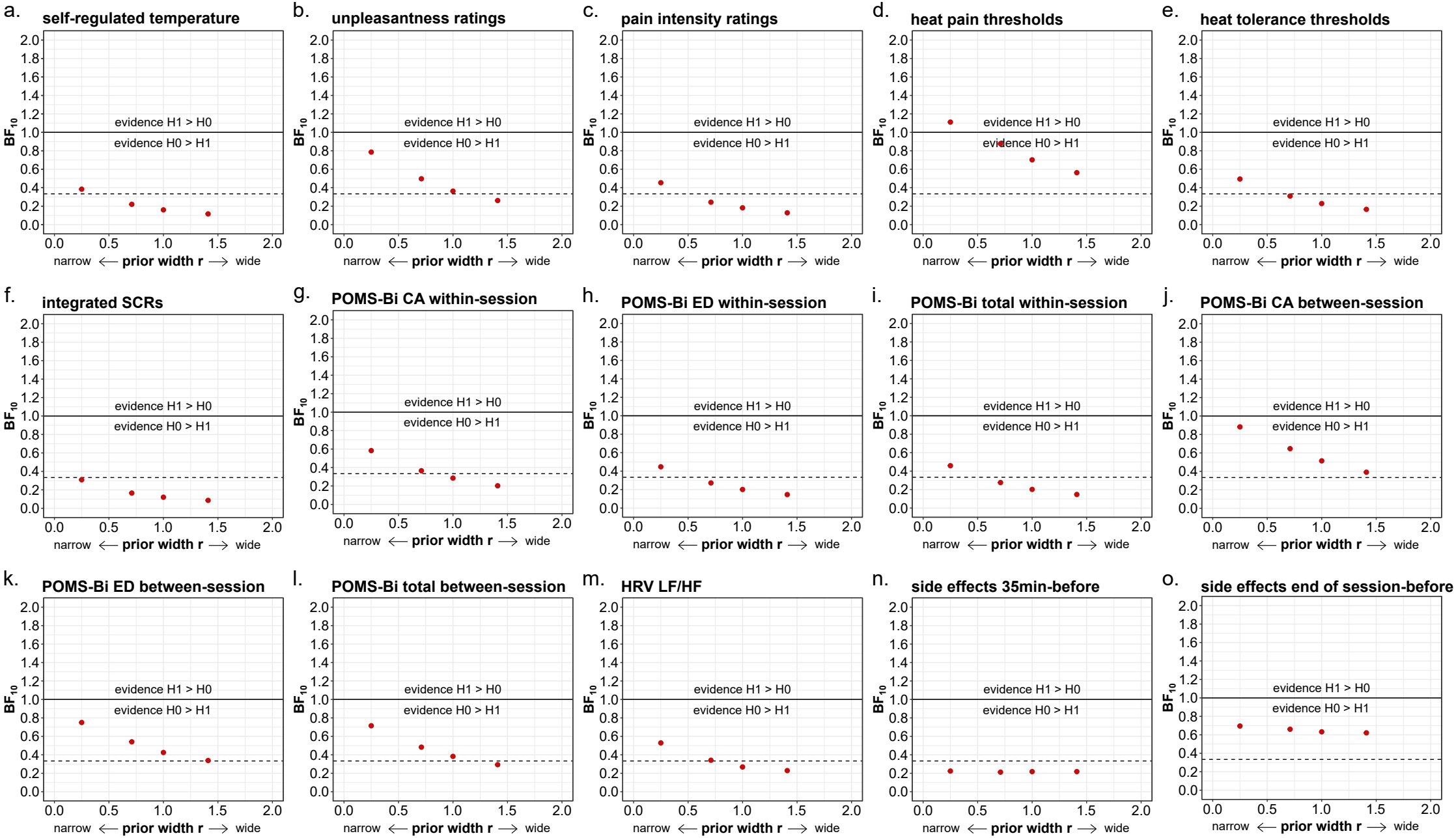
