

Supplementary Information

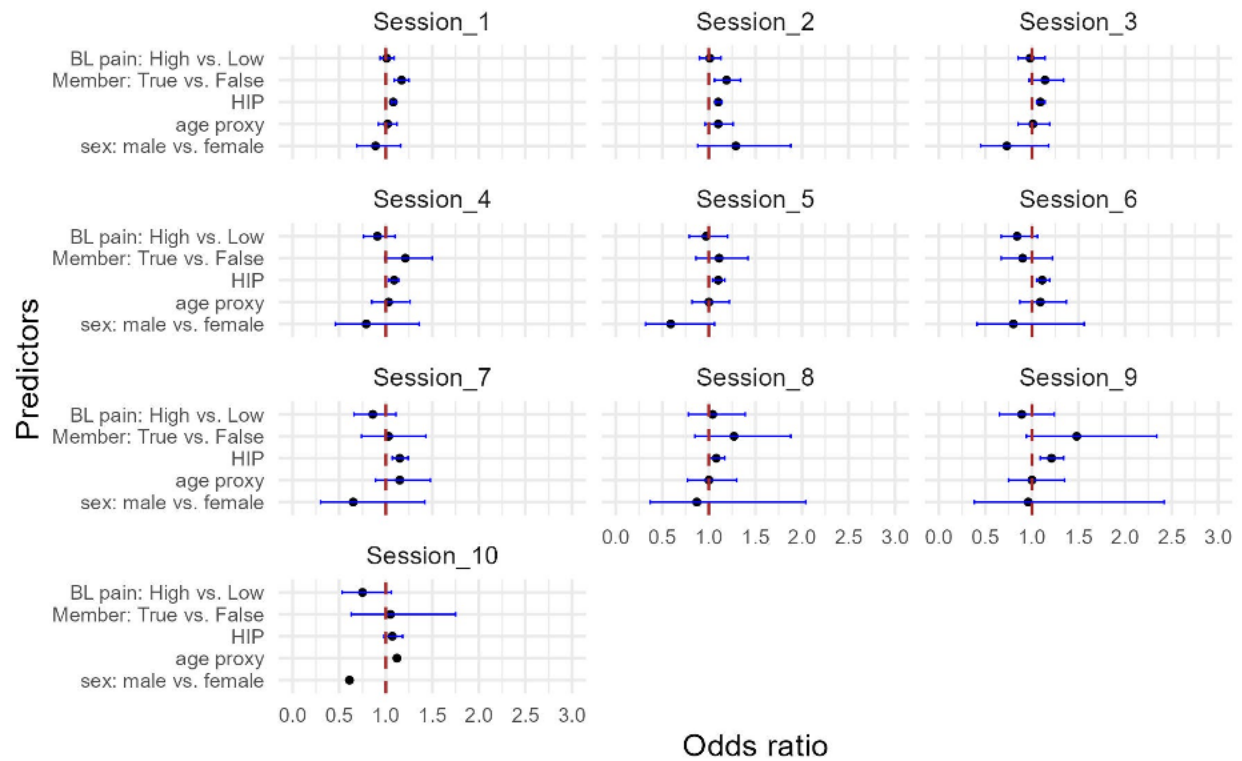
Index

1. Suppl. Results
 - 1.1. Sensitivity analysis by software update
 - 1.2. Sensitivity analysis by completeness of the data
 - 1.3. User participation based on first session response
 - 1.4. Association of hypnotizability (HIP) to exercise type, number of sessions, and membership
 - 1.5. Pain Exercise Type
2. Suppl. Figure 1: Estimated Mean Pain Reduction within Sessions for First 10 Sessions Before and After Software Update.
3. Suppl. Figure 2: Odds Ratio for Achieving a Greater Percent Reduction in Pain.
4. Suppl. Table 1.1: Estimated Mean Pain Reduction and Effect Sizes Before Software Update.
1.2: Estimated Mean Pain Reduction and Effect Sizes After Software Update.
5. Suppl. Table 2: Correlations between HIP Score and Pain Reduction in the First 10 Sessions.

