TECHNICAL COMMENTARY

Group therapies

Introduction

Group therapy refers to any psychosocial therapy that is administered in a group setting. It can include specific cognitive or behavioural therapies and is often utilized in inpatient settings. The usefulness of group therapy has been examined in the context of improving illness outcomes such as symptom severity and quality of life, medication compliance, and particularly social interaction and anxiety.

Method

We have included only systematic reviews (systematic literature search, detailed methodology with inclusion/exclusion criteria) published in full text, in English, from the year 2010 that report results separately for people with a diagnosis of bipolar or related disorders. Reviews were identified by searching the databases of MEDLINE, EMBASE and PsycINFO. Hand searching reference lists of identified reviews was also conducted. When multiple copies of review topics were found, the most recent and/or comprehensive review was included. Reviews with pooled data are prioritized for inclusion.

Review reporting assessment was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist which describes a preferred way to present a meta-analysis[1]. Reviews rated as having less than 50% of items checked have been excluded from the library. The PRISMA flow diagram is a suggested way of providing information about studies included and excluded with reasons for exclusion. Where no flow diagram has been presented by individual reviews, but identified studies have been described in the text, reviews have been checked for this item. Note that early reviews may have been guided by less stringent reporting checklists than the PRISMA, and that some reviews may have been limited by journal guidelines.

Evidence was graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach where high quality evidence such as that gained from randomized controlled trials (RCTs) may be downgraded to moderate or low if review and study quality is limited, if there is inconsistency in results, indirect comparisons, imprecise or sparse data and high probability of reporting bias. It may also be downgraded if risks associated with the intervention or other matter under review are high. Conversely, low quality evidence such as that gained from observational studies may be upgraded if effect sizes are large or if there is a dose dependent response. We have also taken into account sample size and whether results are consistent, precise and direct with low associated risks (see end of table for an explanation of these terms)[2]. The resulting table represents an objective summary of the available evidence, although the conclusions are solely the opinion of staff of NeuRA (Neuroscience Research Australia).

Results

We found three systematic reviews that met our inclusion criteria[3-5].

- Moderate to low quality evidence finds medium to large effects of fewer relapses and improved symptoms and functioning following group therapies. Group therapies consisted of psychoeducation, cognitive behavioural therapy, family-focused treatment, functional remediation, and interpersonal and social rhythm therapy.
- There were also fewer relapses with group mindfulness therapy, social cognition and interaction training, and dialectical behaviour therapy.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

Janis RA, Burlingame GM, Svien H, Jensen J, Lundgreen R

Group therapy for mood disorders: A meta-analysis

Psychotherapy Research 2021; 31: 369-85

View review abstract online

Comparison	**Group psychoeducation, cognitive behavioural therapy, or functional remediation (mostly plus medication) vs. treatment as usual (mostly medication alone).**
Summary of evidence	**Moderate to low quality evidence (medium to large samples, some inconsistency and imprecision, indirect) finds medium-sized improvements in symptoms following group therapy, which may persist for up to two years.**
Symptoms and functioning	
Significant, medium-sized effects of improved symptoms and functioning after group therapy; Symptoms and functioning: 6 studies, N = 442, g = 0.69, 95%CI 0.18 to 1.21, p = 0.008, I^2 = 82% Symptoms: 4 studies, N = 242, g = 0.44, 95%CI 0.08 to 0.80, p = 0.017, I^2 = 50% Removing one study reduced the heterogeneity. There were no other significant moderators. *This effect remained at follow-up (7-24 months post-treatment);* Symptoms and functioning: 6 studies, g = 0.61, 95%CI 0.30 to 0.93, p < 0.001, I^2 = 37% There were no differences in drop-out rates.	
Consistency in results	Inconsistent for post-treatment analyses, consistent for follow-up analysis.
Precision in results	Precise for symptoms and follow-up analyses only.
Directness of results	Indirect (mixed treatment types).

