finally, we note only 33% of the subjects (7/21) reached our original TS goal (30 - 70). A more realistic TS goal of 10 or 20 (comparable to that reported in the literature), would lead to 17/21 (81%), and 10/21 (48%) participants reaching goal, respectively.

Compared to fixed temperature protocols, our individualized protocol is better-suited for within-individual monitoring because it captures TS in most people and minimizes both floor and ceiling effects (so it may be more sensitive to detecting changes). On the other

hand, it may not be ideal for direct between-group comparisons (such as patients versus controls), because the parameters are customized for each individual. Our individualized method thus offers a useful alternative to the fixed-parameter method to obtain TS with an advantage for longitudinal, within-individual monitoring.

Comparison to other individualized protocols

To the best of our knowledge, only two groups have attempted to individually adjust TS parameters using the continuous-contact thermode.[15, 34] As the detailed methods of the individual optimization of TS were abbreviated in the publication, the authors did not report several important parameters, including: threshold for "adequate" TS, the number of trials allowed, and, most importantly, their success rate in achieving TS-like effects. As such, although the magnitude of the group average TS from both studies was similar to ours ($\approx$20),[15, 34] we cannot directly compare their methods with ours.

Our optimization differed from the previous individualized protocols in that we adjusted both the baseline and the peak stimulus temperatures while the previous methods varied the PEAK stimulus temperature only. Both Mauderli et al[21] and Vierck et al[37] established that the baseline skin temperature between each heat pulse contribute critically to the degree of thermal TS (more than the contribution from peak temperature). Therefore, by varying the baseline skin temperature systematically in addition to varying the stimulus temperature, we were likely more able to optimize the magnitude of TS than those varying only the peak temperatures. However, a direct, side-by-side comparison would be needed to demonstrate the superiority of our method.

Between-individual variation of TS magnitude

Our results revealed large variations in the magnitude of TS (Table 1) amongst a relatively homogeneous group of young (average age $\approx$ 29), healthy individuals. This finding was consistent with results from other researchers.[9, 14, 12, 29] The variation was partially attributable to factors including age,[9] sex,[14, 12] fear and anxiety.[14, 29] In our study, we did not find any significant correlation ($r > 0.5$, $P < 0.05$) between TS magnitude and these factors (see item VI of Supplemental Materials for details). We expected this result because each participant had a unique set of TS parameters, so our comparison was not based on equal ground. Finally, the large variation in TS necessitates the recruitment of large number of subjects in clinical trials so that groups can be compared, e.g., patients with chronic pain vs. healthy controls.

The Variation of TS_{mag} over Time

Over the course of 1 hour in the same day, there was no significant variation in TS_{mag}. Specifically, the repeated measure models did not reveal any trial effect. This finding was corroborated by a very high ICC score [ICC(A, 10)=0.97]. A variety of peripheral and central factors, such as sensitization,[18] habituation,[17, 18] and cognitive reappraisal[20] tend to take place immediately after the stimulus, thus could change the TS magnitude if the trials were repeated shortly after one another. We had thus hypothesized that TS would vary more with closer inter-trial intervals than with longer intervals. However, our results from both the mixed effects models and ICC analyses indicated high reliability within one hour and no significant difference in TS magnitudes between inter-trial intervals of 30 minutes and 2 minutes.

Over the course of 1 week, there was a small, but statistically significant increase in TS_{mag}. Specifically, both the repeated measures models (group analysis using all five trials from each experimental day) and the Bland-Altman plot (analysis involving only the first trials from Day 4 and Day 11) showed a significant 5-point increase in TS over 7 days. A paired t-

test on this contrast had a 95% CL of 0.03 to 10, suggesting a 5-point increase was within possible error range. Despite this increase, the test-retest reliability was still excellent, as reflected by an intra-class correlation, ICC (A, 1) of 0.83.

Our results were consistent with previous findings and we further extended the knowledge on the long-term variability of TS. Both Granot et al[15] and Alapattu et al[2] demonstrated high within session reliability but lower between-session reliability without describing the range. With similar findings, we also quantified the mean between-session variability of TS to be 5 VAS points, with a 95% confidence interval between 0.03 and 10 points. A comparison of the degree of reliability (represented by ICC's) of TS obtained by us to those by Granot and Alappattu revealed that ours were higher. This difference may be due to the different methods to quantify TS (see below).

