PSILOGOS
Research Article
open-access

Clinical Superiority or Evidential Mirage? Threshold-Calibrated Decision Stability of Psychological Interventions for Excessive Gaming

Arjen Robben

ABSTRACT

Comparative syntheses of psychological treatments for excessive gaming commonly present a treatment hierarchy, yet a high rank does not itself establish that the difference is large enough to matter or sufficiently supported to guide care. This study examined whether distinctions among eight psychological treatment classes remain decision-stable after imposing clinically relevant superiority thresholds and accounting for replication, enrolment, randomization, comparator stringency, and follow-up coverage. Seventeen controlled studies involving 745 participants contributed eight active treatment classes and 28 active-treatment contrasts. Comparator-stratified standardized mean changes and the complete contrast matrix supplied the quantitative record. A threshold-calibrated contrast reliability analysis converted each contrast and its 95% confidence interval into the probability that one treatment exceeded another by $\delta = 0.0$, 0.3, 0.5, or 0.8 standardized units. A tournament simulation drew 200,000 realizations of all 28 contrasts and ranked treatments by the number of threshold-exceeding wins. A five-component evidence-support coefficient and a 50,000-specification weight perturbation experiment examined whether efficacy order and evidential strength converged. Psychological treatment effects were larger against inactive controls ($g = 1.70$, 95% CI 1.27–2.12; $I^2 = 69\%$) than against active controls ($g = 0.88$, 95% CI 0.21–1.56; $I^2 = 72\%$). At $\delta = 0.5$, cognitive behavioural therapy (CBT) plus mindfulness and CBT plus family treatment had probabilities of ranking first of 0.675 and 0.322, respectively. Their direct contrast provided only a 0.372 probability that CBT plus mindfulness exceeded CBT plus family treatment by at least 0.5 units. Mindfulness had a 0.939 probability of exceeding CBT by that amount, whereas CBT had only a 0.143 probability of exceeding behavioural treatment plus mindfulness. Treatment rank and evidence support were weakly associated (Spearman $\rho = 0.119$, $p = 0.779$). The evidence distinguishes a leading cluster of CBT plus mindfulness, CBT plus family treatment, and mindfulness from several lower-ranked options, but it does not establish a single clinically dominant treatment within that cluster. Treatment selection should therefore reflect patient context and delivery feasibility while confirmatory trials directly compare the leading approaches.

Keywords: gaming disorder; psychological treatment; decision stability; clinical relevance; network meta-analysis; rank uncertainty; cognitive behavioural therapy; mindfulness

Department of Psychiatry, Leiden University Medical Center, Leiden University, Leiden, Netherlands
arj.robben21@gmail.com
– **Published:** 5 November 2024

1. INTRODUCTION

Excessive gaming becomes clinically important when persistent play is accompanied by impaired control, increasing priority assigned to gaming, and continuation despite adverse consequences. The inclusion of gaming disorder in the eleventh revision of the International Classification of Diseases supplied a public-health definition centred on functional impairment rather than on gaming time alone.[1,2] Internet gaming disorder remains a condition for further study in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, reflecting continuing uncertainty about boundaries, assessment, and natural history.[3] These nosological positions are compatible with the observation that intense recreational play is not necessarily pathological. A treatment claim must therefore concern relief of persistent impairment, loss of control, distress, or related symptoms rather than the simple suppression of a popular leisure activity.

Diagnostic convergence has practical consequences for intervention research. International assessment work has emphasized endorsement of the diagnostic features together with clinically significant distress or impairment, while warning against classifying highly engaged but well-functioning players as disordered.[41] The ICD-11 delineation similarly connects diagnosis to management and prevention by requiring a persistent pattern with material consequences.[42] In children and adolescents, developmental context, parental regulation, school demands, and comorbid symptoms can change both presentation and treatment access.[43] Eligibility criteria and outcomes should therefore distinguish symptom endorsement from the functional change that patients, families, and clinicians seek.

The clinical need is substantial but heterogeneous. A synthesis of population studies estimated a pooled prevalence near 3%, although estimates varied markedly with sampling and measurement decisions.[4] Harm may involve disrupted sleep, deteriorating school or occupational performance, family conflict, social withdrawal, financial expenditure, and psychiatric comorbidity.[5,6] Depression, anxiety, attention-deficit/hyperactivity disorder, and other addictive behaviours occur frequently enough to influence both treatment selection and outcome interpretation.[7,8] Cross-sectional associations do not prove that gaming is the primary cause of these difficulties, but they do show why a narrow count of gaming symptoms may be inadequate for evaluating recovery.

Treatment provision has developed faster than consensus about comparative efficacy. Clinicians have used cognitive behavioural therapy (CBT), behavioural modification, motivational interviewing, mindfulness training, family treatment, virtual-reality procedures, and combinations of these approaches. CBT addresses expectancies, triggers, avoidance, reinforcement, and maladaptive routines; mindfulness methods cultivate awareness of urges and affective states without automatic enactment; family work targets communication, boundaries, conflict, and the home contingencies surrounding adolescent play.[9–11] Behavioural procedures may alter cue approach, reward sensitivity, response inhibition, or access patterns.[12,13] These treatments differ not only in content but also in dose, developmental target, professional requirements, comparator conditions, and the outcomes used to judge success.

The controlled evidence illustrates this variation. A multicentre randomized trial examined a manualized short-term treatment for Internet and computer-game addiction, whereas a specialized psychotherapy program for adolescents combined individual work with family participation over a longer course.[26,27] Acceptance and cognitive restructuring were combined for Asian adolescents, and a self-control program targeted gaming symptoms, aggression, and impulsivity in university students.[31,34] Earlier CBT work emphasized self-esteem and adolescent adjustment, showing that the CBT label has encompassed materially different therapeutic targets.[32] A virtual-reality program introduced another delivery mode rather than merely another theoretical orientation.[30] Pooling across these interventions is statistically possible, but clinical interpretation requires attention to what participants actually received.

Behavioural treatment plus mindfulness also spans more than one therapeutic ingredient. Craving-focused interventions have examined longitudinal symptom change, cue-induced neural responses, and intrinsic connectivity associated with treatment outcome.[28,29,33] These studies offer converging process hypotheses, yet repeated publications, small cohorts, and related recruitment streams do not supply the same evidential independence as replications conducted by separate teams. Escapism and coping may further moderate negative gaming outcomes, which supports measuring the psychological function of play rather than assuming that all excessive gaming follows the same maintaining process.[40] Consequently, a treatment class can be mechanistically coherent while its comparative estimate remains uncertain.

