

Electronic Supplementary Material

Sex-Related Differences in Outcomes Following Psilocybin or Escitalopram: A Bayesian Secondary Analysis

Aline Frick, Grace Blest-Hopley, Manesh Girn, David Erritzoe, David Nutt, and Robin Carhart-Harris

Corresponding author: Dr Grace Blest-Hopley

King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom

Email: Grace.blest-hopley@kcl.ac.uk

Contents: Supplementary Tables S1–S4 and Supplementary Figures S1–S2

Supplementary Table S1. Descriptive Statistics by treatment condition and sex

Measure		Escitalopram		Psilocybin	
Outcome	Assessment	Female	Male	Female	Male
Age (years); n; mean (SD)	Baseline	9; 37.56 (10.43)	20; 39.75 (9.61)	11; 38.36 (8.51)	19; 46.16 (12.56)
STAI (n); mean(SD)	Baseline	9; 67.89 (7.01)	20; 64.60 (5.92)	11; 66.73 (4.08)	19; 62.89 (7.45)
	Week 6	9; 57.00 (14.28)	20; 56.50 (10.00)	11; 40.64 (12.88)	19; 50.95 (10.99)
	Change	9; -10.89 (11.55)	20; -8.10 (9.28)	11; -26.09 (15.52)	19; -11.95 (11.37)
SHAPS (n); mean(SD)	Baseline	9; 5.00 (2.55)	20; 6.45 (3.66)	11; 6.64 (3.75)	19; 6.21(4.33)
	Week 6	9; 1.33 (2.18)	20; 4.55 (4.44)	11; 1.09 (2.70)	19; 1.84 (3.47)
	Change	9; -3.67 (3.12)	20; -1.90 (3.29)	11; -5.55 (3.80)	19; -4.37 (3.77)
MADRS (n); mean(SD)	Baseline	9; 26.56 (6.02)	20; 27.00 (5.04)	11; 27.36 (4.43)	17; 27.71 (3.42)
	Week 6	9; 20.78 (9.27)	20; 19.40 (8.39)	11; 11.18 (10.13)	17; 14.12 (10.14)
	Change	9; -5.78 (11.86)	20; -7.60 (8.03)	11; -16.18 (9.37)	17; -13.59 (9.49)
QIDS-SR-16 (n); mean (SD)	Baseline	9; 17.78 (3.23)	20; 15.85 (4.42)	11; 14.91 (3.91)	19; 14.32 (3.96)
	Week 6	9; 10.44 (5.59)	20; 9.90 (5.98)	11; 5.73 (6.05)	19; 7.53 (5.93)
	Change	9; -7.33 (6.12)	20; -5.95 (5.37)	11; -9.18 (5.40)	19; -6.79 (6.52)
BDI-1A (n); mean(SD)	Baseline	9; 30.67 (6.54)	20; 27.80 (7.15)	11; 30.73 (6.60)	19; 28.11 (6.84)
	Week 6	9; 17.89 (12.77)	20; 17.85 (10.61)	11; 7.82 (11.16)	19; 12.32 (10.93)

	Change	9; -12.78 (11.44)	20; -9.95 (9.20)	11; -22.91 (11.17)	19; -15.79 (12.39)
WEMWBS (n); mean(SD)	Baseline	9; 31.11 (6.79)	19; 30.53 (5.52)	10; 28.10 (5.47)	18; 28.83 (6.32)
	Week 6	9; 40.00 (10.62)	19; 36.05 (9.06)	10; 46.90 (10.46)	18; 43.50 (12.44)
	Change	9; 8.89 (10.31)	19; 5.53 (8.69)	10; 18.80 (9.35)	18; 14.67 (13.61)
WBSI (n); mean(SD)	Baseline	9; 55.00 (5.27)	19; 54.58 (6.53)	10; 52.70 (6.07)	18; 54.83 (7.06)
	Week 6	9; 51.44 (10.14)	19; 52.05 (7.43)	10; 42.00 (17.23)	18; 45.89 (12.71)
	Change	9; -3.56 (7.50)	19; -2.53 (9.02)	10; -10.70 (14.80)	18; -8.94 (12.70)

