ORIGINAL ARTICLE

Two-Week Test–Retest Stability of the Cold Pressor Task Procedure at two different Temperatures as a Measure of Pain Threshold and Tolerance

Julian Koenig, MA*; Marc N. Jarczok, MSSc†; Robert J Ellis, PhD‡; Claudia Bach, MA*; Julian F. Thayer, PhD§; Thomas K. Hillecke, Dr. Sc. Hum.*

*School of Therapeutic Sciences, SRH University, Heidelberg; †Mannheim Institute of Public Health Social and Preventive Medicine, Mannheim Medical Faculty, Heidelberg University, Mannheim, Germany; ‡Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Masssachusetts; §Department of Psychology, The Ohio State University, Columbus, Ohio, USA

■ **Abstract**

Background: The cold pressor task (CPT) was originally developed as a clinically indicative cardiovascular test, and quantifies vascular response and pulse excitability when a subject's hand is immersed into ice water. Since the test procedure results in a gradually increasing cold pain, the CPT has been widely used as a nociceptive stimulus in experimental studies on adults and children.

Aim: To evaluate the test-retest stability of response patterns using the CPT as a measure of pain threshold and pain tolerance.

Materials and Methods: In the present study, sixty-one undergraduate students received painful stimulation using the CPT either at 4°C or 6°C. Measurements of pain threshold, pain tolerance and pain intensity ratings using the short form of the McGill pain questionnaire (SF-MPQ), were derived. The assessment was repeated twice over an interval of 2 weeks. Test-Retest stability was assessed within a three-layered approach, using ANOVAs, interclass correlation coefficients and standard error of the mean. A Bland-Altman analysis was also performed. Possible predictors of pain threshold and pain tolerance were assessed using random effect panel regression models.

Results: No significant differences emerged as a function of temperature (4°C or 6°C) on pain threshold, pain tolerance, and pain ratings. Environmental variables (room temperature and humidity) show no impact on measures of pain threshold and pain tolerance.

Conclusion: Consistent with previous findings, regression analysis reveals that age is significantly associated with pain tolerance. The CPT procedure shows excellent 2 week test-retest stability to assess pain threshold and pain tolerance within a student population. ■

Key Words: cold pressor task, cold pressor task, test-retest, pain threshold, pain tolerance

Address for correspondence and reprint request to: Julian Koenig, MA, SRH University Heidelberg, Maaßstraße 26, 69123 Heidelberg. E-mail: julian.koenig@fh-heidelberg.de.

Disclosures: The authors declared no potential conflict of interests with respect to the research, authorship, and/or publication of this article. The work on the present article was funded by the "Gesellschaft der Freunde der SRH Hochschule Heidelberg".

Submitted: August 08, 2013; Revision accepted: September 15, 2013
DOI. 10.1111/papr.12142

© 2013 World Institute of Pain, 1530-7085/13/$15.00

INTRODUCTION

The cold pressor task (CPT) was originally developed as a clinically indicative cardiovascular test and quantifies

vascular response and pulse excitability when a subject's hand is immersed into ice water.[1,2] As the test procedure results in a gradually increasing cold pain, the CPT has also been widely used as pain stimuli in experimental studies on adults[3] and children.[4–6] Whereas in the clinical use, the immersion time is standardized (~1 minute), the time of immersion quantified in seconds is the dependent test variable in the application of the CPT in experimental pain research. Two main measures are derived from the test in the latter case: *pain threshold*, determined by the point at which the subject first reports noticeable pain and the upper limit for endurance of noxious stimulation; *pain tolerance*, determined by the elapsed time from immersion into the ice water to the point subjects report that they can no longer endure the stimulation.[7]

The stability of vascular and myocardial response patterns provoked by the CPT as clinically indicative test has already been investigated in several studies.[8–12] Results show that the responses elicited by the CPT are reliable and show little evidence of attenuation over the test–retest interval, indicating that reactivity to the CPT can be reliably assessed over a 2-week interval (Saab et al.[10]). An important source of variation that must be controlled for is water temperature[13] and the speed and turbulence of water circulation[14], which causes variations in convection heat transfer as a potential source of discrepancies.[15,16]

However, the test–retest stability of measurements of pain threshold and pain tolerance associated with the CPT procedure, although frequently assumed,[3,16] has not been systematically demonstrated to date. This study aimed to investigate the test–retest stability of measures of pain threshold and pain tolerance using the CPT at 2 different temperatures (4°C and 6°C) as a pain-inducing stimulus in a 2-week, repeated-measure design using a healthy sample of young adults.

METHODS

All experiments were conducted in accordance with the Declaration of Helsinki and the Ethical Guidelines for Pain Research in Humans by the Committee on Ethical Issues of the International Association for the Study of Pain (IASP). The study protocol was approved by the university's institutional review board.