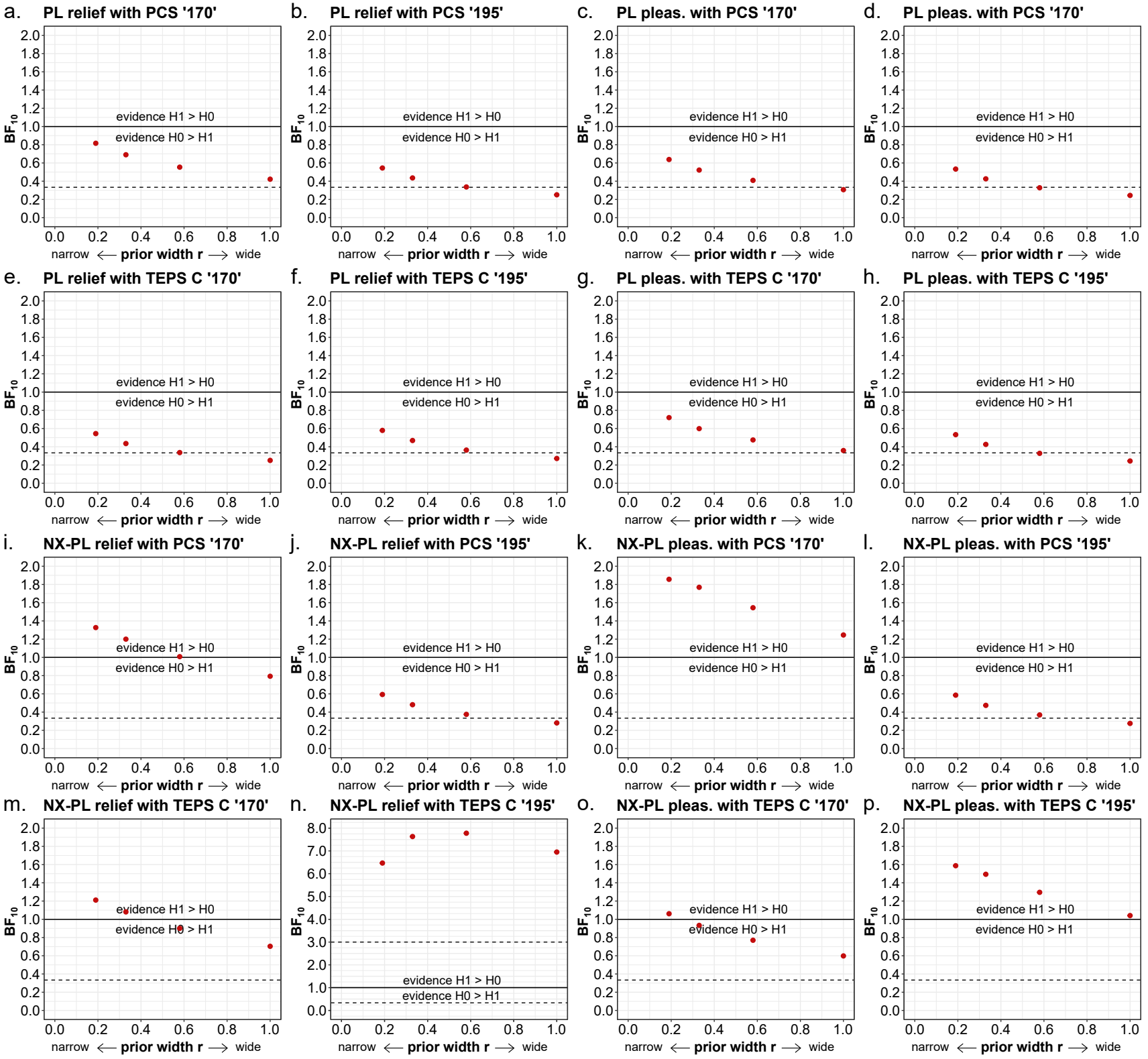