The comparative literature contains encouraging effects alongside recurring methodological limitations. Earlier reviews noted few randomized trials, inconsistent diagnostic thresholds, incomplete treatment descriptions, and limited follow-up.[14,15] Controlled trials range from brief bias-modification tasks to multi-session psychotherapy, and many rely on self-report instruments that are not interchangeable. Although several gaming-disorder scales show acceptable reliability, variation in item content and cut-off selection can alter who enters a trial and what magnitude of change is observed.[16,17] Blinding participants to psychotherapy is rarely feasible, making allocation procedures, assessor masking, treatment fidelity, attrition handling, and comparator credibility especially consequential.

Meta-analysis can pool this diverse evidence, but its conventional outputs answer only part of the clinical question. A standardized mean difference estimates the magnitude of a contrast on a common scale. Network meta-analysis additionally combines direct and indirect evidence to compare multiple interventions and can summarize the relative order through ranking probabilities or the surface under the cumulative ranking curve.[18,19] These quantities are useful when the network is connected and assumptions of transitivity and consistency are credible. They can nevertheless invite overinterpretation. A treatment may rank first because a single small study produced a large imprecise estimate, while a somewhat lower-ranked treatment may be supported by several trials, active comparators, and longer observation. Rank summaries also do not state whether the separation between adjacent treatments exceeds a clinically relevant magnitude.[20,21]

Reporting standards for treatment networks require transparent eligibility decisions, presentation of network geometry, assessment of inconsistency, and separation of direct from indirect evidence.[35] Global agreement does not guarantee that each sparse comparison is well determined; inconsistency tests can have low power, and treatment rankings remain conditional on the exchangeability of studies across comparisons.[37] Graphical tools can make network structure visible, but visual connectivity must not be mistaken for evidential density.[38] The same interpretive issue arises in larger psychiatric networks: rankings are most credible when accompanied by contrast uncertainty, heterogeneity, and confidence assessments.[39] Because I^2 quantifies the proportion of observed variation attributable to between-study heterogeneity rather than sampling error, substantial values additionally caution against presenting a pooled mean as a uniform treatment response.[36]

Clinical decisions require at least three distinctions. Statistical superiority asks whether the contrast is probably greater than zero. Clinically important superiority asks whether the contrast probably exceeds a non-zero threshold. Decision stability asks whether the preferred option remains preferred under reasonable changes in the threshold and the evidential emphasis. Conflating these distinctions produces brittle treatment recommendations. A narrow confidence interval just above zero may not imply a meaningful advantage, while a wide interval around a large point estimate may be too uncertain to justify a categorical hierarchy. The problem is amplified in sparse treatment networks because the ranking of a singleton node can be driven by one comparison.

The clinical question was: *which distinctions among psychological treatments for excessive gaming remain stable after efficacy contrasts are evaluated against non-zero clinical thresholds and the treatment order is examined against the amount and quality of supporting evidence?* Four threshold levels were evaluated, the complete set of active-treatment contrasts was transformed into probabilities of clinically important superiority, and repeated computational experiments quantified first-rank probability, expected rank, and sensitivity to the weighting of evidential properties. The expected pattern was a stable separation between a leading group and lower-ranked approaches but an uncertain order within the leading group.

2. METHODS

This aggregate-data decision study used the intervention taxonomy, study descriptors, comparator-stratified pooled estimates, rank summaries, and league-table contrasts reported for 17 controlled psychological-treatment studies involving 745 participants.[22] The clinical corpus contained eight active classes: CBT plus mindfulness, CBT plus family treatment, mindfulness, CBT, behavioural treatment plus mindfulness, virtual-reality behavioural treatment, behavioural treatment, and motivational interviewing plus behavioural treatment. Control conditions were retained for the comparator-stratified synthesis but excluded from the active-treatment decision tournament because no treatment and waiting-list conditions are not clinical alternatives.

The unit of the decision analysis was the active-treatment contrast. Eight treatments yield $8(8-1)/2 = 28$ unique contrasts. Each contrast supplied a point estimate and 95% confidence interval. Study descriptors comprised treatment class, displayed enrolment, randomized or non-randomized allocation, active or inactive comparator, and presence of post-treatment follow-up. The study rows totalled 715 participants, whereas the synthesis total was 745. Enrolment was therefore used only as a treatment-level information indicator derived from the study rows; 745 was used to describe the complete corpus. No individual participant records were used.

2.1. Comparator-stratified treatment effects

The pooled standardized mean changes against active and inactive controls were retained as complementary estimates of absolute treatment benefit. The active-comparator pool included psychological treatment or credible training conditions, while the inactive pool included no-treatment and waiting-list controls. Hedges' g was interpreted with its confidence interval and heterogeneity rather than by a rigid small–medium–large vocabulary. The difference between comparator strata was treated as clinically informative because inactive controls do not balance attention, expectancy, therapeutic contact, or spontaneous improvement to the same degree as active controls.

2.2. Threshold-calibrated contrast reliability analysis

For contrast j comparing treatments a and b, the network estimate d_j was assumed to be approximately normal. Its standard error was recovered from the 95% confidence limits (L_j, U_j) as

$$\mathrm{SE}_j = \frac{U_j - L_j}{2 \times 1.96}. \tag{1}$$

For a clinical superiority threshold $\delta \geq 0$, the probability that a exceeded b by at least δ was

$$P_{a>b}(\delta) = 1 - \Phi\left(\frac{\delta - d_j}{\mathrm{SE}_j}\right), \tag{2}$$

where Φ is the standard normal cumulative distribution function. The probability that b exceeded a by the same amount was $\Phi[(-\delta - d_j)/\,\mathrm{SE}_j]$; the remaining probability represented a difference lying between $-\delta$ and $+\delta$. Values of $\delta = 0.0$, 0.3, 0.5, and 0.8 standardized units were selected to span any statistical advantage, a modest advantage, a moderate advantage, and a large advantage. The thresholds were not declared universal clinical cutoffs. They were decision tolerances used to expose how strongly the hierarchy depended on the size of advantage demanded.

A treatment-level dominance index summarized the mean probability of beating each of the seven alternatives by at least δ:

$$D_a(\delta) = \frac{1}{7}\sum_{b \neq a} P_{a>b}(\delta). \tag{3}$$

This index ranges from 0 to 1. It differs from a rank curve because it retains the magnitude threshold and permits an indeterminate region in which neither treatment achieves a clinically important advantage.