General Procedures

Healthy undergraduate students were recruited at the SRH University Heidelberg from September 2012 through January 2013. Upon arrival to the laboratory, participants provided written informed consent and completed several questionnaires. Self-rated health (SRH) was measured using the question "How do you rate your current health status?" on a 0 ("very bad") to 6 ("excellent") scale. Only subjects indicating a SRH ≥3 ("fair") were included in the trial. After inclusion in the study, subjects were randomly assigned by permuted-block randomization (1:1) to the 4°C or 6°C group. The randomization sequence was generated by the first author *a priori* for up to 80 participants (40 per temperature group). The final sample of participants was healthy undergraduate students consisting of 46 females and 15 males. The age range was 19 to 45 years; mean 23.71 years (SD = 4.15). The difference of 2°C between the groups was based on a previous finding[3], which reported a significant difference in pain experience.

Subjects were blind to their randomized group. Each participant was given a day of the week (Monday through Friday) and a time (between 9AM and 6PM), 2 weeks apart when both assessments would be made. Date and time of measurements were recorded by a protocol. Socio-demographic variables were assessed using a self-developed questionnaire. The fourth author performed all assessments. Data management and analysis were performed by the first and second author: They were otherwise not involved in participant contact. All participants received class credits or an allowance of 20€ for the completion of the study.

Cold Pressor Task

Cold pressor pain sensitivity was assessed by immersing the nondominant hand up to the wrist in a 15.5 × 19.5 × 15.75 inches acrylic glass (2 cm thick) tank with circulating water (3 floor pumps Conrad Electronic GmbH AP-333, water flow: each 200 l/h) to prevent local warming. Water temperature was controlled as specified constantly with a chilling device (Resun CL 250) and water pump (Conrad Electronic GmbH Item no. 55 16 73, 1400 l/h) and measured with 3 digital thermometers (Electronics Tomorrow Ltd. 2120, Hong Kong, Hong Kong) at different spots (chiller inflow, chiller outflow, tank back top). Minimum and maximum water temperature was recorded in °C for every thermometer and session. Subjects were told to keep their hand open rather than closed in a fist while it was in the water. Before the immersion the subject was told to keep the hand in the water until cold

pressor pain turned intolerable, with a cutoff time of 4 minutes. The latencies to the first pain sensation (*pain threshold*) and to the intolerable pain (*pain tolerance*) were measured with a stopwatch in seconds. The CPT was repeated 2 times, with a 2-week interval in both groups. The ambient temperature and humidity was recorded during both sessions.

Pain Ratings

The short form of the McGill Pain Questionnaire (SF-MPQ[17]) was administered after the cold pressor procedure. The SF-MPQ allows quantitative, multidimensional pain ratings to be obtained in a brief period of time and was derived from the McGill Pain Questionnaire (MPQ[18]). The SF-MPQ consists of 15 descriptors (11 sensory, 4 affective) that are rated on a 4-point intensity scale from 0 ("none") to 3 ("severe"). Three pain scores (sensory, affective, and total descriptors) are derived from the sum of the intensity rank values of the words chosen for descriptors. Pain intensity on hand removal was assessed on a 10 cm visual analog scale (VAS) from 0 to 10, and the overall pain experience is assessed by one descriptor ("no pain", "mild", "discomforting", "distressing", "horrible", or "excruciating"). We used the German version of the SF-MPQ.[19]

Statistical Methods

Statistical comparisons on socio-demographic variables between groups and environmental and test variables between the assessments were made using a one-way analysis of variance for continuous variables and chi-squared test (χ^2-test) for categorical variables (2-sided significance) to ensure equivalent baseline status. A series of repeated-measure ANOVAs and MANOVAs to examine group differences in pain threshold, pain tolerance, and pain rating (SF-MPQ) over time (Group × Time interactions) was performed. The distribution of all continuous variables was determined, and transformations were used as necessary to meet the assumptions of modeling.[20]

Test–retest stability for pain threshold, pain tolerance, and pain intensity was assessed via a 3-layered approach to assess reliability as recommended elsewhere.[21] First, a repeated-measure ANOVA (2-way) was performed as described to examine systematic error. Second, Pearson product-moment correlations were expressed by the interclass correlation coefficient (ICC). Third, standard error of the mean (SEM) of the first and second assessments were determined. The ICC (rho/r) ranges between 0 and 1, where ≥ 0.75 is considered excellent reliability.[21] All tests were deemed significant at the threshold $P < 0.05$. Additionally, differences between both the 2 assessments were presented graphically for each variable using Bland–Altman plots. Possible predictors of pain threshold and pain tolerance were assessed using random effects panel regression models. Data management and analysis were performed using Stata (12.1 MP, College Station, TX, USA) and SPSS (20, IBM, Chicago, IL, USA).

RESULTS

Subjects

Self-rated health was rated as "good/very good" or "excellent" by 78% of participants. Subjects had an average height of 170.75 cm (SD = 8.76 cm) and an average body weight of 67.13 kg (SD = 14.09 kg). Their body mass index (BMI) range was 17.2 to 41.6; mean 22.93 (SD = 4.12). Thirty-four participants (28 female) were allocated to the 4°C group, and 25 (16 female) to the 6°C group. Differences in the above demographics were not significant between the groups. Characteristics are given in Table 1.

