BJPO/2015/000745
doi: 10.1192/bjpo.bp.115.000745

Data supplement
BJPO/2015/000745
doi: 10.1192/bjpo.bp.115.000745

Table DS1. Means and Standard Deviations for CAPS Scores at All Time-points Over All Treatment Groups

Timepoint	***M***	***SD***
Baseline	63.08	18.62
Posttreatment	33.70	21.08
3-month FU	30.82	21.93
6-month FU	25.58	22.13

Note. CAPS = Clinician Administered PTSD Scale.

BJPO/2015/000745
doi: 10.1192/bjpo.bp.115.000745

Table DS2. Correlations Among the Covariates at Baseline

	AxEx	BDI	Diss	CAPS	IIP	NMR
AxEx	1	0.27	0.28	0.16	0.49	-0.48
BDI	0.27	1	0.55	0.48	0.55	-0.6
Diss	0.28	0.55	1	0.47	0.5	-0.38
CAPS	0.16	0.48	0.47	1	0.31	-0.29
IIP	0.49	0.55	0.5	0.31	1	-0.51
NMR	-0.48	-0.6	-0.38	-0.29	-0.51	1
#CTs	0.03	0.09	0	0.22	0.01	-0.12

Note. AxEx = State-trait Anger Expression Inventory; BDI = Beck Depression Inventory- II; Diss = Trauma Symptom Inventory Dissociation Subscale; CAPS = Clinician-administered PTSD Scale; IIP = Inventory of Interpersonal Problems; NMR = Negative Mood Regulation Scale; #CTs = number of childhood traumas.

BJPO/2015/000745
doi: 10.1192/bjpo.bp.115.000745

Table DS3. The Linear Combination of the Six Baseline Patient Characteristics That Has the Highest Test Statistic for Testing the Significance of the Interaction Term Between Treatment, Time, and the Linear Combination

Baseline Characteristics	**GEM**	
	Coefficients of the standardized predictor	**Coefficients of the original predictor**
AxEx	-12.77	-123.62
BDI	2.09	20.40
Diss	1.90	1.77
CAPS	0.96	17.82
IIP	7.58	4.45
NMR	-0.39	-6.17
GEM *p*-value	0.0025	
GEM permutation *p*-value	0.0081	

Note. AxEx = State-trait Anger Expression Inventory; BDI = Beck Depression Inventory- II; Diss = Trauma Symptom Inventory Dissociation Subscale; CAPS = Clinician-administered PTSD Scale; IIP = Inventory of Interpersonal Problems; NMR = Negative Mood Regulation Scale.

BJPO/2015/000745
doi: 10.1192/bjpo.bp.115.000745

Detailed Explanation of Moderator analysis

Standard Moderator Analysis. In a randomized clinical trial RCT, what is commonly referred as a "standard moderator analysis" can be described as follows. Let *Y* denotes the outcome of interest at post-treatment; let *A* denotes the randomised treatment assignment with *A=0* if a subjects is randomized to the control treatment condition and *A=1* for the experimental condition; and let *Z* be a patient-level baseline covariate, such as age, severity of symptoms or a specific comorbid condition. If the outcome *Y* is a continuous variable, reasonably assumed to follow Gaussian distribution conditional on covariates, and the subscript '*i*' indicates the value of the i^{th} patient, standard moderator analysis is based on the following linear model

(1) $$Y_i = \beta_0 + \beta_1 I_i^{\{A=1\}} + \beta_2 Z_i + \beta_3 Z_i I_i^{\{A=1\}} + e_i,$$

where $I_i^{\{A=1\}}$is an indicator for whether subject *i* received treatment *A=1* or not and ε is a random error. A significant interaction term ($\beta_3 \neq 0$) would indicate that *Z* moderates the effect of treatment *A=1* compared to the effect of treatment *A=0*. This is usually expressed shortly as "*Z* is a moderator of the effect of treatment on the outcome".

If the outcome cannot be reasonably assumed to be normally distributed conditional on the covariates, then a generalized linear model (GLM) is applied, where the linear predictor on the right hand side of (1) would model a link function *g(Y)* of the outcome, rather than the outcome *Y* itself. For example, in the case of binary outcome (e.g., remission or non-remission) the link function *g* would be *logit*, i.e., $g(Y) = \log\frac{P(Y=1)}{P(Y=0)}$ and the GLM model would be logistic regression.

If the RCT has more than 2 treatment conditions, as it is in the study presented in the main paper, instead of one indicator for treatment, there will be (*k-1*) treatment indicators where *k* is the number of treatment conditions in the RCT. In our study *k=3*, i.e., *A=0, 1* and *2*, thus instead of (1) the appropriate model is:

(2) $$Y_i = \beta_0 + \beta_1 I_i^{\{A=1\}} + \beta_2 I_i^{\{A=2\}} + \beta_3 Z_i + \beta_4 Z_i I_i^{\{A=1\}} + \beta_5 Z_i I_i^{\{A=2\}} + e_i.$$

In this case *Z* is a moderator of treatment effect if model (2) is better than model (2*)

(2*) $$Y_i = \beta_0 + \beta_1 I_i^{\{A=1\}} + \beta_2 I_i^{\{A=2\}} + \beta_3 Z_i + e_i,$$

usually compared based on a χ^2 likelihood ratio test (LRT)on *k-1* degrees of freedom or based on Analysis of Variance (ANOVA) F test.