Supplemental Digital Content 1. Bayesian analysis robustness check for null effects of naltrexone and null effects of session. Bayes factor analysis was repeated three times using priors of different widths (narrow: r = 0.25, wide: r = 1, ultrawide: r = 1.41) to gain insight into the robustness of the Bayes factors depending on prior choice. In outcomes (a-f) and (n, o), the effect of naltrexone was investigated, in outcomes (g-i), the drug:timepoint interaction effect was examined and in outcomes (j-m), the effect of session was assessed. Outcomes (a-m) are ANOVA designs, while outcomes (n, o) represent Bayesian Wilcoxon signed-rank tests. Red dots represent the BF_{10} values (i.e., ratio of likelihood of the null hypothesis (H0) "absence of an effect" to likelihood of the alternate hypothesis (H1) "presence of an effect" for the different prior widths. BF_{10} values above 1 indicate stronger evidence for H1, while BF_{10} values below 1 indicate stronger evidence for H0. The dotted line at 0.33 represents the threshold for moderate evidence for H0 [24,29]. CA, composed-anxious; ED, elated-depressed; HRV, heart rate variability; LF/HF, ratio of low frequency to high frequency components of HRV; POMS-Bi, Profile of Mood States Bipolar scale; SCR, skin conductance response.

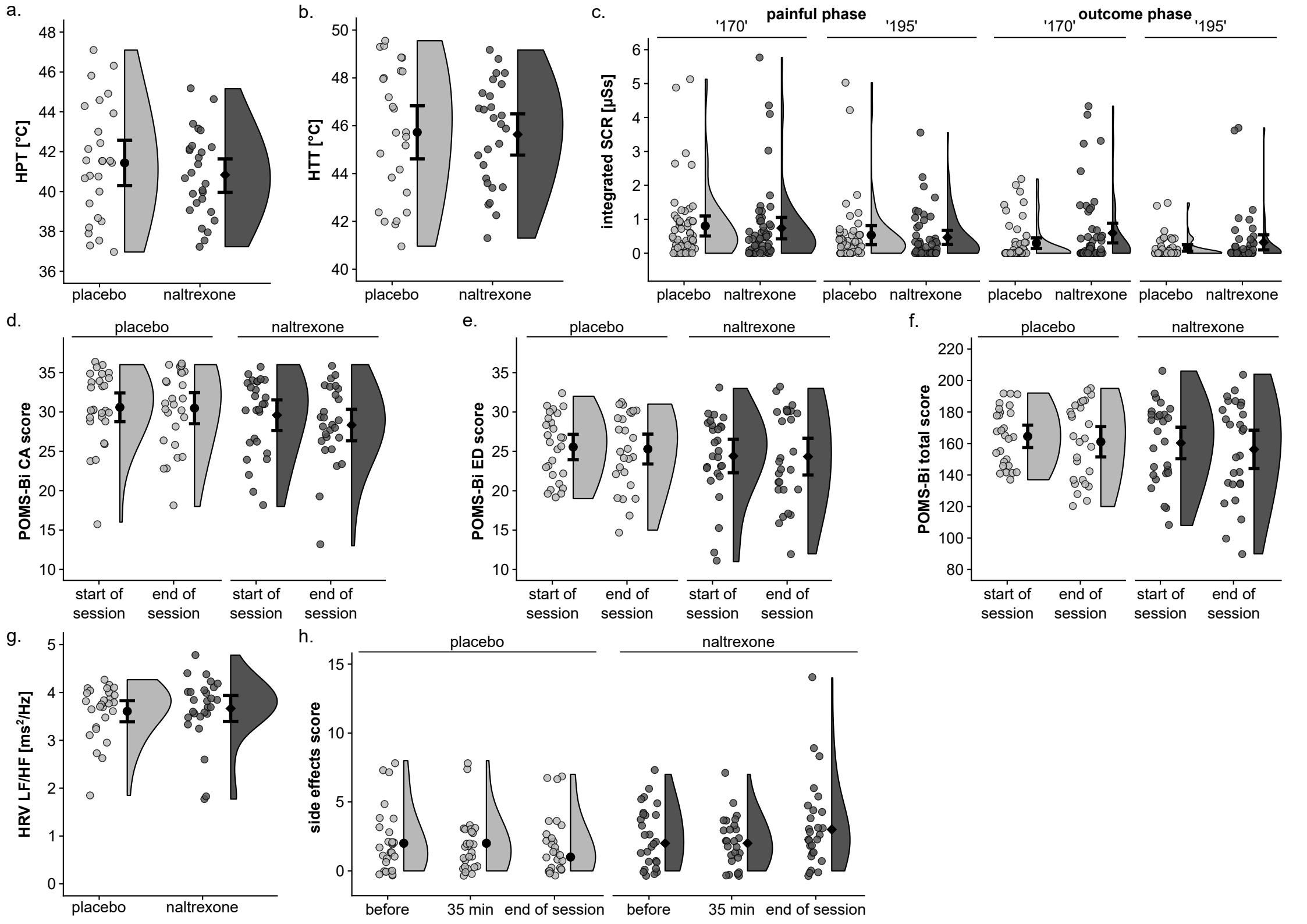


Supplemental Digital Content 2. Bayesian analysis robustness check for null findings in correlations. Bayes factor analysis was repeated three times using priors of different widths (narrow: r = 0.19, wide: r = 0.58, ultrawide: r = 1) to gain insight into the robustness of the Bayes factors depending on prior choice. It was examined whether or not a correlation is present. "Relief"/"pleas." refers to maximal relief/pleasantness ratings in the placebo session (a-h) and "NX-PL relief"/"NX-PL pleas." refers to the differences in maximal relief/ pleasantness ratings between the naltrexone and placebo session (i-p). Red dots represent the BF10 values (i.e., ratio of likelihood of H0 "absence of a correlation" to likelihood of H1 "presence of a correlation") for the different prior widths. BF10 values above 1 indicate stronger evidence for H1, while BF10 values below 1 indicate stronger evidence for H0. The dotted line at 0.33 represents the threshold for moderate evidence for H0, while the dotted line at 3 represents the threshold for moderate evidence for H1 [24,29]. NX, naltrexone; PCS, Pain Catastrophizing Scale; PL, placebo; TEPS C, Temporal Experience of Pleasure Scale for consummatory pleasure.