Pain Threshold, Pain Tolerance

ANOVA results on average pain threshold and pain tolerance (with group and time as factors) are presented

Table 1. Socio-demographic Characteristics

	4°C Group	6°C Group	*P*	Total Sample
n (f/m)	34 (28/6)	25 (16/9)	0.110	59 (44/15)
age, M (SD), years	22.85 (2.46)	24.88 (5.55)	0.063	23.71 (4.15)
height, M (SD), cm	169.68 (7.49)	172.20 (10.23)	0.278	170.75 (8.76)
weight, M (SD), kg	66.97 (14.88)	67.34 (13.24)	0.921	67.13 (14.09)
BMI, M (SD), kg/m^2	23.18 (4.64)	22.59 (3.33)	0.591	22.93 (4.12)
dominant hand, *n/n* (%/%), left/right	2/32 (5.9/94.1)	2/23 (8.0/92.0)	0.749	4/55 (6.8/93.2)

n (f/m) = number of subjects, females/males; BMI (kg/m^2) = body mass index, *P* differences between groups assessed by ANOVA or chi-squared test.

Table 2. Pain-related Measures

	4°C Group			6°C Group		
	1st Assessment	2nd Assessment	*P*	1st Assessment	2nd Assessment	*P*
CPT performance, mean (SD)						
Pain threshold, seconds	18.82 (13.69)	17.00 (9.9)	0.428*	16.78 (15.44)	15.15 (14.31)	0.647*
Pain tolerance, seconds	65.19 (74.75)	63.19 (78.24)	0.530*	43.24 (59.99)	56.01 (76.10)	0.284*
Total immersion time, seconds	84.01 (75.92)	80.19 (80.35)	0.303*	60.02 (66.16)	71.16 (80.03)	0.238*
SF-MPQ, mean (SD)						
Sensory pain index, 0 to 33	2.41 (0.43)	2.32 (0.46)	0.242*	2.55 (0.43)	2.52 (0.38)	0.676*
Affective pain index, 0 to 12	1.68 (2.40)	2.18 (2.59)	0.100	1.76 (1.94)	1.72 (2.32)	0.911
Total pain index, 0 to 45	2.52 (0.46)	2.48 (0.51)	0.572*	2.66 (0.47)	2.62 (0.43)	0.637*
Pain experience intensity, 0 to 5	0.86 (0.26)	0.88 (0.29)	0.678	0.92 (0.23)	0.85 (0.31)	0.178
Pain intensity on hand removal, 0 to 10	5.14 (2.08)	5.28 (2.03)	0.830*	6.08 (2.11)	6.14 (1.81)	0.753*

P differences between assessments assessed by one-way repeated-measurement ANOVA.
*log-transformed for ANOVA.

in Table 2. Pain threshold ($F_{1,116} = 2.93$; $P = 0.089$), pain tolerance ($F_{1,116} = 3.88$; $P = 0.051$), and total immersion time ($F_{1,116} = 3.76$; $P = 0.055$) showed no statistically significant differences between groups. Furthermore, repeated-measure ANOVAs revealed no statistical significant differences for (1) pain threshold in the 4°C ($F_{1,33} = 0.64$; $P = 0.428$) and 6°C group ($F_{1,24} = 0.21$; $P = 0.647$), (2) pain tolerance in the 4°C ($F_{1,33} = 0.40$; $P = 0.530$) and 6°C group ($F_{1,24} = 1.20$; $P = 0.284$), and (3) total immersion time in the 4°C ($F_{1,33} = 1.09$; $P = 0.303$) and the 6°C group ($F_{1,24} = 1.46$; $P = 0.238$) between the 2 assessments. MANOVAs that were performed to examine Group × Assessment interactions revealed no statistical significant differences for pain threshold ($F_{1,118} = 2.93$; $P = 0.090$), pain tolerance ($F_{1,118} = 3.87$; $P = 0.515$), or total immersion time ($F_{1,118} = 0.42$; $P = 0.518$). Figure 1 shows the survival graphs for total immersion time by group and assessment.

Pain Ratings

Results from the SF-MPQ pain intensity rating and sensory, affective, and total descriptors are presented in Table 2. Neither VAS ratings of pain intensity on hand removal ($F_{1,117} = 0.06$; $P = 0.806$), pain experience intensity($F_{1,115} = 5.66$; $P = 0.019$) nor sensory ($F_{1,117} = 4.53$; $P = 0.036$), affective ($F_{1,117} = 0.18$; $P = 0.672$), and total pain index ($F_{1,117} = 2.56$; $P = 0.113$) showed significant differences between the groups. In neither in the 4°C group ($F_{1,32} = 0.05$, $P = 0.830$) nor in the 6°C group ($F_{1,22} = 0.10$; $P = 0.753$), ratings of pain intensity showed significant differences between the 2 assessments, and similar null differences were found for the sensory pain index, the affective pain index, the total pain index, and the pain experience intensity (Table 2). Furthermore, MANOVA (Group × Assessment) ([Pain intensity on hand removal: ($F_{1,117} = 0.79$; $P = 0.376$)]) ([Pain experience intensity: ($F_{1,115} = 0.01$; $P = 0.911$)]), ([Sensory pain index: ($F_{1,117} = 0.12$; $P = 0.731$)]) ([Affective pain index: ($F_{1,117} = 0.39$; $P = 0.536$)]) ([Total pain index: ($F_{1,117} = 0.00$; $P = 0.962$)]) revealed, no scale or index of the SF-MPQ showing any statistical significant differences.