1. Supplementary Results

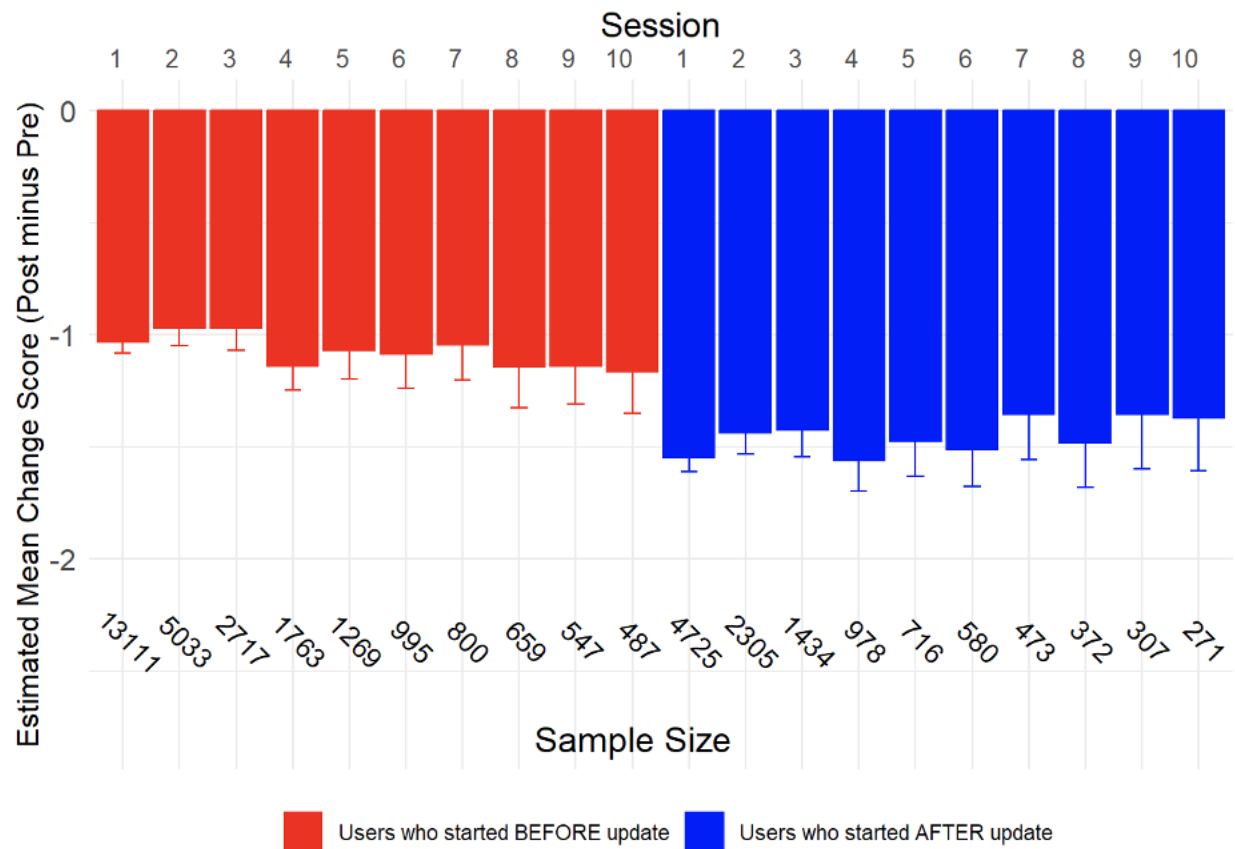
- 1.1. Sensitivity analysis by software update: The LME model analyzing data before the software update, where pre-scores of 10 and post-scores of 1 were discarded, yielded effect sizes ranging from -0.45 to -0.57 for the first 10 sessions (Suppl. Table 1.1). These are lower than the effect sizes for the combined sample (Table 2). Analysis of the first 10 sessions after the software update, where no scores needed to be set to missing, yielded effect sizes ranging from -0.60 to -.71 (Suppl. Table 1.2), which is higher than those of the combined sample (Table 2). This sensitivity analysis thus suggests the estimated change scores for the combined sample may have been underestimated due to our conservative approach to handling the default assignments before the software update (Suppl. Figure 2).
- 1.2. Sensitivity analysis by completeness of the data: LME models limited to only those users with complete data for both pre and post scores showed effect sizes ranging from -0.63 to -0.72. This suggests that the results from Figure 2 and Table 2 may be underestimations caused by those who had missing data, which is partly due to the conservative exclusion of extreme scores prior to the software update (i.e. discarding pre-scores of 10 and post-scores of 1).
- 1.3. User participation based on first session response: To examine whether session 1 pain score is associated with repeat use, we tested the relationships of session 1 change score, pre-score (baseline pain), and post-score with the number of repeat sessions using Spearman Rank-Order correlations. Our analysis revealed no meaningful association between first session change scores and subsequent number of sessions participated in (Spearman ρ = -0.049, 95% CI = -0.071, -0.029). Users with higher first session pre-intervention pain scores did not participate in a meaningfully different number of sessions (Spearman ρ = -0.005, 95% CI = -0.019, 0.009). The association between the number of subsequent sessions and first session post-intervention pain scores was not meaningful (Spearman ρ = -0.040, 95% CI = -0.063, -0.018). These findings suggest that neither first session baseline pain, post-intervention pain, nor change in pain predicted repeat use of the intervention.