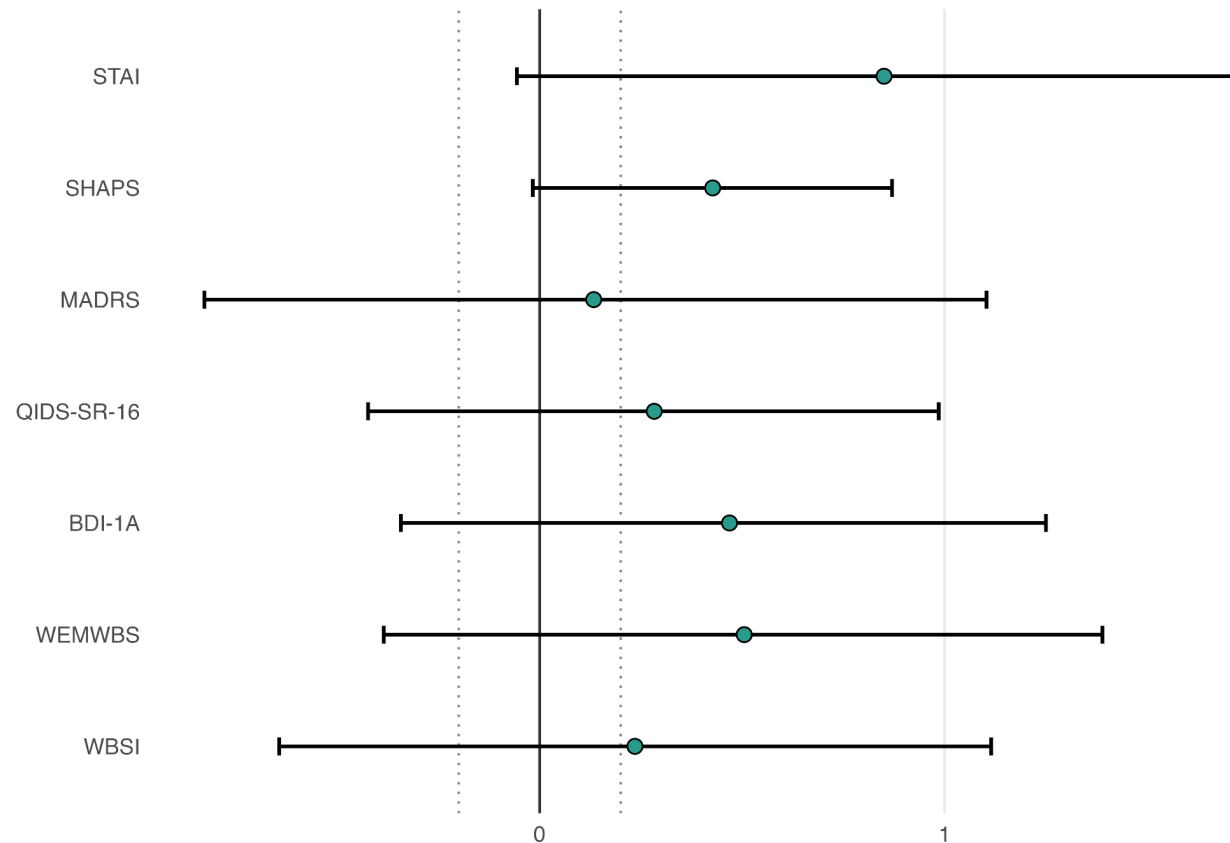
Note: Change was calculated as Week 6 minus baseline; negative change indicates improvement for STAI, SHAPS, MADRS, QIDS-SR-16, BDI-1A, and WBSI, whereas positive change indicates improvement for WEMWBS. MADRS included 57 complete cases; WBSI and WEMWBS each included 56 complete cases. All other continuous outcomes included 59 participants. PRSexDQ-SALSEX was administered only at Week 6.