Environment and Test Variables

All environmental and test variables are presented in Table 3. Room temperature and humidity were recorded at the beginning and the end of each assessment. None of the environmental variables showed significant differences between the groups. Only room temperature at the beginning of the assessments within the 6°C group showed a significant difference when comparing the first and second assessment ($F_{1,24} = 5.62$; $P = 0.026$). Humidity showed significant differences within the 4°C group at the beginning ($F_{1,32} = 0.38$; $P = 0.003$) and the end of assessment ($F_{1,32} = 8.23$; $P = 0.007$), and at the beginning of the assessment within the 6°C group ($F_{1,24} = 4.61$; $P = 0.042$), comparing the first and second assessment. Cold pressor task water temperature was controlled by 3 digital thermometers placed at different spots within the water tank. Maximum and minimum water temperatures were recorded for each subject and assessment. Mean values were calculated based on these recordings. Significant differences between the first and second assessment occured within the 6°C group for the mean tank bottom ($F_{1,24} = 6.05$; $P = 0.022$) and mean tank outflow ($F_{1,24} = 6.06$; $P = 0.021$) water temperature, the max-

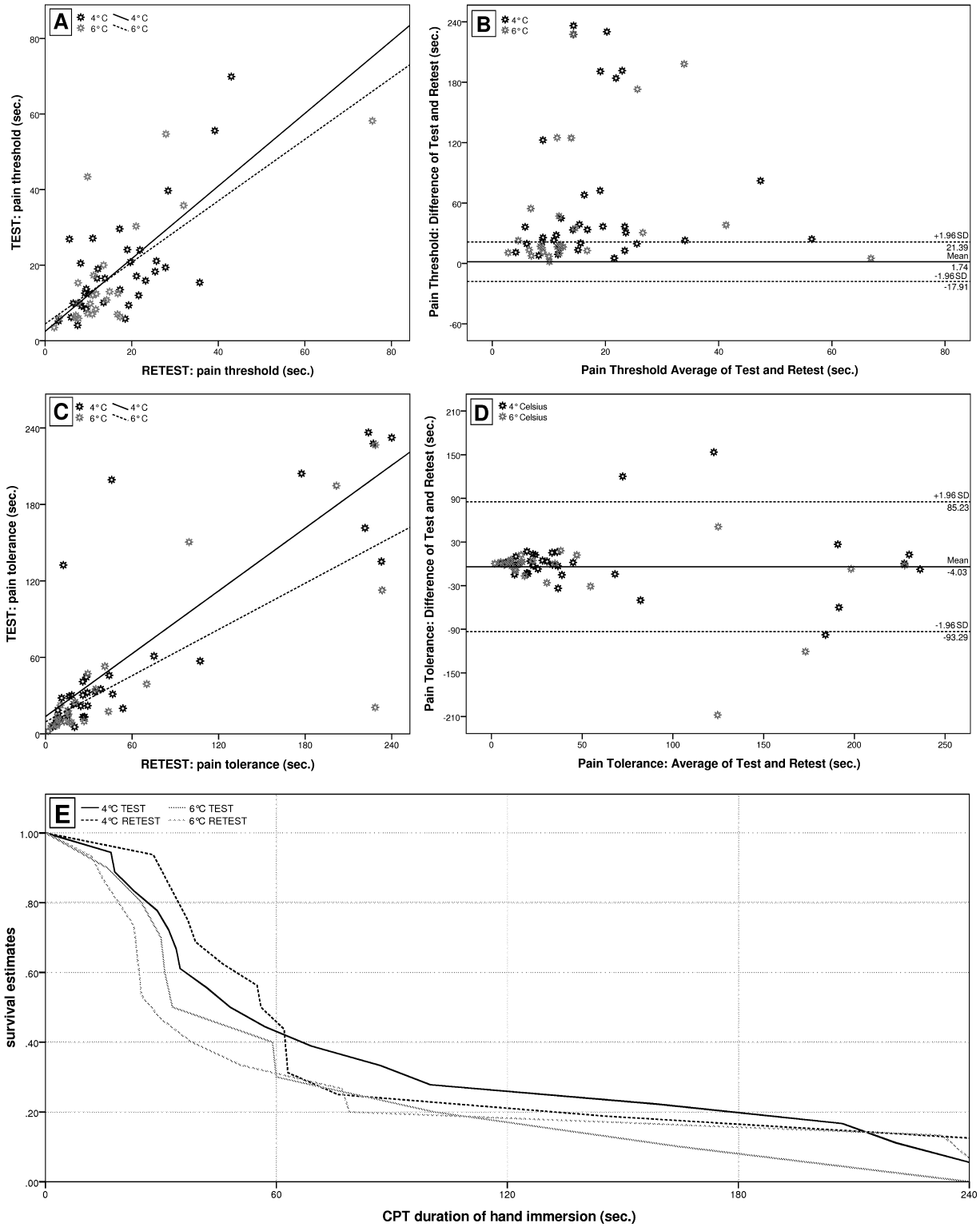

Figure 1. Visual inspection of test–retest stability. (A) Test–retest pain threshold in seconds, by group (4°C and 6°C), 4°C group: $R^2 = 0.48$; 6°C group: $R^2 = 0.57$; (B) Bland–Altman plot on pain threshold; (C) Test–retest pain tolerance in seconds, by group (4°C and 6°C), 4°C group: $R^2 = 0.72$; 6°C group: $R^2 = 0.59$; (D) Bland–Altman plot on pain tolerance; (E) Survival estimates on total CPT performance (duration of hand immersion in seconds) by group and time of assessment; B/D represent Bland–Altman plots on pain threshold and pain tolerance in seconds by group. The average of test and retest and the differences between test and retest, as well as their mean and a $\pm$ 1.96 standard deviation for the total sample are presented to show the agreement between the 2 different assessments.