2.3. Evidence-support coefficient

Five bounded components characterized the support behind treatment a. Replication was $R_a = 1 - \exp(-k_a/2)$, where k_a is the number of contributing studies. Information size was $N_a = 1 - \exp(-n_a/150)$, where n_a is displayed enrolment for that treatment class. Randomization, active-comparator, and follow-up components used Jeffreys-type smoothing: $Q_a = (x_a + 0.5)/(k_a + 1)$, where x_a is the number of studies possessing the relevant property. The treatment-level coefficient was the geometric mean

$$C_a = (R_a N_a Q_{a,\mathrm{RCT}} Q_{a,\mathrm{active}} Q_{a,\mathrm{follow}})^{1/5}. \tag{4}$$

The geometric mean penalizes a treatment for a very weak component that would be obscured by simple addition. The coefficient is descriptive and is not an alternative effect size. To examine how evidential support changes contrast certainty while preserving a coherent probability distribution, the contrast mean was contracted toward zero in proportion to joint treatment support:

$$d^*_{ab} = \sqrt{C_a C_b}\, d_{ab}, \qquad P^*_{a>b}(\delta) = 1 - \Phi\left(\frac{\delta - d^*_{ab}}{\mathrm{SE}_{ab}}\right). \tag{5}$$

The reverse-treatment probability and the indifference probability were derived from the same $N(d^*_{ab}, \mathrm{SE}^2_{ab})$ distribution, ensuring that the three mutually exclusive outcomes summed to one. This contraction does not claim that low-support treatments are ineffective; it prevents a large contrast with limited support from being treated as fully established.

2.4. Tournament simulation

The 28 active-treatment contrasts were sampled 200,000 times from normal distributions defined by their point estimates and recovered standard errors. During each draw, treatment a received one decisive win over b when the sampled contrast exceeded $+\delta$, treatment b received one decisive win when it was below $-\delta$, and neither received a win when the draw lay within the indifference interval. Treatments were ordered by their number of decisive wins; total sampled contrast margin resolved equal win counts. The simulations yielded the probability of ranking first, expected rank, and 2.5th–97.5th percentile rank interval. Contrast covariance was unavailable, so independent sampling was used. The resulting probabilities are consequently decision diagnostics rather than a substitute for a one-stage network model.

2.5. Rank entropy and precision demand

First-rank concentration was quantified with normalized Shannon entropy,

$$H = -\frac{\sum_a p_a \log(p_a)}{\log(8)}, \tag{6}$$

where p_a is the simulated first-rank probability. The effective number of first-rank contenders was $\exp[-\sum_a p_a \log(p_a)]$. A value near one indicates that the decision is dispersed across many treatments, whereas a value near zero indicates concentration on a small set. To translate unresolved contrasts into trial-planning information, a precision-demand calculation estimated the information increase required for the 95% confidence interval to lie entirely above $\delta = 0.5$, assuming the observed contrast remained unchanged. For $d_j > \delta$, the target standard error was $(d_j - \delta)/1.96$, and the required information multiplier was the squared ratio of current to target standard error. No finite multiplier was assigned when the observed contrast itself was at or below 0.5, because additional precision around the unchanged point estimate could not establish a 0.5-unit advantage.

2.6. Weight-perturbation experiment

The association between treatment order and evidence support was first summarized by Spearman correlation. A separate 50,000-specification experiment varied the relative importance of efficacy and evidence properties. Each specification drew five evidence-component weights from a uniform Dirichlet distribution and an efficacy weight λ from $\mathrm{Uniform}(0.35, 0.80)$. Treatment scores combined the logit of the rank summary with the

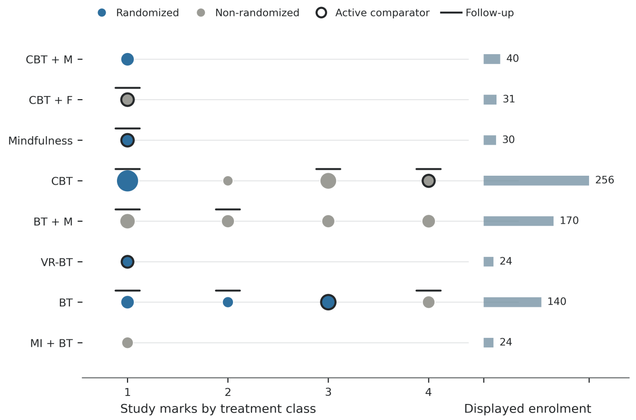

Figure 1. Study distribution across active treatments.

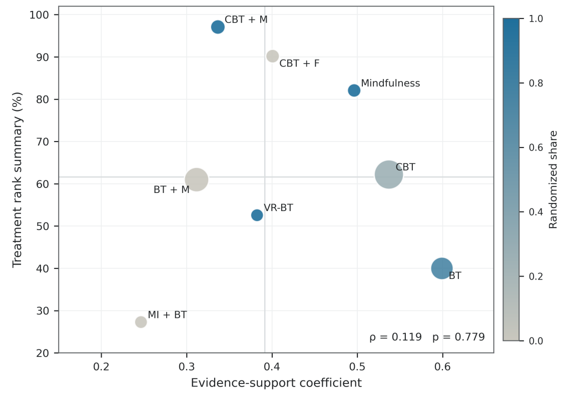

Figure 2. Treatment position versus evidence support.

weighted logits of the five evidence components. First-place frequency and central rank intervals across specifications quantified whether a treatment's position required a narrow choice of weights.

3. RESULTS

3.1. Composition of the evidence

The evidence base was concentrated in three treatment classes. CBT, behavioural treatment, and behavioural treatment plus mindfulness each contributed four studies; every other active class contributed one. Seven of 17 studies were randomized, five used an active comparator, and follow-up beyond immediate post-treatment was uneven. Fourteen studies were conducted in Eastern regions and three in Western regions. Eleven different gaming-related self-report instruments were represented. This distribution created an important asymmetry: some of the highest-ranked combinations were supported by only one trial, whereas CBT and behavioural approaches had more replication but lower rank summaries.

The study ledger in figure 1 makes this imbalance visible at the level of individual studies. Large randomized marks are concentrated in CBT and behavioural treatment, whereas each combined CBT option is represented by one mark. Active comparisons occur for mindfulness, CBT plus family treatment, virtual-reality behavioural treatment, and one study each of CBT and behavioural treatment. Follow-up is common within CBT and behavioural treatment but absent for CBT plus mindfulness. These design features explain why apparent efficacy and evidential support can occupy different positions.