Supplementary Table S2. Baseline score variability by sex

Scale	F, n; mean (SD)	M, n; mean (SD)	SD ratio (F/M)	Brown–Forsythe F (df ₁ , df ₂)	<i>p</i>	Holm- <i>p</i>
STAI	20; 67.25 (5.46)	39; 63.77 (6.67)	0.82	0.81 (1, 57)	.373	1.000
SHAPS	20; 5.90 (3.29)	39; 6.33 (3.95)	0.83	1.24 (1, 57)	.270	1.000
MADRS	20; 27.00 (5.08)	37; 27.32 (4.33)	1.17	0.24 (1, 55)	.623	1.000
QIDS-SR-16	20; 16.20 (3.82)	39; 15.10 (4.22)	0.91	0.40 (1, 57)	.529	1.000
BDI-1A	20; 30.70 (6.40)	39; 27.95 (6.91)	0.93	0.14 (1, 57)	.710	1.000
WEMWBS	19; 29.53 (6.15)	37; 29.70 (5.90)	1.04	0.11 (1, 54)	.744	1.000
WBSI	19; 53.79 (5.67)	37; 54.70 (6.70)	0.85	0.44 (1, 54)	.511	1.000

Note: SD = standard deviation; F = Brown–Forsythe test statistic; df = degrees of freedom. Brown–Forsythe tests compare baseline score variability between females and males across treatment conditions. The female-to-male SD ratio describes the direction and magnitude of variability differences; values greater than 1 indicate greater variability among females, and values below 1 indicate greater variability among males. Holm-adjusted *p* values account for the seven baseline variance comparisons. WBSI and WEMWBS included 19 females and 37 males because three participants had missing baseline scores; MADRS included 20 females and 37 males because Week 6 scores were missing for two male participants. All other outcomes included 20 females and 39 males.

Supplementary Figure S1. Overall adjusted post-treatment sex differences across treatment conditions



Note: Points represent posterior means, and horizontal intervals represent 95% credible intervals for the overall female–male difference at Week 6, adjusted for baseline score and randomized treatment condition. Estimates were standardized using the pooled baseline SD and direction-aligned so that positive values consistently indicate a more favorable adjusted outcome among females. The solid vertical line represents no sex difference, and the dotted vertical lines indicate the boundaries of the region of practical equivalence (-0.20 to $+0.20$ baseline SD).

Supplementary Table S3. Uniform magnitude benchmark and prior specifications by outcome

Outcome	Pooled baseline SD	Magnitude benchmark δ, raw points	Magnitude benchmark δ, standardized	Narrow prior SD, raw equivalent	Primary prior SD, raw equivalent	Wide prior SD, raw equivalent	ROPE half-width, raw equivalent
STAI	6.46	3.23	0.50	3.23	6.46	9.69	1.29
SHAPS	3.72	1.86	0.50	1.86	3.72	5.57	0.74
MADRS	4.56	2.28	0.50	2.28	4.56	6.84	0.91
QIDS-SR-16	4.09	2.04	0.50	2.04	4.09	6.13	0.82
BDI-1A	6.81	3.41	0.50	3.41	6.81	10.22	1.36
WEMWBS	5.93	2.97	0.50	2.97	5.93	8.90	1.19
WBSI	6.33	3.17	0.50	3.17	6.33	9.50	1.27

Note: δ denotes the magnitude benchmark used to calculate BF benchmark. A single distribution-based benchmark of 0.5 pooled baseline SD was applied identically to all seven continuous outcomes; the raw-score equivalent shown for each outcome is 0.5 times its pooled baseline SD from the corresponding complete-case sample. The benchmark is a group-level magnitude criterion and was not interpreted as a validated minimal clinically important difference. The primary female–male contrast prior was normal(0,1) on the standardized outcome scale. The narrower and wider sensitivity priors were Normal(0, 0.5) and Normal(0, 1.5), respectively. The raw-score equivalents shown in the table were obtained by multiplying the standardized prior SD by the corresponding pooled baseline SD; the models themselves were fitted on the standardized scale. The ROPE was defined as ± 0.20 standardized units, with the raw-score half-width calculated as $0.20 \times \text{SD baseline}$.