Alternative Methods for Quantifying TS and Impact on the Test-retest Reliability

A variety of indices exist to quantify TS: the highest pain rating minus the first pain rating,[9, 10, 6, 16, 28] the last pain rating minus the first pain rating,[15, 34] the pain rating of the fifth heat pulse alone,[7, 36] or the slope of the pain ratings versus time during the first 3[16], 5[2], or 6[6] pulses. We chose the first option because it reflects the maximum amount of TS obtained[10, 28] and is less prone to variations in rating/reaction time as is the case with slope measures based on our pilot studies.

We compared our method to quantify TS (TS_{mag}, "max-first") to two other commonly used methods using the "last pain minus 1st pain rating" (TS_{mag1}) and the "slope of the first 5 pulses" (TS_{mag2}). We found high levels of correlations between (TS_{mag}) and these other two: r=0.85 with TS_{mag1} and 0.91 with TS_{mag2}.

TS_{mag1} and TS_{mag2} showed similar within-session stability as TS_{mag} but less between-session stability. Analysis by mixed effect models revealed no statistically significant trial effect but a slight, 5/100 VAS point day effect. Similarly, ICC (A, 10) on all ten trials from Days 4 and 11 showed high reliability for TS_{mag1} (0.95) and TS_{mag2} (0.92). However, correlation analysis of the first trials from Days 4 and 11 showed relatively poor ICC (A, 1) of 0.58 for TS_{mag1}, and 0.25 for TS_{mag2}.

A closer look at the between-session ICC results from previous authors suggests similar values to our calculations if we consider the different methods of calculating TS. ICC from Alappattu[2] showed a value of 0.3 if TS were calculated by the slope of pain from the first 5 pulses vs time (compared to ICC of 0.25 in our study for TS_{mag2}). The ICC becomes higher (~0.4) if the difference between the first and the fifth pulse were taken as TS (compared to 0.58 for TS_{mag1} in our study). Granot used a different method to generate TS (20 rather than 10 heat pulses), calculate TS (pain from 20th pulse minus pain from first pulse, or log pain divided by log time for all 20 pulses) and to represent reliability (95% repeatability).[15] These differences make a direct comparison difficult but the overall message was similar: significant between-session variability when TS was calculated by slope or change score of last rating minus 1st rating.

We therefore conclude that compared to TS_{mag}, TS_{mag1} and TS_{mag2} showed similar, high within-session reliability. For between-sessions, TS_{mag} demonstrated more reliability than TS_{mag1} and TS_{mag2}. Therefore, if one desires to use TS for between-session contrasts such as in longitudinal studies, it may be preferable to use TS_{mag} (max-first) rather than TS_{mag1} (last-first) or TS_{mag2} (slope) to quantify TS, as the former showed the most temporal stability.

Limitations and Further Studies

We recognize limitations in both the individual optimization and the test-retest portions of the study, as well as in our target population.

The limitations of individual optimization are two-fold. First, we focused on adjusting temperature settings and did not test other parameters, such as frequency or pulse duration, factors also known to influence the degree of TS.[37, 21] Previous studies reported optimal frequencies between 0.33-0.5Hz and pulse durations between 0.5-0.75 seconds.[25, 37, 21, 15, 33, 29] We therefore selected an optimal frequency of 0.5Hz and pulse duration of 0.5 seconds in keeping with prior work. Second, it took a little over an hour to optimize the temperatures in the current study. Researchers may not be able to afford this time in complex clinical trials. However, the total optimization time can reduced to 20 minutes by decreasing the inter-trial interval from 15 to 2 minutes, as our study showed high test-retest reliability of TS whether the trials are 15 or 2 minutes apart.

The test-retest part of the study is limited by duration. We examined 7 days as the longest period; larger TS variations may exist over longer periods. The relevance of this limitation is that many treatments may not show beneficial effects on chronic pain until several weeks or months later. Therefore, an appropriate follow-up study may focus on the stability of TS over longer periods of time, e.g., months.

Finally, our study was carried out in healthy participants, but we believe the customized TS settings will be especially useful for chronic pain patients, given the large variations in peripheral and central sensitivity in these individuals. It would also be important in the future to characterize the stability of thermal TS in these patients whose pain sensitivity often varies over time.

Significance for Clinical Research

Despite the limitations, this study introduced a novel method to individually optimize TS that captured TS in a high proportion of subjects, and provided important evidence for the stability of thermal TS within-day and over the course of one week. Additionally, by contrasting several methods to quantify TS, this study demonstrated that subtracting pain ratings of the first pulse from the most painful pulse provides the most stable measure of TS.