The coefficient in table 1 did not track the efficacy order closely. CBT and behavioural treatment had the largest coefficients (both 0.589) because they combined replication with larger enrolment and some follow-up, while CBT plus mindfulness had a coefficient of 0.357 despite its high rank summary. Mindfulness achieved 0.553 because its sole study was randomized, used CBT as an active comparator, and included follow-up. Across eight treatments, rank summary and evidence support had Spearman $\rho = 0.119$ ($p = 0.779$). The absence of alignment means that reported efficacy position cannot serve as a proxy for evidential maturity.

The treatment-level pattern also shows why raw study count is insufficient. Behavioural treatment plus mindfulness had four records and 170 displayed participants, yet no contributing comparison was randomized or active-controlled, producing a coefficient of 0.365. Mindfulness had only one study and 30 displayed participants, but the randomized active comparison and follow-up increased its coefficient to 0.553. CBT plus family treatment occupied an intermediate position because active comparison and follow-up were offset by non-randomized allocation. The geometric aggregation therefore distinguished replication from design rather than rewarding either property alone.

The two-dimensional map in figure 2 confirms that a treatment's location on the efficacy axis does not predict its location on the support axis. CBT plus mindfulness lies highest on efficacy but left of the median support line. CBT and behavioural treatment occupy the rightmost support positions while lying in the middle and lower portions of the efficacy axis. Mindfulness is the only treatment in the upper-right quadrant, but its circle remains small because the position rests on one study. Circle size and colour therefore prevent a high coordinate from being read without its information and randomization context.

3.2. Comparator stringency

The pooled benefit against inactive controls was $g = 1.70$ (95% CI 1.27–2.12), nearly twice the active-comparator estimate of $g = 0.88$ (95% CI 0.21–1.56). Heterogeneity was substantial in both strata ($I^2 = 69\%$ and 72%, respectively). The confidence intervals indicate that psychological treatment was beneficial under both comparisons, but the active-control result is the more relevant estimate for choosing among plausible services. Its wider interval also shows that precision remains limited when nonspecific attention and credible intervention exposure are better balanced.

The density silhouettes in figure 3 place both estimates

Table 1. Treatment-level evidence support

Treatment	Studies	n	RCT	Active	Follow-up	Rank summary (%)	C_a
CBT + mindfulness	1	40	1	0	0	97.1	0.357
CBT + family	1	31	0	1	1	90.2	0.423
Mindfulness	1	30	1	1	1	82.1	0.553
CBT	4	256	1	1	3	62.2	0.589
BT + mindfulness	4	170	0	0	2	61.0	0.365
Virtual-reality BT	1	24	1	1	0	52.6	0.403
BT	4	140	3	1	3	40.0	0.589
MI + BT	1	24	0	0	0	27.3	0.215

n, enrolment displayed in the study-characteristics rows; RCT, number randomized; Active, number with active comparator; Follow-up, number reporting post-treatment observation; CBT, cognitive behavioural therapy; BT, behavioural treatment; MI, motivational interviewing; C_a, evidence-support coefficient.

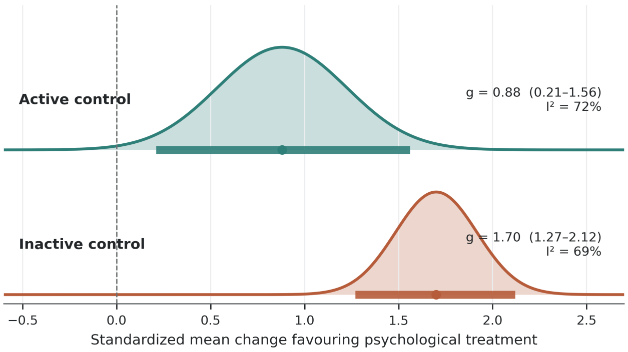

Figure 3. Treatment effects by comparator stringency.

on a common standardized scale. Neither 95% interval crosses zero, yet the active-control distribution is wider and centred markedly closer to zero. The overlap between the curves is compatible with genuine psychological benefit under both comparisons, while the displaced centres show that the expected effect depends materially on what the comparison group receives. Heterogeneity near 70% in each stratum indicates that comparator type alone does not explain the full variation among studies.

3.3. Clinically important pairwise distinctions

The three leading treatments were not equally distinguishable from one another. CBT plus mindfulness exceeded CBT plus family treatment by 0.23 units (95% CI -1.39 to 1.85), corresponding to a 0.610 probability of any advantage and only a 0.372 probability of an advantage greater than 0.5. The comparison of CBT plus mindfulness with mindfulness was larger at 1.11 (-0.47 to 2.68); its probabilities were 0.916 for any advantage and 0.776 for an advantage greater than 0.5. CBT plus family treatment exceeded mindfulness by 0.88 (-0.37 to 2.13), yielding corresponding probabilities of 0.916 and 0.724.

Separation from the middle of the hierarchy was more convincing. Mindfulness exceeded CBT by 1.16 (0.32–1.99), with a 0.939 probability of an advantage greater than 0.5. CBT plus family treatment exceeded CBT by 2.03 (1.11–2.96), and CBT plus mindfulness exceeded CBT by 2.26 (0.93–3.59); both probabilities above 0.5 were greater than 0.995. In contrast, CBT exceeded behavioural treatment plus mindfulness by only 0.17 (-0.43 to 0.78), producing a 0.143 probability that the difference was greater than 0.5. The numerical ordering of these two approaches therefore conveys more separation than the contrast supports.

The probabilities in table 2 separate evidence of direction from evidence of clinically important magnitude. Several contrasts with a high probability of being positive had a much lower probability of clearing 0.5. This distinction was most pronounced for CBT versus behavioural treatment plus mindfulness and for virtual-reality behavioural treatment versus behavioural treatment. A binary declaration based only on whether the point estimate was positive would conceal this decision uncertainty.

The direction–magnitude distinction was less consequential for comparisons between the leading cluster and the lowest treatments. CBT plus mindfulness, CBT plus family treatment, and mindfulness each had probabilities above 0.99 of exceeding behavioural treatment and motivational interviewing plus behavioural treatment by 0.5 in most corresponding contrasts. These stable separations account for the narrow central rank ranges of the three leaders. By comparison, estimates among CBT, behavioural treatment plus mindfulness, and virtual-reality behavioural treatment produced substantial indifference probability. The evidence thus forms zones of separation rather than eight individually resolved positions.