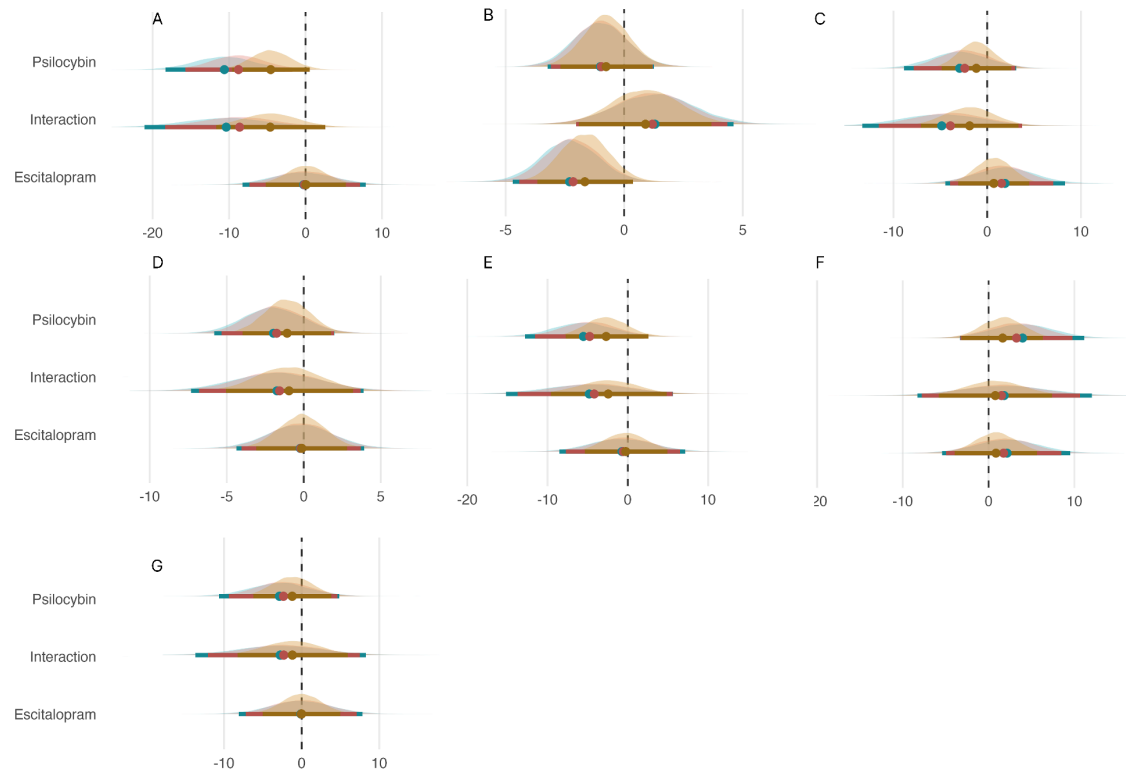
Supplementary Table S4. Sensitivity Analysis under alternative prior specifications