This study has several implications for clinical trials with TS as a measure, specifically for longitudinal studies and interventional studies. First, we recommend individualized optimization of TS parameters to maximize detection of within-individual changes. Second, we recommend using firstMax method to quantify TS as this is most stable over time. Third, one measure at each time point is sufficient given the high within-session stability. Fourth, a small increase of TS may occur over time. Finally, as we calibrate TS to the intermediate range in our individualized method, the sensitivity measure would then be the temperatures used to generate the calibrated TS if one is interested in between-individual comparison.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Aarts E, Roelofs A, van Turennout M. Attentional control of task and response in lateral and medial frontal cortex: brain activity and reaction time distributions. Neuropsychologia. 2009; 47:2089–2099. [PubMed: 19467359]

2. Alappattu MJ, Bishop MD, Bialosky JE, George SZ, Robinson ME. Stability of behavioral estimates of activity-dependent modulation of pain. Journal of pain research. 2011; 4:151–157. [PubMed: 21811390]

3. Arendt-Nielsen L, Petersen-Felix S. Wind-up and neuroplasticity: is there a correlation to clinical pain? European journal of anaesthesiology Supplement. 1995; 10:1–7. [PubMed: 7641635]

4. Arendt-Nielsen L, Yarnitsky D. Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. The journal of pain: official journal of the American Pain Society. 2009; 10:556–572. [PubMed: 19380256]

5. Beissner F, Brandau A, Henke C, Felden L, Baumgartner U, Treede RD, Oertel BG, Lotsch J. Quick discrimination of A(delta) and C fiber mediated pain based on three verbal descriptors. PloS one. 2010; 5:e12944. [PubMed: 20886070]

6. Bhalang K, Sigurdsson A, Slade GD, Maixner W. Associations among four modalities of experimental pain in women. The journal of pain: official journal of the American Pain Society. 2005; 6:604–611. [PubMed: 16139779]

7. Bialosky JE, Bishop MD, Robinson ME, Zeppieri G Jr. George SZ. Spinal manipulative therapy has an immediate effect on thermal pain sensitivity in people with low back pain: a randomized controlled trial. Physical therapy. 2009; 89:1292–1303. [PubMed: 19797305]

8. Bishop MD, Craggs JG, Horn ME, George SZ, Robinson ME. Relationship of intersession variation in negative pain-related affect and responses to thermally-evoked pain. The journal of pain: official journal of the American Pain Society. 2010; 11:172–178. [PubMed: 19853524]

9. Edwards RR, Fillingim RB. Effects of age on temporal summation and habituation of thermal pain: clinical relevance in healthy older and younger adults. The journal of pain: official journal of the American Pain Society. 2001; 2:307–317. [PubMed: 14622810]

10. Edwards RR, Ness TJ, Weigent DA, Fillingim RB. Individual differences in diffuse noxious inhibitory controls (DNIC): association with clinical variables. Pain. 2003; 106:427–437. [PubMed: 14659526]

11. Eide PK. Wind-up and the NMDA receptor complex from a clinical perspective. Eur J Pain. 2000; 4:5–15. [PubMed: 10833550]

12. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL 3rd. Sex, gender, and pain: a review of recent clinical and experimental findings. The journal of pain: official journal of the American Pain Society. 2009; 10:447–485. [PubMed: 19411059]

13. Geoffrey, R.; Norman, DLS. Health Measurement Scales: A practical guide to their development and use. 4th Edition. Oxford University Press; 2008. Reliability; p. 167-210.Chapter 8

14. George SZ, Wittmer VT, Fillingim RB, Robinson ME. Sex and pain-related psychological variables are associated with thermal pain sensitivity for patients with chronic low back pain. The journal of pain: official journal of the American Pain Society. 2007; 8:2–10. [PubMed: 17207739]

15. Granot M, Granovsky Y, Sprecher E, Nir RR, Yarnitsky D. Contact heat-evoked temporal summation: tonic versus repetitive-phasic stimulation. Pain. 2006; 122:295–305. [PubMed: 16540248]

16. Greenspan JD, Slade GD, Bair E, Dubner R, Fillingim RB, Ohrbach R, Knott C, Mulkey F, Rothwell R, Maixner W. Pain sensitivity risk factors for chronic TMD: descriptive data and empirically identified domains from the OPPERA case control study. The journal of pain: official journal of the American Pain Society. 2011; 12:T61–74. [PubMed: 22074753]