Macheiner T, Skavantzos A, Pilz R, Reininghaus EZ

A meta-analysis of adjuvant group-interventions in psychiatric care for patients with bipolar disorders

Journal of Affective Disorders 2017; 222: 28-31

View review abstract online

Comparison	**Group psychoeducation, cognitive behavioural therapy, family-focused treatment, or interpersonal and social rhythm therapy vs. treatment as usual.**
Summary of evidence	**Moderate quality evidence (large sample, inconsistent, precise, indirect) suggests medium to large effect of reduced risk of relapse with group therapies.**
Relapse	
Significant, medium-large effect of reduced risk of relapse with group therapies; 23 RCTs, N = 2,265, RR = 0.71, 95%CI 0.62 to 0.80, $p < 0.01$, $I^2 = 50\%$, $p = 0.003$	
Consistency in results	Inconsistent
Precision in results	Precise
Directness of results	Indirect (mixed treatment types), however the subgroup analyses revealed similar effect sizes for each treatment type.

Oud M, Mayo-Wilson E, Braidwood R, Schulte P, Jones SH, Morriss R, Kupka R, Cuijpers P, Kendall T

Psychological interventions for adults with bipolar disorder: systematic review and meta-analysis

British Journal of Psychiatry 2016; 208: 213-22

View review abstract online

Comparison 1	**Group psychoeducation, cognitive behavioural therapy, mindfulness therapy, social cognition and interaction training, or dialectical behaviour therapy vs. treatment as usual.**
Summary of evidence	**Moderate to low quality evidence (medium to large samples, consistent, some imprecision, indirect) suggests a medium-sized reduced risk of relapse to depression, mania, or a mixed episode after of 21 weeks of group therapy compared to treatment as usual. This effect was maintained at 1-2 year follow-up. However, there were no differences between group therapy and treatment as usual in symptom severity or hospital admissions.**
Relapse	
Significant, medium-sized reduced risk of relapse to depression, mania or a mixed episode with	

group therapies post-treatment and at follow-up; Depression relapse after 21 weeks of treatment: 2 RCTs, N = 170, RR = 0.39, 95%CI 0.19 to 0.78, $p < 0.05$, $I^2 = 0\%$, $p = 0.50$ Depression relapse at 52-124 weeks following treatment: 5 RCTs, N = 333, RR = 0.62, 95%CI 0.45 to 0.88, $p < 0.05$, $I^2 = 44\%$, $p = 0.13$ Mania relapse after 21 weeks of treatment: 2 RCTs, N = 170, RR = 0.48, 95%CI 0.28 to 0.82, $p < 0.05$, $I^2 = 0\%$, $p = 0.37$ Mixed episode relapse at 52-124 weeks following treatment: 4 studies, N = 274, RR = 0.48, 95%CI 0.30 to 0.77, $p < 0.05$, $I^2 = 0\%$, $p = 0.50$ *There were no significant differences between groups when any relapse was assessed together:* Any relapse after 21 weeks of treatment: 2 RCTs, N = 170, RR = 0.48, 95%CI 0.22 to 1.04, $p > 0.05$, $I^2 = 59\%$, $p = 0.12$ Any relapse at 52-124 weeks following treatment: 5 RCTs, N = 395, RR = 0.86, 95%CI 0.61 to 1.20, $p > 0.05$, $I^2 = 81\%$, $p = 0.0003$
Hospital admissions
There were no significant differences between groups in hospital admissions: After 14-21 weeks of treatment: 3 RCTs, N = 205, RR = 0.45, 95%CI 0.10 to 2.09, $p > 0.05$, $I^2 = 49\%$, $p = 0.14$ At 78-124 weeks following treatment: 3 RCTs, N = 200, RR = 0.48, 95%CI 0.16 to 1.45, $p > 0.05$, $I^2 = 56\%$, $p = 0.13$
Depression symptoms
There were no significant differences between groups in depression symptoms: After 8-52 weeks of treatment: 8 RCTs, N = 423, SMD = -0.24, 95%CI -0.64 to 0.16, $p > 0.05$, $I^2 = 73\%$, $p < 0.001$ At 52-61 weeks following treatment: 3 RCTs, N = 219, SMD = 0.22, 95%CI -0.05 to 0.49, $p > 0.05$, $I^2 = 0\%$, $p = 0.62$
Mania symptoms
There were no significant differences between groups in mania symptoms: After 8-52 weeks of treatment: 6 RCTs, N = 375, SMD = -0.08, 95%CI -0.33 to 0.16, $p > 0.05$, $I^2 = 11\%$, $p = 0.35$ At 52-61 weeks following treatment: 3 RCTs, N = 219, SMD = 0.16, 95%CI -0.10 to 0.43, $p > 0.05$, $I^2 = 0\%$, $p = 0.68$