Outcome	Prior	Escitalopram				Psilocybin				Interaction			
		Estimate (95% CrI)	BFany	BFbenchmark	BFROPE	Estimate (95% CrI)	BFany	BFbenchmark	BFROPE	Estimate (95% CrI)	BFany	BFbenchmark	BFROPE
STAI	Narrow: N(0, 0.5)	0.06 (−5.20 to 5.26)	0.82	0.61	1.36	−4.56 (−9.59 to 0.55)	3.41 ^a	5.02 ^a	0.23 ^b	−4.62 (−11.69 to 2.58)	1.80	2.13	0.50
	Primary: N(0, 1)	−0.15 (−7.34 to 7.13)	0.57	0.38	1.96	−8.77 (−15.73 to −1.73)	11.18 ^a	9.46 ^a	0.09 ^b	−8.62 (−18.36 to 1.30)	2.55	2.57	0.38
	Wide: N(0, 1.5)	−0.24 (−8.24 to 7.87)	0.43	0.28 ^b	2.64	−10.62 (−18.32 to −2.71)	14.61 ^a	10.36 ^a	0.07 ^b	−10.38 (−21.06 to 0.41)	2.43	2.28	0.40
SHAPS	Narrow: N(0, 0.5)	−1.66 (−3.66 to 0.37)	2.08	1.56	0.48	−0.76 (−2.69 to 1.18)	0.72	0.35	1.65	0.90 (−1.91 to 3.68)	0.64	0.42	1.74
	Primary: N(0, 1)	−2.15 (−4.43 to 0.14)	1.73	0.95	0.62	−0.96 (−3.09 to 1.21)	0.44	0.17 ^b	2.96	1.19 (−2.02 to 4.36)	0.40	0.22 ^b	2.92
	Wide: N(0, 1.5)	−2.30 (−4.70 to 0.13)	1.40	0.65	0.82	−1.00 (−3.23 to 1.26)	0.31 ^b	0.11 ^b	4.28 ^a	1.30 (−1.98 to 4.61)	0.28 ^b	0.15 ^b	4.28 ^a
MADRS	Narrow: N(0, 0.5)	0.71 (−3.09 to 4.49)	0.91	0.76	1.13	−1.16 (−4.86 to 2.57)	1.01	0.99	0.98	−1.87 (−7.08 to 3.35)	1.06	1.06	0.94
	Primary: N(0, 1)	1.53 (−3.95 to 7.07)	0.72	0.58	1.50	−2.39 (−7.85 to 3.01)	0.88	0.79	1.20	−3.92 (−11.57 to 3.72)	1.01	0.98	0.97
	Wide: N(0, 1.5)	1.91 (−4.47 to 8.30)	0.56	0.44	1.83	−2.94 (−8.87 to 3.11)	0.71	0.63	1.37	−4.85 (−13.34 to 3.60)	0.84	0.76	1.24
QIDS-SR-16	Narrow: N(0, 0.5)	−0.12 (−3.08 to 2.81)	0.74	0.44	1.59	−1.08 (−3.97 to 1.77)	0.94	0.80	1.10	−0.96 (−5.04 to 3.20)	0.84	0.66	1.31
	Primary: N(0, 1)	−0.19 (−4.04 to 3.71)	0.49	0.27 ^b	2.46	−1.76 (−5.33 to 1.96)	0.73	0.54	1.49	−1.57 (−6.80 to 3.70)	0.55	0.41	1.99
	Wide: N(0, 1.5)	−0.23 (−4.37 to 3.92)	0.35	0.18 ^b	3.48 ^a	−1.97 (−5.81 to 1.98)	0.54	0.38	2.01	−1.74 (−7.32 to 3.88)	0.39	0.27 ^b	2.79
BDI-1A	Narrow: N(0, 0.5)	−0.26 (−5.36 to 4.94)	0.79	0.53	1.44	−2.70 (−7.74 to 2.56)	1.33	1.50	0.70	−2.44 (−9.58 to 4.84)	0.95	0.88	1.07