17. Greffrath W, Baumgartner U, Treede RD. Peripheral and central components of habituation of heat pain perception and evoked potentials in humans. Pain. 2007; 132:301–311. [PubMed: 17533117]
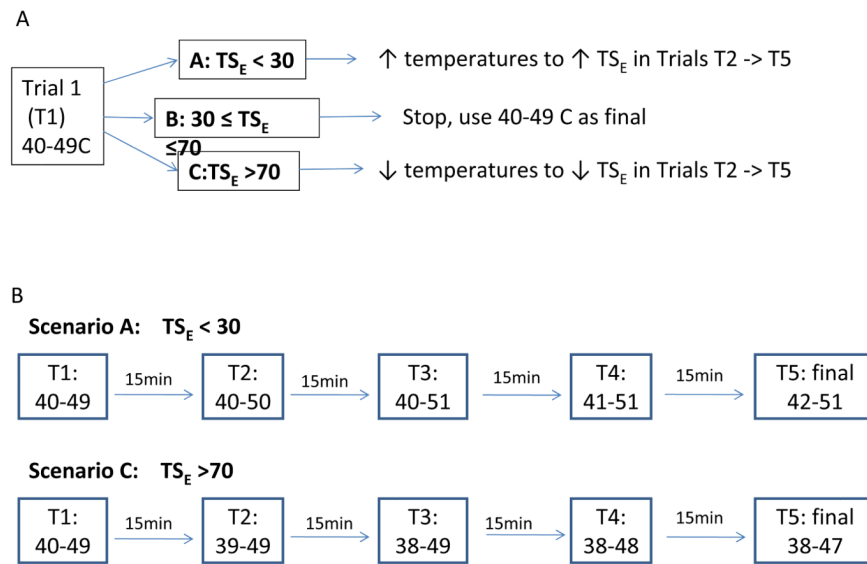
18. Hollins M, Harper D, Maixner W. Changes in pain from a repetitive thermal stimulus: the roles of adaptation and sensitization. Pain. 2011; 152:1583–1590. [PubMed: 21454015]

19. Konrad A, Vucurevic G, Musso F, Stoeter P, Winterer G. Correlation of brain white matter diffusion anisotropy and mean diffusivity with reaction time in an oddball task. Neuropsychobiology. 2009; 60:55–66. [PubMed: 19752579]

20. Lawrence JM, Hoeft F, Sheau KE, Mackey SC. Strategy-dependent dissociation of the neural correlates involved in pain modulation. Anesthesiology. 2011; 115:844–851. [PubMed: 21934411]

21. Mauderli AP, Vierck CJ Jr. Cannon RL, Rodrigues A, Shen C. Relationships between skin temperature and temporal summation of heat and cold pain. Journal of neurophysiology. 2003; 90:100–109. [PubMed: 12843304]

22. Miller, RD. Miller's Anesthesia. 7th Edition. Churchill & Livingston; 2009. Anatomy of the Peripheral Nerve.

23. Neziri AY, Haesler S, Petersen-Felix S, Muller M, Arendt-Nielsen L, Manresa JB, Andersen OK, Curatolo M. Generalized expansion of nociceptive reflex receptive fields in chronic pain patients. Pain. 2010; 151:798–805. [PubMed: 20926191]

24. Price DD, Dubner R. Mechanisms of first and second pain in the peripheral and central nervous systems. J Invest Dermatol. 1977; 69:167–171. [PubMed: 326991]

25. Price DD, Hu JW, Dubner R, Gracely RH. Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. Pain. 1977; 3:57–68. [PubMed: 876667]

26. Price DD, Staud R, Robinson ME, Mauderli AP, Cannon R, Vierck CJ. Enhanced temporal summation of second pain and its central modulation in fibromyalgia patients. Pain. 2002; 99:49–59. [PubMed: 12237183]

27. Raphael KG, Janal MN, Anathan S, Cook DB, Staud R. Temporal summation of heat pain in temporomandibular disorder patients. Journal of orofacial pain. 2009; 23:54–64. [PubMed: 19264036]

28. Ribeiro-Dasilva MC, Goodin BR, Fillingim RB. Differences in suprathreshold heat pain responses and self-reported sleep quality between patients with temporomandibular joint disorder and healthy controls. Eur J Pain. 2012