Supplemental Digital Content 3		Results of potential confounders											
read-out	**trial**	**naltrexone**	**placebo**	**effect drug**		**effect intensity**		**effect outcome**				**Potential influential observations**	**Change in statistical inference**
		mean ± SD	mean ± SD	F	p	F	p	F	p			N/total N	
temperature [°C]													
	'170' relief	34.6 ± 5.1	34.8 ± 4.7	0.74	0.39	55.6	< 0.001	0.00036	0.98			20/208	no
	'170' pleasantness	34.7 ± 5.1	34.8 ± 4.6										
	'195' relief	36.0 ± 5.1	36.8 ± 5.4										
	'195' pleasantness	36.0 ± 5.2	36.7 ± 5.4										
unpleasantness													
	'170' relief	64.9 ± 19.8	67.4 ± 17.9	2.54	0.11	38.19	< 0.001	0.0024	0.96			18/207	no
	'170' pleasantness	64.9 ± 19.8	66.9 ± 18.5										
	'195' relief	75.3 ± 22.0	77.5 ± 19.9										
	'195' pleasantness	78.0 ± 19.1	76.7 ± 19.9										
intensity													
	'170' relief	169.7 ± 7.4	169.1 ± 6.9	1.42	0.23	246.45	< 0.001	0.14	0.71			9/207	no
	'170' pleasantness	166.7 ± 11.4	168.6 ± 5.9										
	'195' relief	184.6 ± 18.8	187.8 ± 9.8										
	'195' pleasantness	186.6 ± 18.8	188.0 ± 8.4										
										effect gender			
thresholds [°C]										F	p		
HPT (females)		40.0 ± 1.8	40.5 ± 2.7	3.01	0.095					5.23	0.031	7/54	no
HPT (males)		41.7 ± 2.2	42.5 ± 2.8										
HTT (females)		44.7 ± 2.0	44.4 ± 2.5	0.22	0.64					7.52	0.011	7/53	no
HTT (males)		46.7 ± 1.9	47.0 ± 2.4										
										effect phase			
integrated SCRs [µSs]										F	p		
painful phase	'170' relief	0.71 ± 1.19	0.70 ± 1.03	1.04	0.31	17.80	< 0.001	0.27	0.60	31.96	< 0.001	35/402	no
	'170' pleasantness	0.77 ± 1.13	0.90 ± 1.12										
outcome phase	'170' relief	0.57 ± 1.10	0.27 ± 0.54										
	'170' pleasantness	0.61 ± 1.02	0.32 ± 0.60										
painful phase	'195' relief	0.49 ± 0.82	0.59 ± 1.06										
	'195' pleasantness	0.44 ± 0.62	0.48 ± 0.88										
outcome phase	'195' relief	0.34 ± 0.77	0.12 ± 0.20										
	'195' pleasantness	0.30 ± 0.77	0.18 ± 0.41										