The full probability surface in figure 4 shows these zones without reducing them to selected comparisons. Dark blue cells across the first three rows mark high probabilities that the leading treatments exceed the middle and lower classes by 0.5. The 0.37 cell for CBT plus mindfulness over CBT plus family treatment and the 0.19 reverse cell leave considerable probability inside the indifference interval. The middle block is more heterogeneous: CBT has a 0.14 probability of exceeding behavioural treatment plus mindfulness by 0.5 but a 0.77 probability of

Table 2. Decision probabilities for selected treatment contrasts

Treatment A	Treatment B	Difference (95% CI)	$P(A > B)$	$P(A > B + 0.5)$
CBT + mindfulness	CBT + family	0.23 (-1.39, 1.85)	0.610	0.372
CBT + mindfulness	Mindfulness	1.11 (-0.47, 2.68)	0.916	0.776
CBT + family	Mindfulness	0.88 (-0.37, 2.13)	0.916	0.724
Mindfulness	CBT	1.16 (0.32, 1.99)	0.997	0.939
CBT	BT + mindfulness	0.17 (-0.43, 0.78)	0.709	0.143
CBT	BT	0.71 (0.14, 1.27)	0.993	0.767
BT + mindfulness	BT	0.53 (-0.06, 1.12)	0.961	0.540
VR-BT	BT	0.45 (-0.58, 1.49)	0.803	0.462

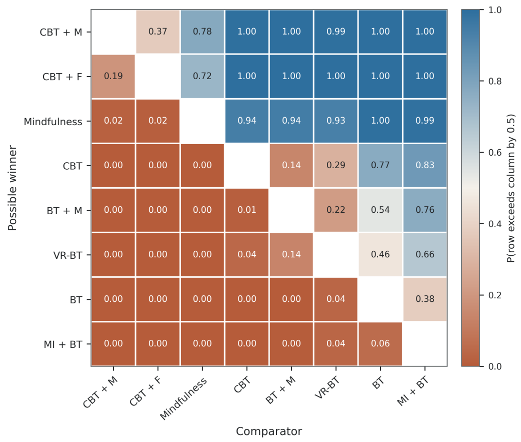

Figure 4. Probability of a 0.5-unit treatment advantage.

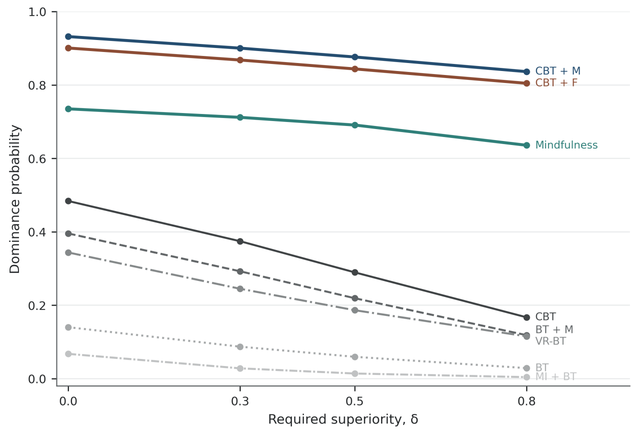

Figure 5. Dominance across superiority thresholds.

exceeding behavioural treatment by the same amount. The matrix therefore supports clustered, not fully ordinal, interpretation.

3.4. Treatment-level decision stability

At $\delta = 0$, CBT plus mindfulness had a dominance index of 0.932 and CBT plus family treatment had 0.901. Requiring a 0.5-unit advantage reduced these values to 0.876 and 0.844. Mindfulness declined from 0.735 to 0.691, remaining clearly above the middle group. CBT declined more sharply from 0.484 to 0.290 because several of its advantages over adjacent treatments were smaller than 0.5. Behavioural treatment plus mindfulness, virtual-reality behavioural treatment, behavioural treatment, and motivational interviewing plus behavioural treatment had dominance indices of 0.219, 0.187, 0.060, and 0.014, respectively, at $\delta = 0.5$.

The trajectories in figure 5 retain the same three leading treatments throughout the full threshold range. Their gradual declines differ from the steep loss for CBT and the middle classes, showing that the leaders' advantages are distributed across several large contrasts. No trajectory crosses into the leading group as δ rises. The plot thus establishes stability of the broad treatment clusters while revealing sensitivity of the absolute dominance probability to the magnitude required.

The tournament in table 3 allocated nearly all first-place probability to the two CBT combinations. CBT plus mindfulness ranked first in 67.5% of draws and CBT plus family treatment in 32.2%; mindfulness ranked first in 0.4%. The result should not be read as a 67.5% probability that CBT plus mindfulness is the universally best clinical choice. It is the probability of winning the specified aggregate tournament under the normal approximation, independent contrast draws, and a 0.5-unit tolerance. The 1–2 rank intervals for both CBT combinations show that their internal order is unstable even though their separation from lower treatments is comparatively stable.

The rank intervals in figure 6 make the asymmetry of uncertainty explicit. The two largest circles occupy overlapping 1–2 intervals, so first-place probability is divided between them rather than spread across the treatment set. Mindfulness is narrowly confined to ranks 2–3 but has little first-place mass. By contrast, CBT, behavioural treatment plus mindfulness, and virtual-reality behavioural treatment share overlapping intervals from ranks 4 to 7, confirming that their exact order is weakly determined. Behavioural treatment and motivational interviewing plus behavioural treatment remain concentrated near the bottom despite their differing evidence support.

First-rank entropy was 0.313 at $\delta = 0.5$, equivalent to 1.918 effective contenders. Values were similar at $\delta = 0$ (0.337; 2.015 contenders), 0.3 (0.318; 1.937 contenders), and 0.8 (0.308; 1.896 contenders). Increasing the superiority requirement therefore did not disperse the decision

Table 3. Treatment order at a 0.5-unit superiority threshold

Treatment	Dominance	Calibrated dominance	P(rank 1)	Expected rank (95% range)
CBT + mindfulness	0.876	0.646	0.675	1.34 (1–2)
CBT + family	0.844	0.656	0.322	1.69 (1–2)
Mindfulness	0.691	0.481	0.004	2.97 (2–3)
CBT	0.290	0.154	0.000	4.49 (4–6)
BT + mindfulness	0.219	0.111	0.000	5.19 (4–7)
Virtual-reality BT	0.187	0.142	0.000	5.52 (4–7)
BT	0.060	0.049	0.000	7.00 (6–8)
MI + BT	0.014	0.070	0.000	7.80 (6–8)