29. Robinson ME, Bialosky JE, Bishop MD, Price DD, George SZ. Supra-threshold scaling, temporal summation, and after-sensation: relationships to each other and anxiety/fear. Journal of pain research. 2010; 3:25–32. [PubMed: 21197307]

30. Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, Binder A, Birbaumer N, Birklein F, Botefur IC, Braune S. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. Pain. 2006; 123:231–243. [PubMed: 16697110]

31. Sarlani E, Garrett PH, Grace EG, Greenspan JD. Temporal summation of pain characterizes women but not men with temporomandibular disorders. Journal of orofacial pain. 2007; 21:309–317. [PubMed: 18018992]

32. Sibillie, KT.; Goodin, BR.; Herrera, DG.; Riley, JL., 3rd; Fillingim, RB. Ethnic Differences in Temoral Summation of Thermal Pain; American Pain Society Annual Conference; Austin, Texas. 2011;

33. Staud R, Price DD, Fillingim RB. Advanced continuous-contact heat pulse design for efficient temporal summation of second pain (windup). The journal of pain: official journal of the American Pain Society. 2006; 7:575–582. [PubMed: 16885014]

34. Staud R, Bovee CE, Robinson ME, Price DD. Cutaneous C-fiber pain abnormalities of fibromyalgia patients are specifically related to temporal summation. Pain. 2008; 139:315–323. [PubMed: 18538477]

35. Staud R, Vierck CJ, Cannon RL, Mauderli AP, Price DD. Abnormal sensitization and temporal summation of second pain (wind-up) in patients with fibromyalgia syndrome. Pain. 2001; 91:165–175. [PubMed: 11240089]

36. Valencia C, Fillingim RB, George SZ. Suprathreshold heat pain response is associated with clinical pain intensity for patients with shoulder pain. The journal of pain: official journal of the American Pain Society. 2011; 12:133–140. [PubMed: 20692209]

37. Vierck CJ Jr. Cannon RL, Fry G, Maixner W, Whitsel BL. Characteristics of temporal summation of second pain sensations elicited by brief contact of glabrous skin by a preheated thermode. Journal of neurophysiology. 1997; 78:992–1002. [PubMed: 9307129]

38. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. Journal of strength and conditioning research / National Strength & Conditioning Association. 2005; 19:231–240.

39. Williamson DF, Parker RA, Kendrick JS. The box plot: a simple visual method to interpret data. Annals of internal medicine. 1989; 110:916–921. [PubMed: 2719423]

A

Trial 1
(T1)
40-49C

A: TS$_E$ < 30 → ↑ temperatures to ↑ TS$_E$ in Trials T2 -> T5

B: 30 ≤ TS$_E$ ≤70 → Stop, use 40-49 C as final

C:TS$_E$ >70 → ↓ temperatures to ↓ TS$_E$ in Trials T2 -> T5

B

Scenario A: TS$_E$ < 30

| T1:
40-49 | 15min → | T2:
40-50 | 15min → | T3:
40-51 | 15min → | T4:
41-51 | 15min → | T5: final
42-51 |

Scenario C: TS$_E$ >70

| T1:
40-49 | 15min → | T2:
39-49 | 15min → | T3:
38-49 | 15min → | T4:
38-48 | 15min → | T5: final
38-47 |

Figure 1.
Algorithm to individually optimize TS.
a) Overall Scheme after Trial 1 (T1).
b) Specific algorithms for scenario A and C, where TS is <30 or >70, respectively. T1 = 1st optimization trial, T2 -> T5 = 2nd to 5th optimization trials. Each participant is allowed to have no more than 5 optimization trials.

Day 1	5 optimization trials to determine final Base$_{temp}$ and Peak$_{temp}$ to be used on Day 4 and 11

Day 4 | Trial1 | →30min→ | Trial2 | →15min→ | Trial3 | →5min→ | Trial4 | →2min→ | Trial5