	timepoint			interaction effect drug x timepoint						
POMS-Bi within-session				F	p					
CA	session start	29.6 ± 4.9	30.6 ± 4.6	0.76	0.39				12/108	no
	session end	28.3 ± 5.1	30.5 ± 5.0							
ED	session start	24.4 ± 5.4	25.6 ± 4.1	0.026	0.87				16/108	no
	session end	24.3 ± 5.9	25.3 ± 4.8							
total	session start	160.3 ± 25.3	164.5 ± 18.1	0.017	0.90				14/108	no
	session end	156.3 ± 30.8	161.1 ± 24.3							
				effect session						
POMS-Bi between-session				F	p					
CA	session start	29.6 ± 4.9	30.6 ± 4.6	2.22	0.15				8/54	no
ED	session start	24.4 ± 5.4	25.6 ± 4.1	1.68	0.21				6/54	no
total	session start	160.3 ± 25.3	164.5 ± 18.1	1.47	0.24				10/54	no
HRV [ms^2/Hz]										
LF/HF	session start	3.67 ± 0.69	3.61 ± 0.56	0.58	0.45				6/54	no

If not reported, interaction or gender effects were not significant and removed from the model. Integrated SCRs were log10(x+1) transformed to meet requirements of general linear mixed models. For interpretation purposes, untransformed values are reported. CA, composed-anxious; ED, elated-depressed; HPT, heat pain threshold; HRV, heart rate variability; HTT, heat tolerance threshold; LF/HF, ratio of low frequency to high frequency components of HRV; POMS-Bi, Profile of Mood States Bipolar scale; SCR, skin conductance response; SD, standard deviation.