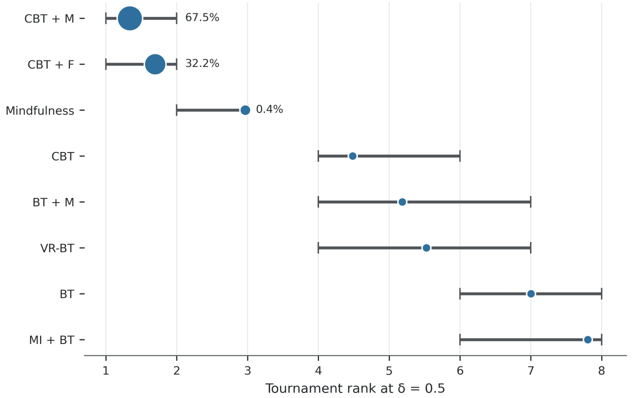

Figure 6. Tournament ranks at $\delta = 0.5$.

across the full treatment set; it concentrated uncertainty within the same two-treatment contest. The top-two first-rank share at $\delta = 0.5$ was 0.996.

Precision-demand calculations clarified which questions are answerable by larger trials. If the contrast estimates remained unchanged, the CBT plus mindfulness–mindfulness comparison would require approximately 6.67 times the current information for its 95% interval to lie entirely above 0.5, while CBT plus family treatment–mindfulness would require 10.82 times the current information. Mindfulness versus CBT required a smaller 1.60-fold increase. No finite information increase could establish a 0.5-unit advantage for CBT plus mindfulness over CBT plus family treatment at the observed difference of 0.23, or for CBT over behavioural treatment plus mindfulness at the observed difference of 0.17. Those questions require either a larger observed separation or acceptance that the treatments are clinically close.

Evidence calibration compressed rather than confirmed the hierarchy. At $\delta = 0.5$, calibrated dominance was 0.646 for CBT plus mindfulness and 0.656 for CBT plus family treatment. Mindfulness followed at 0.481. The 0.010 reversal between the top combinations is too small to support a clinical preference; its importance lies in showing that their apparent separation disappears once limited replication and design properties contract the contrast means. The calibrated values are certainty-adjusted decision probabilities, not alternative efficacy estimates.

3.5. Weight perturbation

Across 50,000 combinations of efficacy weight and evidence-component weights, CBT plus mindfulness ranked first in 90.5% of specifications, CBT plus family treatment in 7.5%, CBT in 1.4%, and mindfulness in 0.6%. The median positions were first, second, fourth, and third, respectively. CBT plus mindfulness ranged from first to third across the central 95% of specifications, while CBT plus family treatment ranged from first to fourth. The high first-place frequency for CBT plus mindfulness reflects its very large rank summary; the non-zero first-place frequency for CBT demonstrates that greater replication and enrolment can compensate when evidential properties receive more weight. No lower treatment became a frequent first choice.

The perturbation distribution also guards against overreacting to one chosen coefficient formula. CBT plus mindfulness remained first under most combinations, but its central interval extended to third, confirming that the position is not immutable. CBT plus family treatment extended to fourth, and CBT occasionally moved to first when replication and information size dominated the weights. This behaviour is informative because the purpose of perturbation is to reveal which conclusions depend on value choices rather than to preserve the initial order by construction.

4. DISCUSSION

The analysis answered a different question from a conventional treatment hierarchy: not which intervention occupies the highest numerical rank, but which treatment distinctions survive a requirement for clinically important separation and scrutiny of evidential support. Three findings are central. Psychological treatments showed benefit against both inactive and active comparators, although the effect against active comparators was smaller and less precise. A leading cluster comprising CBT plus mindfulness, CBT plus family treatment, and mindfulness was consistently separated from several lower-ranked approaches. Within that cluster, however, the order was not sufficiently stable to support an unconditional declaration of one best treatment.

The distinction between inactive and active comparators has direct clinical meaning. Waiting-list and no-treatment groups can estimate improvement relative to delayed or absent care, but they do not balance expectancy, clinician attention, structured activity, or the passage into a credible treatment setting. The decline from $g = 1.70$ against inactive controls to $g = 0.88$ against active controls is therefore not evidence that treatment lacks value. It indicates that a material part of the larger contrast may be attributable to nonspecific or comparator-related processes. Both estimates remained heterogeneous, suggesting that average effects conceal differences in population, treatment dose, measurement, and study conduct. Future efficacy trials should prefer credible active controls when the aim is to identify treatment-specific benefit, while pragmatic trials should compare complete services when the aim is to guide delivery.

The top-cluster result is mechanistically plausible without proving mechanism. CBT may help patients identify triggers, challenge gaming-related expectancies, restructure routines, and develop alternative reinforcement. Mindfulness may add tolerance of craving, distress, boredom, and negative affect without immediate escape into play. Family treatment may be particularly relevant for adolescents whose gaming occurs within cycles of conflict, inconsistent limits, criticism, or reduced shared activity.[11,23] The combination treatments could therefore cover complementary pathways. Yet each leading combination was represented by one study, and mechanistic plausibility cannot replace replication. The appropriate inference is that these combinations deserve direct testing, not that their superiority is established.

Mindfulness occupied an informative position. Its rank summary was below the two combined treatments, but its evidence-support coefficient was higher because the contributing study was randomized, used CBT as an active comparator, and included follow-up. Its probability of exceeding CBT by more than 0.5 was 0.939. This result supports mindfulness as a serious active option within the available network, while the singleton evidence still prevents broad generalization. Evidence from mindfulness-based treatments in substance use and gambling suggests that urge awareness and affect regulation can be clinically relevant across addictive behaviours;[24,25] excessive gaming trials should test whether changes in craving, experiential avoidance, or emotion regulation mediate symptom and functional outcomes.

The middle of the hierarchy was less orderly than a rank list suggests. CBT and behavioural treatment plus mindfulness had similar rank summaries, and the estimated CBT advantage of 0.17 corresponded to only a 0.143 probability of exceeding 0.5. Virtual-reality behavioural treatment was also difficult to distinguish from adjacent approaches. These treatments should be considered statistically unresolved rather than assigned sharply different clinical status. A rank label converts continuous uncertainty into an ordinal statement; the threshold probabilities recover the information lost by that conversion.