Table 3. Environmental and Test Variables

	4°C Group			6°C Group		
	1st Assessment	2nd Assessment	*P*	1st Assessment	2nd Assessment	*P*
Room temperature, °C, mean (SD)						
At the beginning of assessment	21.64 (1.45)	21.67 (1.55)	0.878	22.30 (2.35)	21.36 (1.03)	**0.026**
At the end of assessment	21.69 (1.42)	22.24 (3.51)	0.349	22.32 (2.36)	21.56 (0.94)	0.077
Humidity,%, mean (SD)						
At the beginning of assessment	41.29 (4.70)	37.45 (5.65)	**0.003**	43.59 (6.02)	40.96 (5.37)	**0.042**
At the end of assessment	41.50 (4.86)	38.41 (4.96)	**0.007**	44.00 (6.07)	41.61 (5.41)	0.063
Water temperature, °C, mean (SD)						
Tank bottom	4.24 (0.23)	4.28 (0.25)	0.534	6.07 (0.33)	5.76 (0.64)	**0.022**
Tank inflow	4.19 (0.26)	4.31 (0.26)	0.183	6.10 (0.30)	5.96 (0.63)	0.308
Tank outflow	4.38 (0.30)	4.40 (0.27)	0.565	6.15 (0.33)	5.86 (0.59)	**0.021**
Maximum water temperature, °C, mean (SD)						
Tank bottom	4.34 (0.24)	4.35 (0.26)	0.923	6.20 (0.32)	5.90 (0.62)	**0.017**
Tank inflow	4.30 (0.30)	4.41 (0.23)	0.289	6.24 (0.31)	6.10 (0.59)	0.272
Tank outflow	4.53 (0.34)	4.54 (0.28)	0.795	6.32 (0.35)	5.99 (0.62)	**0.011**
Minimum water temperature,°C, mean (SD)						
Tank bottom	4.14 (0.26)	4.21 (0.26)	0.277	5.94 (0.40)	5.62 (0.68)	**0.033**
Tank inflow	4.07 (0.26)	4.21 (0.26)	0.139	5.95 (0.33)	5.82 (0.69)	0.378
Tank outflow	4.22 (0.32)	4.26 (0.28)	0.415	5.98 (0.38)	5.73 (0.59)	0.059

P differences between assessments assessed by one-way repeated-measurement ANOVA.
Bold values indicate a statistical significant difference.

imum tank outflow ($F_{1,24} = 6.59$; $P = 0.017$) and maximum tank bottom ($F_{1,24} = 7.50$; $P = 0.011$) and the minimum tank bottom ($F_{1,24} = 5.14$; $P = 0.033$) water temperature likewise. MANOVA (Group × Assessment) ($F_{1,117} = 2.47$; $P = 0.120$) revealed, no statistical significant differences on any of the environmental variables.

Test–Retest Reliability

Pearson product-moment correlations and standard error of the mean (SEM) of the first and second assessment for both groups were calculated as presented in Table 4. Figure 1 presents several graphical analyses. Figure 1A shows test (*y*-axis) and retest (*x*-axis) pain threshold (Figure 1A) and pain tolerance (Figure 1B) (non-log-transformed data), separately for each group, with associated Bland–Altman plots (Figure 1C,D). Figure 1E presents a survival analysis of total performance time by group and assessment. Coefficients of determination (R^2, simple linear regression) for the graphs by group are given within the figure legends.

Predictors of Pain Threshold and Pain Tolerance

Possible predictors of pain threshold and pain tolerance were assessed using random effect panel regression models. Hausman tests indicated no systematic difference between fixed effect and random effect models (pain threshold χ^2 1.21 $P = 0.55$; pain tolerance χ^2 0.29 $P = 0.87$), thus allowing for inferences on the infinite population. The unadjusted population average of pain threshold was 17 s and 57.9 s for pain tolerance.

Table 4. ICC and SEM

	4°C Group		6°C Group	
	ICC	CORR	ICC	CORR
Pain threshold, seconds	0.79	0.62***	0.86	0.73***
Pain tolerance, seconds	0.92	0.81***	0.85	0.86***
Total immersion time, seconds	0.92	0.83***	0.87	0.85***
SEM	1st Assessment	2nd Assessment	1st Assessment	2nd Assessment
Pain threshold, seconds	2.35	1.70	3.09	2.86
Pain tolerance, seconds	12.82	13.42	12.19	15.53
Total immersion time, seconds	13.02	13.78	13.23	16.01

ICC, intraclass correlations for one-way random effects model on absolute agreement; SEM standard error of the mean; CORR, Pearson correlation.
***$P < 0.001$.