Comparison 2	**Group therapies vs. active controls (mixed interventions).**
Summary of evidence	**Low quality evidence (small sample) is unable to determine any benefit of group therapies compared to active controls.**
Depressive symptoms	
There were no significant differences between groups in depression symptoms: After 12 weeks of treatment: 1 RCT, N = 61, SMD = -0.35, 95%CI -0.85 to 0.16, $p > 0.05$ At 26 weeks following treatment: 1 RCT, N = 61, SMD = 0.11, 95%CI -0.39 to 0.61, $p > 0.05$	
Mania symptoms	
There were no significant differences between groups in mania symptoms after treatment, but there was a medium-sized effect of reduced mania symptoms at follow-up: After 12 weeks of treatment: 1 RCT, N = 61, SMD = -0.17, 95%CI -0.68 to 0.33, $p > 0.05$ At 26 weeks following treatment: 1 RCT, N = 61, SMD = -0.53, 95%CI -1.05 to -0.02, $p < 0.05$	
Consistency in results	Mostly consistent (inconsistent only for comparison 1, depression symptoms post-treatment and any relapse at follow-up).
Precision in results	Imprecise for RRs, precise for SMDs.
Directness of results	Indirect (mixed group therapies and mixed controls).

Explanation of acronyms

CI = confidence interval, g = Hedges standardized mean difference, N = number of participants, I^2 = the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance), p = statistical probability of obtaining that result ($p < 0.05$ generally regarded as significant), RR = risk ratio, SMD = standardized mean differences, vs. = versus

Explanation of technical terms

* Bias has the potential to affect reviews of both RCT and observational studies. Forms of bias include; reporting bias – selective reporting of results; publication bias - trials that are not formally published tend to show less effect than published trials, further if there are statistically significant differences between groups in a trial, these trial results tend to get published before those of trials without significant differences; language bias – only including English language reports; funding bias - source of funding for the primary research with selective reporting of results within primary studies; outcome variable selection bias; database bias - including reports from some databases and not others; citation bias - preferential citation of authors. Trials can also be subject to bias when evaluators are not blind to treatment condition and selection bias of participants if trial samples are small[6].

† Different effect measures are reported by different reviews.

Prevalence refers to how many existing cases there are at a particular point in time. Incidence refers to how many new cases there are per population in a specified time period. Incidence is usually reported as the number of new cases per 100,000 people per year. Alternatively some studies present the number of new cases that have accumulated over several years against a person-years denominator. This denominator is the sum of individual units of time that the persons in the population are at risk of becoming a case. It takes into account the size of the underlying population sample and its age structure over the duration of observation.

Reliability and validity refers to how accurate the instrument is. Sensitivity is the proportion of actual positives that are correctly identified (100% sensitivity = correct identification of all actual positives) and specificity is the proportion of negatives that are correctly identified (100% specificity = not identifying anyone as positive if they are truly not).