Greater replication did not guarantee a higher position. CBT and behavioural treatment had the strongest evidence-support coefficients but lower efficacy ranks. Several explanations are possible. Repeated studies may yield more conservative estimates than isolated early trials. Active comparators and longer observation may reduce apparent effects. Treatment labels may also conceal variation in dose and content. CBT ranged across different session counts and populations, while behavioural treatment included distinct procedures. The result does not show that replicated treatments are inferior. It shows that evidential maturity and estimated efficacy are separate axes that should be reported together.

Measurement diversity further limits clinical interpretation. Eleven self-report instruments were used across 17 studies. Standardization places changes on a common statistical scale but does not ensure that each instrument captures the same construct or functional consequence. Some scales emphasize preoccupation and control, others adapt general internet-addiction items, and diagnostic endorsement measures may differ from continuous severity scales. Self-report can be informative, particularly when gaming behaviour is private, but demand characteristics and perceived stigma may influence responses. Clinical interviews, collateral information where appropriate, behavioural indicators, and functional outcomes should complement symptom scales. Reliable instruments with demonstrated convergent and discriminant validity should be preferred.[17]

This measurement problem also affects the meaning of a 0.5-unit contrast. Standardization removes the original score units but does not equalize content validity or responsiveness. A half-standard-deviation change on a scale dominated by preoccupation may not represent the same clinical gain as the same standardized change on an instrument that requires impairment. The observed threshold probabilities should therefore be interpreted as comparative evidence on the available scales, not as direct probabilities of recovery. Anchor-based work linking score change to restored sleep, attendance, relationships, and self-directed control would make future treatment contrasts more clinically interpretable.

The threshold analysis is deliberately transparent. A 0.5-unit tolerance is not presented as a universally validated minimal important difference for gaming-disorder treatment. It is a probe of decision robustness. The broad conclusions changed little from $\delta = 0.3$ to 0.8: the two CBT combinations dominated first place, mindfulness remained third, and the ordering of adjacent middle treatments was uncertain. A future empirically anchored minimal important difference could replace these probes. Such work would require linking symptom-scale change to patient-reported recovery, functional improvement, or clinician-rated change.

The precision-demand findings prevent an unproductive call for "more research" without specifying what additional information could accomplish. When a point estimate already exceeds 0.5, greater information can in principle establish a moderate advantage if the effect persists. When the point estimate is below 0.5, precision alone will narrow the interval around a clinically small difference. In the latter case, a future trial should be designed for equivalence or non-inferiority, incorporate costs and acceptability, and ask which treatment is easier to deliver rather than presuming that one must become superior. This distinction is particularly important for the two leading CBT combinations, for which the present difference is much smaller than its uncertainty.

Several limitations qualify the findings. The computation used aggregate contrasts and could not recover their covariance. Independent sampling may understate or overstate rank uncertainty because network contrasts sharing a comparator are correlated. The evidence-support coefficient combines five defensible properties, but its functional forms and saturation constant are choices rather than clinical laws. The weight perturbation experiment addresses this concern by varying the balance across a broad range, although it cannot test properties absent from the record, such as therapist competence, protocol adherence, recruitment setting, or differential attrition. Displayed enrolment totalled 715 across study rows while the synthesis described 745 participants; enrolment was consequently used only as a bounded information component. Sparse singleton nodes remain vulnerable to small-study effects even when funnel tests across the full network are non-significant.

The treatment literature would benefit most from direct, adequately powered comparisons among the leading cluster. A three-arm trial of CBT plus mindfulness, CBT plus family treatment, and mindfulness would be especially informative if stratified by developmental stage and family involvement. For adults living independently, the family component may require adaptation to partners or other support figures; for adolescents, parent participation may be central. Trials should preregister a primary gaming outcome, define diagnostic eligibility with validated instruments and clinical assessment, report allocation concealment, monitor treatment fidelity, and retain follow-up of at least six months. Depression, anxiety, sleep, school or work functioning, family conflict, quality of life, and adverse events should be analysed as distinct outcomes rather than folded into a single symptom score.

Clinical application should remain individualized. The available evidence supports offering a structured psychological treatment rather than no care, but it does not justify denying a feasible treatment because another class occupies a higher numerical rank. A patient with strong avoidance and craving may prefer mindfulness-enhanced CBT; an adolescent in escalating family conflict may benefit from structured family participation; a patient with limited access may receive a briefer behavioural or CBT intervention while awaiting a more intensive service. Shared decision-making can integrate these circumstances with the strength and uncertainty of comparative evidence.

5. CONCLUSION

The paper's research question was whether the comparative order of psychological treatments for excessive gaming remains stable when treatment differences must clear clinically relevant thresholds, efficacy is considered alongside evidential support, and unresolved contrasts are translated into confirmatory information demands. The answer is qualified. The available contrasts consistently identify CBT plus mindfulness, CBT plus family treatment, and mindfulness as a leading cluster and distinguish that cluster from several lower-ranked approaches. They do not reliably distinguish CBT plus mindfulness from CBT plus family treatment by a moderate margin, and the observed difference between those combinations is too small for precision alone to establish such an advantage. The clinically defensible decision is therefore to select among the leading approaches according to patient characteristics and service feasibility while prioritizing direct, replicated comparisons that measure durable functional recovery and are designed for superiority only where a clinically important separation is plausible.

CONFLICT OF INTEREST

The author declares no conflict of interest.

BIBLIOGRAFIA / REFERENCES

1. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, 11th revision. Geneva: World Health Organization; 2019.
2. Rumpf H-J, Achab S, Billieux J, Bowden-Jones H, Carragher N, Demetrovics Z, et al. Including gaming disorder in the ICD-11: the need to do so from a clinical and public health perspective. J Behav Addict. 2018;7:556–561.
3. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
4. Stevens MW, Dorstyn D, Delfabbro PH, King DL. Global prevalence of gaming disorder: a systematic review and meta-analysis. Aust N Z J Psychiatry. 2021;55:553–568.
5. King DL, Delfabbro PH. The concept of harm in Internet gaming disorder. J Behav Addict. 2018;7:562–564.
6. Mihara S, Higuchi S. Cross-sectional and longitudinal epidemiological studies of Internet gaming disorder: a systematic review. Psychiatry Clin Neurosci. 2017;71:425–444.
7. Ho RC, Zhang MWB, Tsang TY, Toh AH, Pan F, Lu Y, et al. The association between internet addiction and psychiatric co-morbidity: a meta-analysis. BMC Psychiatry. 2014;14:183.
8. González-Bueso V, Santamaría JJ, Fernández D,