Adjustments for experimental condition, gender, age, BMI, or environmental condition (ie, room temperature, average humidity, and average water temperature) did not contribute significantly to the explanation of pain threshold, but age did contribute significantly in the pain tolerance models (Table 5). All else being equal, pain tolerance diminishes on average by 2.4 seconds per age year increase in the final model.

DISCUSSION

The reproducibility of the conscious experience and psychophysical assessments of pain are critical factors in pain research.[22] This is the first study systematically investigating test–retest reliability of the widely used CPT as a measure of pain tolerance and threshold. Sixty-one students received nociceptive stimulation by the CPT with 1 of 2 water temperatures (4°C or 6°C). The assessment was repeated after 2 weeks. Pain threshold, pain tolerance, and ratings of pain intensity were derived for each assessment. We controlled for socio-demographic (eg, gender, age) and environmental variables (eg, room temperature, humidity). Despite a strong design, with blinding of the participants, the study has several limitations.

Participants consistent of students and were predominantly female. As gender differences may have substantial contributions to measures of experimentally induced pain tolerance,[3,23,24] this must be taken into account considering the results.[39] However, gender analysis revealed no statistical significant differences within our study (Table 5).

Due to the method of randomization, groups were not balanced according to the number of subjects per group. Other studies that addressed test–retest reliability of different methods of pain stimulation (eg, pressure pain) in comparable subjects report way smaller total sample sizes,[25] then we do per group. Furthermore, a rational for the sample size is given by the findings of Jensen et al.[26] that demonstrated that intra-individual variation of pain threshold in induced pressure pain can be estimated with 80% power at the 0.05 significance level in groups of 10 subjects. However, differences between the groups were not significant, revealing equivalent baseline status. We were not able to blind study personnel, as cooling of the water needed to be adjusted and recorded for each assessment. Participants were blinded to the water temperature. We did not obtain ratings on the water temperature expectation by the participants, what might have been of interest. Furthermore, we did not record outdoor temperature that might be of interest, contributing to possible differences in thermal stimulation.

Three different thermometers at different spots within the tank controlled water temperature of the CPT at the beginning and the end of the task. However, instead of recording maximum and minimum temperature for each assessment, longitudinal records of the water temperature during the time of hand immersion might lead to more appropriate data. As sex of the experimenter has been shown to influence experimental pain thresholds,[27] and a female experimenter (fourth author) took all our assessments, a possible but systematic bias may be present in our data.

Table 5. Possible Predictors of Pain Threshold (seconds) and Pain Tolerance (seconds) in Random Effect Panel Regression Models

	Pain Threshold				Pain Tolerance			
Model	M0	M1	M2	M3	M0	M1	M2	M3
Group		−1.948	−4.01	−6.872		−14.362	−11.492	8.205
Gender			6.583	6.915			17.355	17.339
BMI (KG/m^2)			−0.280	−0.309			−0.108	−0.332
Age (Years)			0.348	0.395			−3.011**	−2.458*
Water temp. (°C)				1.92				−11.342
Room temp. (°C)				−1.054				−2.983
Humidity (%)				−0.080				−0.676
Constant	17.086	17.912	17.119	20.759	57.903	63.988	58.360	87.571
r2 within	0		0.001	0.033	0		0	0.023
r2 between	0	0.006	0.066	0.088	0	0.011	0.05	0.082
r2 overall	0	0.005	0.057	0.080	0	0.01	0.045	0.075
sigma_u	11.102	11.176	11.021	11.189	65.871	66.121	66.527	66.501
sigma_e	7.134	7.134	7.314	7.38	32.056	32.056	32.267	33.791

Group (4° vs. 6°C); gender (male vs. female); BMI: body mass index; water temp. (mean tank bottom temperature), room temp. (room temperature at beginning of assessment), and humidity (humidity at beginning of assessment).
Legend:*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Our findings on average pain threshold and pain tolerance in relation to the water temperature reveal some interesting questions compared with the work by other authors. Mitchell et al.[3] that used the CPT with 4 temperature conditions (1°C, 3°C, 5°C, and 7°C) found that tolerance time increases as temperature increases. Within our study, mean pain threshold and pain tolerance as well as total immersion time showed no significant differences between the 4°C and the 6°C group. Our findings do not support earlier findings,[3,14] which show variations in temperature as small as 2°C when using the CPT might result in significantly different pain experience. Again, variations in the construction and methods of CPT[3] might be responsible for these inconsistent results. Further studies are needed, addressing these issues of the CPT use for nociceptive stimulation. However, the main focus of the study was to assess test–retest reliability of the CPT with 2 different water temperatures.