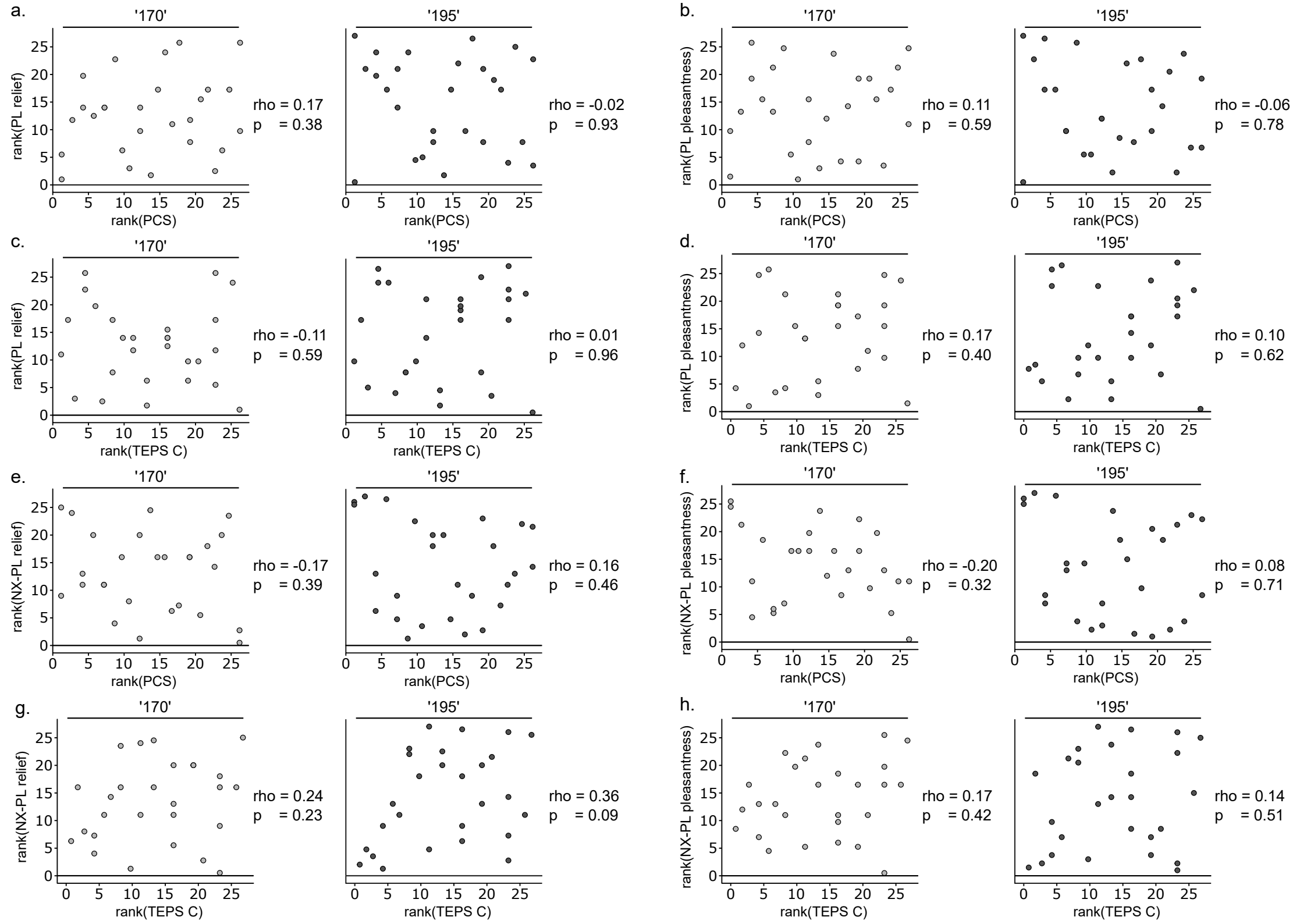

a.
HPT [°C]
placebo
naltrexone
b.
HTT [°C]
placebo
naltrexone
c.
painful phase
outcome phase
'170'
'195'
'170'
'195'
integrated SCR [µSs]
placebo
naltrexone
d.
POMS-Bi CA score
placebo
naltrexone
start of session
end of session
e.
POMS-Bi ED score
f.
POMS-Bi total score
g.
HRV LF/HF [ms²/Hz]
h.
side effects score
before
35 min
end of session

Supplemental Digital Content 4. Outcomes with null effects of naltrexone and null effects of session. Naltrexone had no effect on heat pain thresholds (a), heat tolerance thresholds (b), integrated SCRs after the onset of the painful phase and after the onset of the outcome phase (c), change of POMS-Bi scores over the course of the sessions (d-f), and side effects (h). "Baseline" (i.e., start of the session) POMS-Bi scores (d-f) and HRV LF/HF component (g) did not differ between the placebo and naltrexone sessions. The raincloud plots [2] display the raw data (grey dots), means and 95% confidence intervals (black dots/diamonds and bars) and probability distributions (vertical "clouds"), except for (h) where black dots/diamonds represent the medians. In (c), each dot represents the integrated SCRs averaged over two trials with the same outcome phase (i.e., relief or pleasantness) at the respective target pain intensity resulting in two datapoints per participant (N = 27) per raincloud plot. CA, composed-anxious; ED, elated-depressed; HPT, heat pain threshold; HRV, heart rate variability; HTT, heat tolerance threshold; LF/HF, ratio of low frequency to high frequency components of HRV; POMS-Bi, Profile of Mood States Bipolar scale; SCR, skin conductance response.

Supplemental Digital Content 5		Results of Bayes factor analysis and random sampling		
Bayes factors (BF_{10} < 1 supports H0, BF_{10} > 1 supports H1)				
read-out		**BF_{10} effect drug**	**BF_{10} interaction effect drug:timepoint**	**BF_{10} effect session**
temperature [°C]		0.22 ± 0.93 %		
unpleasantness		0.50 ± 2.07 %		
intensity		0.24 ± 1.93 %		
HPT [°C]		0.88 ± 1.52 %		
HTT [°C]		0.31 ± 1.12 %		
integrated SCRs [µSs]		0.17 ± 2.51 %		
POMS-Bi within-session				
CA			0.36 ± 1.48 %	
ED			0.27 ± 1.24 %	
total			0.28 ± 1.26 %	
POMS-Bi between-session				
CA				0.65 ± 1.59 %
ED				0.54 ± 1.24 %
total				0.48 ± 1.05 %
HRV [ms^2/Hz]				
LF/HF				0.34 ± 1.51 %
			Wilcoxon signed-rank test	
			V	**p**
side effects	35min – before	0.21	104.5	0.72
	end of session - before	0.66	106	0.13
correlation	**intensity**	**BF_{10} correlation**		
placebo relief with PCS	'170'	0.69 ± 0 %		
	'195'	0.44 ± 0 %		
placebo pleasantness with PCS	'170'	0.52 ± 0 %		
	'195'	0.43 ± 0 %		
placebo relief with TEPS C	'170'	0.44 ± 0 %		
	'195'	0.47 ± 0 %		
placebo pleasantness with TEPS C	'170'	0.60 ± 0 %		
	'195'	0.43 ± 0 %		
delta NX-PL relief with PCS	'170'	1.20 ± 0 %		
	'195'	0.48 ± 0 %		