Day 11 | Trial1 | →30min→ | Trial2 | →15min→ | Trial3 | →5min→ | Trial4 | →2min→ | Trial5
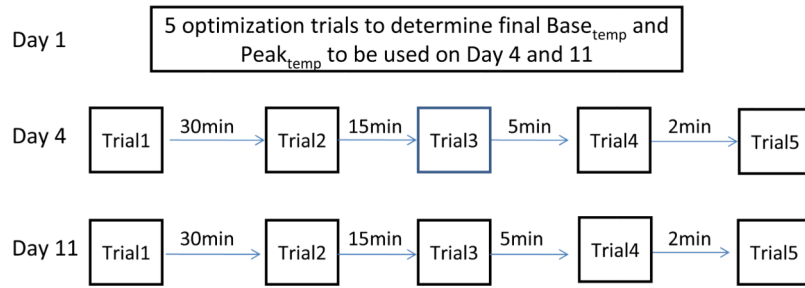
Figure 2.
Timeline to assess the test-retest reliability of thermal TS generated by individually optimized parameters. Base$_{temp}$ refers to the baseline temperature and Peak$_{temp}$ to the stimulus temperature.
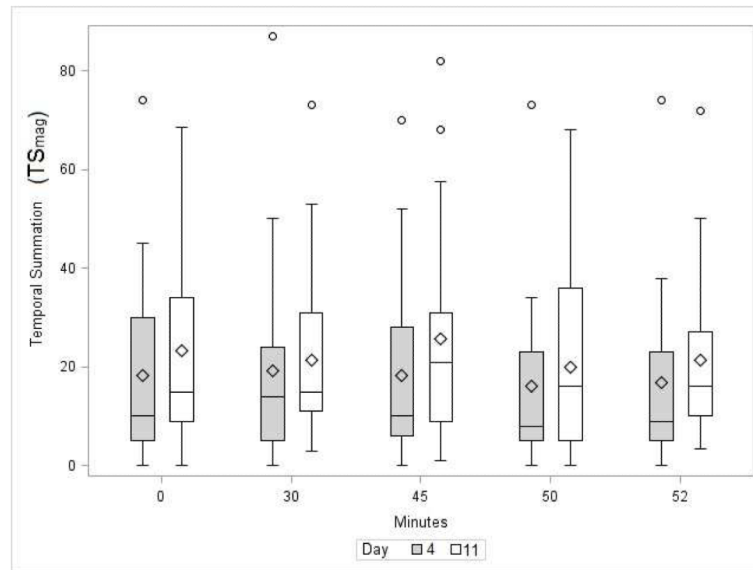
Figure 3.
Box plot of mean and distribution of TS from five trials on Day 4 and five on Day 11. N = 19. ◇ denotes mean. Elongated rectangle denotes 25 to 75 quartiles. Horizontal bar within rectangle denotes median. ○ outside the rectangle denotes outlier. Vertical line outside rectangle denotes range.
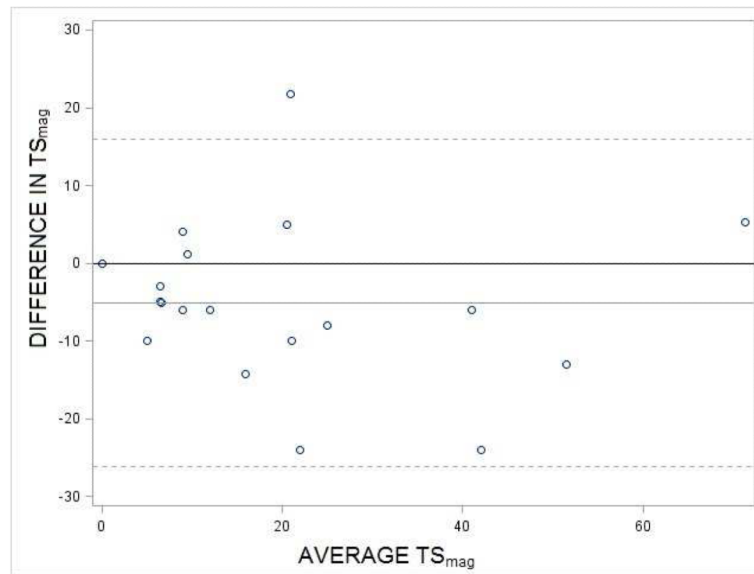
Figure 4.
Bland-Altman plot of TS from the first trials on Day 4 and Day 11. Each circle represents one of the 19 participants. Mean difference in 1st TS of (Day 4 - Day 11) = −5.1 (gray line). The 95% CL is between −0.03 and −10.1. The hashed lines represent 2 standard deviations above and below the mean.