In RTCs where the outcome Y is assessed over time (as in our study – at immediately post-treatment, at 3 months follow up and 6 months follow up), the standard moderator analysis includes time and is based on a "standard longitudinal moderator model", which in the case of only two treatment conditions is

(3) $$Y_{it} = \beta_0 + \beta_1 I_i^{\{A=1\}} + \beta_2 Z_i + \beta_3 t + \beta_4 Z_i I_i^{\{A=1\}} + \beta_5 t I_i^{\{A=1\}} + \beta_6 Z_i t + \beta_7 Z_i t I_i^{\{A=1\}} + error_i,$$

where $error_i$includes random subject effects (e.g., intercepts and slopes) plus a random error. Generalized linear mixed effects models are typically used to fit models similar to (3). The significance of the 3-way interaction term β_7 would indicate that the comparison between treatment *A=0* and *A=1* with respect to the course of the outcome *Y* over time depends on *Z*. If β_7 is not different from 0, then one can look at the coefficient of the interaction between *Z* and the treatment indicator – a significantly different from zero β_4 would indicate that *Z* moderates

the effect of treatment on the average outcome, but does not moderate how the effect of treatment changes over time.

The specific longitudinal mixed effects model for the outcome *Y* (CAPS scores) at the time points post-treatment, 3m FU and 6m FU employed in this study with 3 treatment conditions is

$$(4)\quad Y_{it} = \beta_0 + \beta_1 I_i^{\{A=STAIR\}} + \beta_2 I_i^{\{A=Ex\}} + \beta_3 Z_i + \beta_4 t + \beta_5 Z_i I_i^{\{A=STAIR\}} + \beta_6 Z_i I_i^{\{A=Ex\}} + \beta_7 t I_i^{\{A=STAIR\}} + \beta_8 t I_i^{\{A=Ex\}} + \beta_9 Z_i t + \beta_{10} Z_i t I_i^{\{A=STAIR\}} + \beta_{11} Z_i t I_i^{\{A=Ex\}} + error_i.$$

Testing whether *Z* is a moderator of treatment effects was based on the comparison of (4) against model

$$(4^*)\quad Y_{it} = \beta_0 + \beta_1 I_i^{\{A=STAIR\}} + \beta_2 I_i^{\{A=Ex\}} + \beta_3 Z_i + \beta_4 t + \beta_5 Z_i I_i^{\{A=STAIR\}} + \beta_6 Z_i I_i^{\{A=Ex\}} + \beta_7 t I_i^{\{A=STAIR\}} + \beta_8 t I_i^{\{A=Ex\}} + \beta_9 Z_i t + error_i$$

using LRT (a 2 degrees of freedom χ^2test). Each of the individual six baseline predictors is subjected to moderator analysis using the comparison of models (4) and (4*). The p-values of the LRT are reported in Table 1 of the main text.

Finding the GEM

Let denote the six baseline predictors by X_1, X_2, X_3, X_4, X_5 and X_6. We are looking for a linear combination (Z) of these variables

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_4 + \alpha_5 X_5 + \alpha_6 X_6,$$

such that the test statistics for comparing model (4) against model (4*) ($\chi^2(2)$ LRT) has the largest value, and equivalently, has the smallest p-value. This linear combination was determined numerically using the function `stats::optim` and `BB::spg` in R.[35]

GEM p-value

The GEM approach seeks to determine a linear combination of predictors that maximizes the evidence of an interaction effect. If there are no interaction effects between predictors and treatment indicators, then the GEM approach would tend to generate anti-conservative p-values when testing for an interaction in the estimated GEM models. A straightforward remedy to this problem is to generate GEMs from many "permuted" data sets(i.e., sets) obtained by randomly permuting the treatment labels among the data points. For each permuted data set, a solution for the linear combination is obtained; then a GEM corresponding to this solution is constructed and a *p*-value for the interaction of that GEM with treatment can be computed. Thus, a valid *p*-value for testing for an interaction effect can be calculated by

Permutation p-value = {Proportion of permuted p-values < original p-value.}

In our case, we obtain the permuted p-values for the LRT to compare models (4) and (4*).

Testing for difference between the average outcome under a given treatment vs. the average outcome under treatment decisions based on BP

Notice that the average outcome under a treatment decision based on BP is obtained from the outcomes of patients in levels 1 and 2 of BR treated with SC/Ex and the outcomes of patients in

levels 3 and 4 of BP treated with STAIR/SC. The average outcome under a decision to treat everyone with SC/Ex is the obtained from the outcomes of all patients (levels 1, 2, 3 and 4 of BP) treated with SC/Ex. Thus the comparison between the average outcomes of the decision "treat everyone with SC/Ex" vs. the decision "treat everyone according to BP" is the difference between the average outcomes of subjects in levels 3 and 4 of BR treated with SC/Ex and the average outcomes of subjects in levels 3 and 4 of BR treated with STAIR/SC. In other words, we are looking at the outcomes only of subjects in levels 3 and 4 of BR (subjects who do better on STAIR/SC. The test is a test for comparison of the efficacies of STAIR/SC and SC/Ex among subjects in levels 3 and 4 of BR.