- 1.4. Association of hypnotizability (HIP) to exercise type, number of sessions, and membership: We evaluated associations between HIP score and exercise type, membership, and whether users did more than one session. Of the 19,063 users who used any pain management exercise, only 854 users (4.5%) were “switchers”, logging more than one exercise type. Therefore, over 95% of users used a single exercise type. Due to the lack of within-user variation in exercise type, an LME model could not reliably estimate the effect of exercise type for the entire sample. Thus, a primary analysis used linear regression including the entire sample, assuming independence of the data, and a sensitivity analysis used an LME model restricted to only the switchers. Based on the linear regression model, those who used the “Pleasant Place” ($\beta = 0.37$, 95% CI = 0.16, 0.59) or “Slow Moving Liquid” ($\beta = 0.29$, 95% CI = 0.10, 0.47) exercise types had higher HIP scores than those who used the standard type. Among the switchers, there was no association between HIP and the pain exercise type, suggesting the exercise type these users selected was not associated with their hypnotizability score. User hypnotizability as measured by the HIP was higher for users who tried more than one exercise type than those who tried only one ($\beta = 0.45$, 95% CI = 0.19, 0.71). Additionally, higher hypnotizability users were likely to participate in more sessions ($\beta = 0.37$, 95% CI = 0.25, 0.49). Members had higher HIP scores than non-members ($\beta = 0.97$, 95% CI = 0.85, 1.09).
- 1.5. Pain Exercise Type: Three additional non-standard exercise types were available to users, all sharing the key components of the standard protocol, with minor differences. The “Be Kind” protocol involved suggestions that encouraged developing a compassionate attitude to one’s body, imagining one’s body as a child coping with discomfort. The “Pleasant Place” protocol focused more heavily on guided imagery to dissociate one’s attention from his or her pain by imagining being somewhere else that is pleasant and far away from their body. The “Slow Moving Liquid” protocol involved suggestions encouraging one to imagine his or her pain as a malleable substance that can be moved and transformed, thereby enhancing a sense of agency. The vast majority of users only utilized the standard type. Pain reduction was not statistically different across exercise types in the first 10 sessions, though we observed weaker pain reduction in the “Pleasant Place” exercise relative to the standard exercise type in sessions 2 ($\beta = .39$, 95% CI = 0.01, 0.78) and 9 ($\beta = .95$, 95% CI = 0.06, 1.84). These isolated session-specific findings should be interpreted as exploratory.