Previous studies report good reliability for the assessment of pain thresholds, using pressure pain stimuli inducted by mechanical algometers in healthy[28–31] and clinical[32,33] subjects. Although the measurements may be significantly lower with repeated measures over a short period time[34], pressure pain thresholds were reported highly consistent over 4[34] and 30 days.[35] The reliability of pain threshold and pain tolerance to thermal pain has only been investigated in a few studies. One study found the assessment of threshold and tolerance to pain induced by ice cubes,[36] over assessments separated by 5 days in healthy subjects, reliable. Another study, using ice cubes in sealed plastic satchels that were held against the wrist of the participants (young adults), reports acceptable reliability of the present measures for within-session repeat assessments.[25] Another study[37] investigated test–retest reliability of cold pain threshold and heat pain threshold with a 48-hour interval between 2 assessments. Thermal pain was induced by a thermode, in 13 young participants with chronic nonspecific low back. The authors report an ICC for the cold pain threshold of 0.89, when pain stimulation is applied to the local lumbar area. Another group of studies[13] investigating the reproducibility of quantitative sensory testing (QST) and addressing the reliability of cold pain threshold report good reliability. A recent study investigating the striatal μ-opioid receptor availability showed no significant differences in CPT pain threshold repeated measures of both hands with a 5-minute interval.[38] However, our study is the first to report test–retest stability for the CPT procedure with a longer interval between assessments. Pain threshold assessments in the 4°C (ICC = 0.79) and the 6°C group (ICC = 0.86) show both excellent test–retest reliability (ICC >.75) over assessments separated by 2 weeks, comparable the findings on pressure pain stimuli.[28–33] Pain tolerance assessments in both groups also report excellent reliability. However, results indicate a slightly better stability for the CPT procedure used with a water temperature of 6°C.

Consistent with previous findings[39,40] our regression analysis (Table 5) reveals that pain threshold varies by age. No other environmental or socio-demographic variable showed an impact on measures of pain threshold and pain tolerance. However, gender differences did not show significance in the present analysis and this might be due to the small sample size and the given disparity in group sizes. As volunteers were recruited at the SRH University in Heidelberg, participants were predominantly female, reflecting the majority of female students on campus. While the analysis did show significance when controlled for age but not for gender, age differences in male and female participants need to be taken into account. In group 1 (4°C), analysis of variance revealed statistical significant differences ($F_{1,33} = 6.44$; $P = 0.016$) on mean age between male ($n = 6$, 25.00 years, SD = 2.39) and female ($n = 28$, 22.39 years, SD = 2.39) subjects. In group 2 (6°C), no significant differences ($F_{1,24} = 0.83$; $P = 0.775$) on mean age between male ($n = 9$, 24.44 years, SD = 2.96) and female ($n = 16$, 25.13 years, SD = 6.66) subjects were present. As these differences might contribute to the results of the present regression analysis, future studies should address gender and age interactions by carefully balancing group allocation. Weighted sample sizes of at least 10 male and 10 female participants per group are needed do address such specific questions that are beyond the scope of the present analysis as the main aim of the study was to address the test–retest reliability of the CPT procedure.

The results demonstrate excellent 2-week test–retest reliability for measures of pain threshold and pain tolerance using the CPT as pain stimulus in healthy subjects. However, as the CPT produces a gradually increasing cold pain, other methods, using multiple brief stimuli[22] might be more appropriate for experimental use. Still the CPT procedure has several advantages for it seems to be a cost-efficient, safe, ethical acceptable, and reliable measure of pain threshold and pain tolerance. However, further studies need to address different test setups, methods of cooling, and environmental influ-

ences that may contribute to variations of the measures derived using the CPT for nociceptive stimulation.