Merino L, Montero E, Ribas J. Association between Internet gaming disorder or pathological video-game use and comorbid psychopathology: a comprehensive review. Int J Environ Res Public Health. 2018;15:668.
9. Stevens MW, King DL, Dorstyn D, Delfabbro PH. Cognitive-behavioral therapy for Internet gaming disorder: a systematic review and meta-analysis. Clin Psychol Psychother. 2019;26:191–203.
10. Li W, Garland EL, McGovern P, O'Brien JE, Tronnier C, Howard MO. Mindfulness-oriented recovery enhancement for Internet gaming disorder in U.S. adults: a stage I randomized controlled trial. Psychol Addict Behav. 2017;31:393–402.
11. Schneider LA, King DL, Delfabbro PH. Family factors in adolescent problematic Internet gaming: a systematic review. J Behav Addict. 2017;6:321–333.
12. He J, Pan T, Nie Y, Zheng Y, Chen S. Behavioral modification decreases approach bias in young adults with Internet gaming disorder. Addict Behav. 2021;113:106686.
13. Zheng Y, He J, Fan L, Qiu Y. Reduction of symptoms after a combined behavioral intervention for reward sensitivity and rash impulsiveness in Internet gaming disorder: a comparative study. J Psychiatr Res. 2022;153:159–166.
14. Zajac K, Ginley MK, Chang R. Treatments of Internet gaming disorder: a systematic review of the evidence. Expert Rev Neurother. 2020;20:85–93.
15. King DL, Delfabbro PH, Wu AMS, Doh YY, Kuss DJ, Pallesen S, et al. Treatment of Internet gaming disorder: an international systematic review and CONSORT evaluation. Clin Psychol Rev. 2017;54:123–133.
16. King DL, Haagsma MC, Delfabbro PH, Gradisar M, Griffiths MD. Toward a consensus definition of pathological video-gaming: a systematic review of psychometric assessment tools. Clin Psychol Rev. 2013;33:331–342.
17. Yoon S, Yang Y, Ro E, Ahn W-Y, Kim J, Shin S-H, et al. Reliability and convergent and discriminant validity of gaming disorder scales: a meta-analysis. Front Psychol. 2021;12:764209.
18. Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. Evid Based Ment Health. 2015;18:40–46.
19. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol. 2011;64:163–171.
20. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. PLoS One. 2014;9:e99682.
21. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: software for semiautomated assessment of the confidence in the results of network meta-analysis. Campbell Syst Rev. 2020;16:e1080.
22. Kim J, Lee S, Lee D, Shim S, Balva D, Choi K-H, et al. Psychological treatments for excessive gaming: a systematic review and meta-analysis. Sci Rep. 2022;12:20485.
23. Bonnaire C, Phan O. Relationships between parental attitudes, family functioning and Internet gaming disorder in adolescents attending school. Psychiatry Res. 2017;255:104–110.
24. Cavicchioli M, Movalli M, Maffei C. The clinical efficacy of mindfulness-based treatments for alcohol and drug use disorders: a meta-analytic review of randomized and nonrandomized controlled trials. Eur Addict Res. 2018;24:137–162.
25. Maynard BR, Wilson AN, Labuzienski E, Whiting SW. Mindfulness-based approaches in the treatment of disordered gambling: a systematic review and meta-analysis. Res Soc Work Pract. 2018;28:348–362.
26. Wölfling K, Müller KW, Dreier M, Ruckes C, Deuster O, Batra A, et al. Efficacy of short-term treatment of Internet and computer game addiction: a randomized clinical trial. JAMA Psychiatry. 2019;76:1018–1025.
27. Torres-Rodríguez A, Griffiths MD, Carbonell X, Oberst U. Treatment efficacy of a specialized psychotherapy program for Internet gaming disorder. J Behav Addict. 2018;7:939–952.
28. Deng L-Y, Liu L, Xia C-C, Lan J, Zhang J-T, Fang X-Y. Craving behavior intervention in ameliorating college students' Internet game disorder: a longitudinal study. Front Psychol. 2017;8:526.
29. Zhang J-T, Yao Y-W, Potenza MN, Xia C-C, Lan J, Liu L, et al. Effects of craving behavioral intervention on neural substrates of cue-induced craving in Internet gaming disorder. Neuroimage Clin. 2016;12:591–599.
30. Park SY, Kim SM, Roh S, Soh M-A, Lee SH, Kim H, et al. The effects of a virtual reality treatment program for online gaming addiction. Comput Methods Programs Biomed. 2016;129:99–108.
31. Choi OY, Son CN. Effects of a self-control training program on online game addiction, aggression, and impulsivity among college students. Korean J Clin Psychol. 2011;30:723–745.
32. Kang HY, Son CN. The effects of self-esteem enhancement cognitive behavioral therapy for adolescents' Internet and game addiction. Korean J Health Psychol. 2010;15:143–159.
33. Liu L, Yao Y-W, Li C-SR, Zhang J-T, Xia C-C, Lan J, et al. Altered intrinsic connectivity distribution in Internet gaming disorder and its associations with psychotherapy treatment outcomes. Addict Biol. 2021;26:e12917.
34. Kuriala GK, Reyes MES. Efficacy of the acceptance and cognitive restructuring intervention program on Internet gaming disorder symptoms of selected Asian adolescents. J Technol Behav Sci. 2020;5:238–244.
35. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting systematic reviews incorporating network meta-analyses. Ann Intern Med.

2015;162:777–784.

36. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539–1558.
37. Dias S, Welton NJ, Caldwell DM, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. Med Decis Making. 2013;33:641–656.
38. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One. 2013;8:e76654.
39. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. A network meta-analysis of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry. 2020;19:92–107.
40. Bowditch L, Chapman J, Naweed A. Do coping strategies moderate the relationship between escapism and negative gaming outcomes in World of Warcraft players? Comput Hum Behav. 2018;86:69–76.
41. Petry NM, Rehbein F, Gentile DA, Lemmens JS, Rumpf H-J, Mößle T, et al. An international consensus for assessing Internet gaming disorder using the new DSM-5 approach. Addiction. 2014;109:1399–1406.
42. Saunders JB, Hao W, Long J, King DL, Mann K, Fauth-Bühler M, et al. Gaming disorder: its delineation as an important condition for diagnosis, management, and prevention. J Behav Addict. 2017;6:271–279.
43. Paulus FW, Ohmann S, von Gontard A, Popow C. Internet gaming disorder in children and adolescents: a systematic review. Dev Med Child Neurol. 2018;60:645–659.