Table 1

Results from the individualized optimization of TS on Day 1 (n=19)

Study ID	Gender	Age	# of trials	Base$_{temp}$	Peak$_{temp}$	Estimated TS	Time to peak (# of pulses)
TRS 01	F	45	5	42	51	10	10
TRS 02	M	21	2	39	49	50	10
TRS 03	M	26	5	40.5	51	60	4
TRS 04	M	27	2	40	50	60	10
TRS 06	F	21	5	42	51	8	2
TRS 07	F	39	2	39	49	60	5
TRS 08	F	25	3	40	49	36	3
TRS 09	F	19	5	39	49	18	10
TRS 011	M	40	5	41	51	23	2
TRS 012	F	18	5	40	49	24	7
TRS 014	F	45	3	40	49	27	8
TRS 015	M	29	3	40	49	64	5
TRS 016	F	18	4	40	49	18	10
TRS 018	M	50	3	40	49	33	4
TRS 019	M	22	4	38	49	15	5
TRS 021	M	43	5	38.5	49	7	2
TRS 025	M	20	4	40	50	17	5
TRS 026	M	19	5	40	47	11	4
TRS 027	F	27	5	40	47	11	10
AVERAGE		**29.2**	**3.9**	**39.9**	**49.3**	**29.1**	**6.1**
STND DEVIATION		**10.5**	**1.1**	**1.0**	**1.1**	**19.5**	**3.0**

Abbreviations: Base$_{temp}$ represents baseline temperature, Peak$_{temp}$ represents stimulus temperature.

Table 2

TS$_E$ from Day 1 and TS$_{mag}$ from Day 4, Day 11. N=19

Subj ID	Day 1 -- TS$_E$	DAY 4 -- TS$_{mag}$							Day 11 -- TS$_{mag}$						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Mean	SD	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Mean	sd
1	10	9	15	8	7	11	9.9	3.2	23	24	27	30	27	26.2	2.8
2	50	11	0	7	4	9	6.2	4.3	7	3	4	1	9	4.7	3.3
3	60	38	23	28	23	5	23.4	12	44	31	28	17	16	27.2	11.4
4	60	30	50	34	28	33	35.1	8.7	54	73	68	43	50	57.6	12.6
6	8	6	11	9	5	9	8.1	2.6	12	18	8	5	14	11.4	5.1
7	60	45	35	52	34	28	38.8	9.6	58	34	36	41	42	42.2	9.4
8	36	0	3	0	0	0	0.6	1.3	10	11	12	9	11	10.5	1.1
9	18	16	27	22	5	19	17.8	8.2	26	17	19	11	11	16.7	6.3
11	23	10	12	0	17	8	9.5	6.2	34	13	58	41	19	32.9	17.8
12	24	23	24	0	8	16	14.2	10.2	18	5	31	18	10	16.4	9.8
14	27	32	14	15	5	5	14.1	10.9	10	15	17	21	19	16.4	4.2
15	64	74	87	70	73	74	75.6	6.6	69	53	82	68	72	68.7	10.4
16	18	5	8	5	5	0	4.6	2.9	8	11	9	16	21	13	5.4
18	33	9	4	12	13	38	15.2	13.2	15	12	21	16	25	17.8	5.2
19	15	21	19	23	32	23	23.6	5	29	21	27	36	34	29.4	5.9
21	7	0	3	7	5	4	3.8	2.6	0	6	1	0	5	2.4	2.9
25	17	4	5	6	19	18	10.5	7.4	9	39	3	0	3	10.9	16.1
26	11	4	17	10	4	8	8.6	5.4	9	11	23	6	4	10.6	7.3
27	11	10	5	37	19	9	16	12.8	9	7	13	0	16	8.9	6.1
AVG	29.1						17.7	7.0						22.3	7.5

Table 3

Correlations between TS magnitudes calculated by 3 different methods (n=19).

	TS_{mag}	TS_{mag1}	TS_{mag2}
TS_{mag}	r = 1	r = 0.85	r = 0.91
		p < .0001	p < .0001
TS_{mag1}	r = 0.85	r = 1	r = 0.72
	p < .0001		p = 0.0007
TS_{mag2}	r = 0.91	r = 0.72	r = 1
	p < .0001	p = 0.0007	

Abbreviations:

r: Pearson Correlation Coefficient

p: Probability > |r| under H0: Rho=0

TS_{mag}: Magnitude of TS calculated by subtracting pain of first pulse from peak pain.

TS_{mag1}: Magnitude of TS calculated by subtracting pain of first pulse from that of last pulse.

TS_{mag2}: Magnitude of TS calculated by taking the slope of the change in pain over the change in time, between the first and the last pulse.