REFERENCES

1. Hines E, Brown G. A standard stimulus for measuring vasomotor reactions: its application in the study of hypertension. *Proceedings Staff Meetings of the Mayo Clinic*. 1932;7:332–335.
2. Mourot L, Bouhaddi M, Regnard J. Effects of the cold pressor test on cardiac autonomic control in normal subjects. *Physiol Res*. 2009;58:83–91.
3. Mitchell LA, MacDonald RA, Brodie EE. Temperature and the cold pressor test. *J Pain*. 2004;5:233–237.
4. Birnie KA, Noel M, Chambers CT, von Baeyer CL, Fernandez CV. The cold pressor task: is it an ethically acceptable pain research method in children? *J Pediatr Psychol*. 2011;36:1071–1081.
5. Birnie KA, Petter M, Boerner KE, Noel M, Chambers CT. Contemporary use of the cold pressor task in pediatric pain research: a systematic review of methods. *J Pain*. 2012;13:817–826.
6. von Baeyer CL, Piira T, Chambers CT, Trapanotto M, Zeltzer LK. Guidelines for the cold pressor task as an experimental pain stimulus for use with children. *J Pain*. 2005;6:218–227.
7. Edens JL, Gil KM. Experimental induction of pain: utility in the study of clinical pain. *Behav Ther*. 1995;26:197–216.
8. Evans JA, Rubitsky HJ, Bartels CC, Bartels EC. Re-evaluation of the reliability of pharmacologic and cold pressor studies in hypertension and pheochromocytoma. *Am J Med*. 1951;11:448–467.
9. Durel LA, Kus LA, Anderson NB, et al. Patterns and stability of cardiovascular responses to variations of the cold pressor test. *Psychophysiology*. 1993;30:39–46.
10. Saab PG, Llabre MM, Hurwitz BE, et al. The cold pressor test: vascular and myocardial response patterns and their stability. *Psychophysiology*. 1993;30:366–373.
11. Fasano ML, Sand T, Brubakk AO, Kruszewski P, Bordini C, Sjaastad O. Reproducibility of the cold pressor test: studies in normal subjects. *Clin Auton Res*. 1996;6:249–253.
12. Wirch JL, Wolfe LA, Weissgerber TL, Davies GA. Cold pressor test protocol to evaluate cardiac autonomic function. *Appl Physiol Nutr Metab*. 2006;31:235–243.
13. Moloney NA, Hall TM, Doody CM. Reliability of thermal quantitative sensory testing: a systematic review. *J Rehabil Res Dev*. 2012;49:191–207.
14. Hirsch MS, Liebert RM. The physical and psychological experience of pain: the effects of labeling and cold pressor temperature on three pain measures in college women. *Pain*. 1998;77:41–48.
15. von Baeyer CL, Torvi D, Hemingson H, Beriault D. Commentary: water circulation and turbulence in the cold pressor task: Unexplored sources of variance among experimental pain laboratories. *Pediatric Pain Letter*. 2011;13:13–16.
16. Modir JG, Wallace MS. Human experimental pain models 2: the cold pressor model. *Methods Mol Biol*. 2010;617:165–168.
17. Melzack R. The short-form McGill Pain Questionnaire. *Pain*. 1987;30:191–197.
18. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain*. 1975;1:277–299.
19. Tal A. Assessment: short-Form McGill Pain Questionnaire. *Physiopraxis*. 2008;6:38–39.
20. Tukey JW. *Exploratory Data Analysis*. 1 ed. Massachusetts: Addison-Wesley; 1977.
21. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231–240.
22. Rosier EM, Iadarola MJ, Coghill RC. Reproducibility of pain measurement and pain perception. *Pain*. 2002;98:205–216.
23. Fillingim RB, Maixner W. Gender differences in the responses to noxious stimuli. *Pain Forum*. 1995;4:209–221.
24. Riley JL 3rd, Robinson ME, Wise EA, Myers CD, Fillingim RB. Sex differences in the perception of noxious experimental stimuli: a meta-analysis. *Pain*. 1998;74:181–187.
25. Cathcart S, Pritchard D. Reliability of pain threshold measurement in young adults. *J Headache Pain*. 2006;7:21–26.
26. Jensen R, Rasmussen BK, Pedersen B, Lous I, Olesen J. Cephalic muscle tenderness and pressure pain threshold in a general population. *Pain*. 1992;48:197–203.
27. Gijsbers K, Nicholson F. Experimental pain thresholds influenced by sex of experimenter. *Percept Mot Skills*. 2005;101:803–807.
28. Reeves JL, Jaeger B, Graff-Radford SB. Reliability of the pressure algometer as a measure of myofascial trigger point sensitivity. *Pain*. 1986;24:313–321.
29. Nussbaum EL, Downes L. Reliability of clinical pressure-pain algometric measurements obtained on consecutive days. *Phys Ther*. 1998;78:160–169.
30. Potter L, McCarthy C, Oldham J. Algometer reliability in measuring pain pressure threshold over normal spinal muscles to allow quantification of anti-nociceptive treatment effects. *Int J Osteopath Med*. 2006;9:113–119.
31. Lacourt TE, Houtveen JH, van Doornen LJP. Experimental pressure-pain assessments: test–retest reliability, convergence and dimensionality. *Scand J Pain*. 2012;3:31–37.
32. Sand T, Zwart JA, Helde G, Bovim G. The reproducibility of cephalic pain pressure thresholds in control subjects and headache patients. *Cephalalgia*. 1997;17:748–755.
33. Park G, Kim CW, Park SB, Kim MJ, Jang SH. Reliability and usefulness of the pressure pain threshold measurement in patients with myofascial pain. *Ann Rehabil Med*. 2011;35:412–417.
34. Jones DH, Kilgour RD, Comtois AS. Test-retest reliability of pressure pain threshold measurements of the

upper limb and torso in young healthy women. *J Pain*. 2007;8:650–656.

35. Persson AL, Brogardh C, Sjolund BH. Tender or not tender: test-retest repeatability of pressure pain thresholds in the trapezius and deltoid muscles of healthy women. *J Rehabil Med*. 2004;36:17–27.

36. Marlowe NI. Pain sensitivity and headache: an examination of the central theory. *J Psychosom Res*. 1992;36:17–24.

37. Paungmali A, Sitilertpisan P, Taneyhill K, Pirunsan U, Uthaikhup S. Intrarater reliability of pain intensity, tissue blood flow, thermal pain threshold, pressure pain threshold and lumbo-pelvic stability tests in subjects with low back pain. *Asian J Sports Med*. 2012;3:8–14.

38. Hagelberg N, Aalto S, Tuominen L, et al. Striatal mu-opioid receptor availability predicts cold pressor pain threshold in healthy human subjects. *Neurosci Lett*. 2012;521:11–14.

39. Woodrow KM, Friedman GD, Siegelaub AB, Collen MF. Pain tolerance: differences according to age, sex and race. *Psychosom Med*. 1972;34:548–556.

40. Jensen K, Andersen HO, Olesen J, Lindblom U. Pressure-pain threshold in human temporal region. Evaluation of a new pressure algometer. *Pain*. 1986;25:313–323.