delta NX-PL pleasantness with PCS	'170'	1.77 ± 0 %			
	'195'	0.47 ± 0 %			
delta NX-PL relief with TEPS C	'170'	1.08 ± 0 %			
	'195'	7.63 ± 0 %			
delta NX-PL pleasantness with TEPS C	'170'	0.93 ± 0 %			
	'195'	1.49 ± 0 %			
random sampling					
correlation	**intensity**	**observed correlation**	**random correlations**		
		rho	mean rho	97.5% quantile	2.5% quantile
placebo relief with delta NX-PL relief	'170'	-0.36	-0.50	-0.31	-0.68
	'195'	-0.49	-0.51	-0.33	-0.68
placebo pleasantness with delta NX-PL pleasantness	'170'	-0.56	-0.62	-0.47	-0.76
	'195'	-0.43	-0.54	-0.32	-0.72

Bayes factor analysis results are reported for medium prior widths (i.e., r = 0.71 for ANOVA designs and Bayesian Wilcoxon signed-rank test, r = 0.33 for correlational designs). BF_{10} expresses the likelihood of the alternate hypothesis (H1) to the likelihood of the null hypothesis (H0). A value for BF_{10} between 1 and 3 is considered anecdotal (i.e., weak) evidence for H1, while a value between 3 and 10 suggests moderate evidence for H1. BF_{10} values between 1 and 0.33 represent anecdotal evidence for H0 and values between 0.33 and 0.1 are considered moderate evidence for H0 [24, 29]. A robustness check of Bayes factors depending on different prior widths (for ANOVA designs and Bayesian Wilcoxon signed-rank test: narrow: r = 0.25, wide: r = 1, ultrawide: r = 1.41; for correlational designs: narrow: r = 0.19, wide: r = 0.58, ultrawide: r = 1) is displayed in Supplemental Digital Content 1. Random sampling was performed to compare observed correlation coefficients with random correlation coefficients given the flaw of A ~ B-A and an inherent correlation of A and B. The mean, the 97.5% quantile and the 0.25% quantile of the distribution of the 10'000 random correlation coefficients are presented. If the observed correlation coefficients were outside of the limits of the 0.25% and the 97.5% quantiles of the random distribution, the observed effects would be deemed to be different from an effect expected by the flaw of A ~ B-A with the probability of 5% to commit a Type I error. CA, composed-anxious; ED, elated-depressed; HPT, heat pain threshold; HRV, heart rate variability; HTT, heat tolerance threshold; LF/HF, ratio of low frequency to high frequency components of HRV; NX, naltrexone; PCS, Pain Catastrophizing Scale; PL, placebo; POMS-Bi, Profile of Mood States Bipolar scale; SCR, skin conductance response; TEPS C, Temporal Experience of Pleasure Scale for consummatory pleasure.


a.
'170'
'195'
rank(PL relief)
rank(PCS)
rho = 0.17
p = 0.38
rho = -0.02
p = 0.93
b.
rank(PL pleasantness)
rho = 0.11
p = 0.59
rho = -0.06
p = 0.78
c.
rank(TEPS C)
rho = -0.11
p = 0.59
rho = 0.01
p = 0.96
d.
rho = 0.17
p = 0.40
rho = 0.10
p = 0.62
e.
rank(NX-PL relief)
rho = -0.17
p = 0.39
rho = 0.16
p = 0.46
f.
rank(NX-PL pleasantness)
rho = -0.20
p = 0.32
rho = 0.08
p = 0.71
g.
rho = 0.24
p = 0.23
rho = 0.36
p = 0.09
h.
rho = 0.17
p = 0.42
rho = 0.14
p = 0.51

Supplemental Digital Content 6. Outcomes with null findings in correlations. Spearman's rho correlations (based on ranks) are shown between maximal relief/pleasantness ratings in the placebo session and PCS (a, b), as well as TEPS consummatory scores (c, d) at both target pain intensities (i.e.,'170' and '195'). Further, Spearman's rho correlations are shown between delta NX-PL relief/pleasantness ratings (i.e., the differences in maximal relief/pleasantness ratings between the naltrexone and placebo session) and PCS (e, f), as well as TEPS consummatory scores (g, h) at both target pain intensities (i.e.,'170' and '195'). NX, naltrexone; PCS, Pain Catastrophizing Scale; PL, placebo; TEPS C, Temporal Experience of Pleasure Scale for consummatory pleasure.