	Primary: N(0, 1)	−0.57 (−7.71 to 6.50)	0.54	0.33	2.19	−4.76 (−11.56 to 2.06)	1.31	1.20	0.74	−4.19 (−13.73 to 5.59)	0.73	0.63	1.38
	Wide: N(0, 1.5)	−0.73 (−8.52 to 7.15)	0.40	0.24 ^b	2.85	−5.54 (−12.82 to 1.84)	1.13	0.98	0.88	−4.81 (−15.17 to 5.62)	0.55	0.46	1.86
WEMWBS	Narrow: N(0, 0.5)	0.86 (−3.95 to 5.65)	0.87	0.71	1.23	1.64 (−3.15 to 6.33)	1.03	1.03	0.93	0.78 (−5.82 to 7.40)	0.83	0.68	1.26
	Primary: N(0, 1)	1.75 (−4.96 to 8.48)	0.65	0.50	1.63	3.25 (−3.29 to 9.79)	0.90	0.79	1.11	1.50 (−7.79 to 10.67)	0.60	0.46	1.84
	Wide: N(0, 1.5)	2.15 (−5.42 to 9.53)	0.51	0.36	2.18	3.96 (−3.33 to 11.15)	0.75	0.64	1.34	1.81 (−8.30 to 12.05)	0.44	0.33 ^b	2.51
WBSI	Narrow: N(0, 0.5)	−0.02 (−5.02 to 4.96)	0.81	0.56	1.39	−1.21 (−6.22 to 3.81)	0.91	0.81	1.14	−1.19 (−8.26 to 5.93)	0.85	0.73	1.22
	Primary: N(0, 1)	−0.04 (−7.20 to 7.09)	0.58	0.39	1.93	−2.35 (−9.40 to 4.60)	0.69	0.54	1.56	−2.31 (−12.08 to 7.49)	0.63	0.50	1.66
	Wide: N(0, 1.5)	−0.09 (−8.08 to 7.83)	0.42	0.26 ^b	2.72	−2.86 (−10.64 to 4.84)	0.53	0.41	1.92	−2.76 (−13.69 to 8.27)	0.47	0.36	2.30

Note: Estimates are adjusted female–male differences in original outcome units and are presented as posterior means (95% credible intervals). The interaction is the female–male difference under psilocybin minus the corresponding difference under escitalopram. Negative values indicate lower scores among females; this is favorable for symptom measures, whereas positive values are favorable for WEMWBS. "Any" denotes BFany, which compares a non-zero contrast with the point-null hypothesis. "Benchmark" denotes "BFbenchmark," which compares an absolute standardized contrast reaching or exceeding the uniform distribution-based magnitude benchmark, $|\Delta z| \geq 0.50$, with a contrast below that benchmark, $|\Delta z| < 0.50$. The 0.50-SD benchmark was applied uniformly to all seven continuous outcomes and was not interpreted as a validated minimal clinically important difference. "ROPE" denotes BFROPE, which compares a contrast within the region of practical equivalence, $|\Delta z| \leq 0.20$, with a contrast outside this region. All Bayes factors are oriented toward the hypothesis named by the column. ^a indicates evidence supporting the named hypothesis ($BF > 3$), ^b indicates evidence supporting the complementary hypothesis ($BF < 1/3$). Values between 1/3 and 3 were considered inconclusive. Classifications were determined using the unrounded Bayes-factor values.

Supplementary Figure S2. Prior-sensitivity analysis of treatment-specific sex contrasts



Note: Posterior distributions are shown for the adjusted female–male difference under escitalopram, the adjusted female–male difference under psilocybin, and the sex-by-treatment interaction under three prior specifications. Panels A–G correspond to STAI (A), SHAPS (B), MADRS (C), QIDS-SR-16 (D), BDI-1A (E), WEMWBS (F), and WBSI (G). Points indicate posterior means and horizontal intervals indicate 95% credible intervals. The vertical dashed line represents no difference. Models were fitted using questionnaire scores standardized by the corresponding pooled baseline mean and SD; estimates in the figure were subsequently converted back to the original questionnaire units. The skeptical $\text{Normal}(0, 0.5)$ prior imposed stronger regularization toward zero than the primary $\text{Normal}(0, 1)$ prior, whereas the wider $\text{Normal}(0, 1.5)$ prior allowed greater prior probability for large sex differences. The interaction was calculated as the female–male contrast under psilocybin minus the corresponding contrast under escitalopram. Negative values indicate more favourable outcomes among females for symptom measures, whereas positive values favour females for WEMWBS. For the interaction, values in the favourable direction indicate that the female–male contrast was more favourable under psilocybin than under escitalopram.