Note: Standard errors for age and sex in session 10 could not be estimated due to the absence of an outcome category (pain increased).

Supplementary Figure 1: Odds Ratio for Achieving a Greater Percent Reduction in Pain.



Supplementary Figure 2: Estimated Mean Pain Reduction within Sessions for First 10 Sessions Before and After Software Update.

Session	N LME	Pre-score Mean (SD)	Change Score Mean	Std Err	95% LB	95% UB	Effect Size
1	13111	4.98 (2.22)	-1.032	0.025	-1.080	-0.983	-0.47
2	5033	5.12 (2.16)	-0.969	0.040	-1.047	-0.891	-0.45
3	2717	5.18 (2.15)	-0.970	0.051	-1.071	-0.870	-0.45
4	1763	5.24 (2.10)	-1.141	0.054	-1.247	-1.036	-0.54
5	1269	5.27 (2.10)	-1.071	0.064	-1.196	-0.945	-0.51
6	995	5.32 (2.10)	-1.086	0.077	-1.236	-0.935	-0.52
7	800	5.34 (2.06)	-1.044	0.080	-1.201	-0.887	-0.51
8	659	5.49 (2.05)	-1.144	0.093	-1.326	-0.962	-0.56
9	547	5.4 (2.04)	-1.139	0.086	-1.308	-0.969	-0.56
10	487	5.35 (2.05)	-1.166	0.095	-1.352	-0.980	-0.57

Supplementary Table 1.1. Estimated Mean Pain Reduction and Effect Sizes Before Software Update.

Session	N LME	Pre-score Mean (SD)	Change Score Mean	Std Err	95% LB	95% UB	Effect Size
1	4725	5.43 (2.23)	-1.548	0.032	-1.610	-1.486	-0.69
2	2305	5.3 (2.27)	-1.439	0.048	-1.533	-1.345	-0.64
3	1434	5.4 (2.28)	-1.427	0.060	-1.544	-1.310	-0.63
4	978	5.38 (2.21)	-1.562	0.069	-1.697	-1.426	-0.71
5	716	5.26 (2.24)	-1.476	0.079	-1.631	-1.321	-0.66
6	580	5.3 (2.19)	-1.512	0.084	-1.677	-1.348	-0.69
7	473	5.28 (2.24)	-1.356	0.102	-1.556	-1.156	-0.60
8	372	5.21 (2.14)	-1.481	0.102	-1.681	-1.280	-0.69
9	307	5.24 (2.18)	-1.354	0.124	-1.597	-1.112	-0.62
10	271	5.06 (2.19)	-1.372	0.120	-1.608	-1.137	-0.63

Supplementary Table 1.2: Estimated Mean Pain Reduction and Effect Sizes After Software Update.

Session	N	Spearman rho (ρ)	Upper CI	Lower CI
1	3636	-0.14 ***	-0.17	-0.10
2	1794	-0.16 ***	-0.21	-0.12
3	1119	-0.15 ***	-0.20	-0.09
4	800	-0.15 ***	-0.22	-0.08
5	629	-0.11 ***	-0.19	-0.03
6	513	-0.16 ***	-0.24	-0.07
7	426	-0.27***	-0.36	-0.18
8	349	-0.15***	-0.26	-0.04
9	275	-0.30***	-0.41	-0.19
10	255	-0.14***	-0.26	-0.01

Supplementary Table 2: Correlations between HIP Score and Pain Reduction in the First 10 Sessions. There were small, statistically significant effect sizes ($-0.3 < \rho < -0.11$) in sessions 1-10. N indicates the number of participants with (1) pain change score and (2) baseline HIP score, at each session (i.e., those with missing data for either change score or HIP score were excluded from this analysis). The symbol ‘***’ indicates statistical significance at the 0.05 alpha level.