Weighted mean difference scores refer to mean differences between treatment and comparison groups after treatment (or occasionally pre to post treatment) and in a randomised trial there is an assumption that both groups are comparable on this measure prior to treatment. Standardised mean differences are divided by the pooled standard deviation (or the standard deviation of one group when groups are homogenous) which allows results from different scales to be combined and compared. Each study's mean difference is then given a weighting depending on the size of the sample and the variability in the data. 0.2 represents a small effect, 0.5 a medium effect, and 0.8 and over represents a large effect[6].

Odds ratio (OR) or relative risk (RR) refers to the probability of a reduction (< 1) or an increase (> 1) in a particular outcome in a treatment group, or a group exposed to a risk factor, relative to the comparison group. For example, a RR of 0.75 translates to a reduction in risk of an outcome of 25% relative to those not receiving the treatment or not exposed to the risk factor. Conversely, a RR of 1.25 translates to an increased risk of 25% relative to those not receiving treatment or not having been exposed to a risk factor. A RR or OR of 1.00 means there is no difference between groups. A medium effect is considered if RR > 2 or < 0.5 and a large effect if RR > 5 or < 0.2[7]. lnOR stands for logarithmic OR where a lnOR of 0 shows no difference between groups. Hazard ratios measure the effect of an explanatory variable on the hazard or risk of an event.

Correlation coefficients (eg, r) indicate the strength of association or relationship between variables. They can provide an indirect indication of prediction, but do not confirm causality due to possible and often

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

unforseen confounding variables. An r of 0.10 represents a weak association, 0.25 a medium association and 0.40 and over represents a strong association. Unstandardised (*b*) regression coefficients indicate the average change in the dependent variable associated with a 1 unit change in the independent variable, statistically controlling for the other independent variables. Standardised regression coefficients represent the change being in units of standard deviations to allow comparison across different scales.

‡ Inconsistency refers to differing estimates of effect across studies (i.e. heterogeneity or variability in results) that is not explained by subgroup analyses and therefore reduces confidence in the effect estimate. I^2 is the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance) - 0% to 40%: heterogeneity might not be important, 30% to 60%: may represent moderate heterogeneity, 50% to 90%: may represent considerable heterogeneity and over this is considerable heterogeneity. I^2 can be calculated from Q (chi-square) for the test of heterogeneity with the following formula[6];

$$I^2 = \left(\frac{Q - df}{Q}\right) \times 100\%$$

§ Imprecision refers to wide confidence intervals indicating a lack of confidence in the effect estimate. Based on GRADE recommendations, a result for continuous data (standardised mean differences, not weighted mean differences) is considered imprecise if the upper or lower confidence limit crosses an effect size of 0.5 in either direction, and for binary and correlation data, an effect size of 0.25. GRADE also recommends downgrading the evidence when sample size is smaller than 300 (for binary data) and 400 (for continuous data), although for some topics, these criteria should be relaxed[8].

‖ Indirectness of comparison occurs when a comparison of intervention A versus B is not available but A was compared with C and B was compared with C that allows indirect comparisons of the magnitude of effect of A versus B. Indirectness of population, comparator and/or outcome can also occur when the available evidence regarding a particular population, intervention, comparator, or outcome is not available and is therefore inferred from available evidence. These inferred treatment effect sizes are of lower quality than those gained from head-to-head comparisons of A and B.

Margarete Ainsworth Building, Barker Street, Randwick NSW 2031. Phone: 02 9399 1000. Email: info@neura.edu.au

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMAGroup (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 151: 264-9.
2. GRADEWorkingGroup (2004): Grading quality of evidence and strength of recommendations. *British Medical Journal* 328: 1490.
3. Macheiner T, Skavantzos A, Pilz R, Reininghaus EZ (2017): A meta-analysis of adjuvant group-interventions in psychiatric care for patients with bipolar disorders. *Journal of Affective Disorders* 222: 28-31.
4. Oud M, Mayo-Wilson E, Braidwood R, Schulte P, Jones SH, Morriss R*, et al.* (2016): Psychological interventions for adults with bipolar disorder: systematic review and meta-analysis. *British Journal of Psychiatry* 208: 213-22.
5. Janis RA, Burlingame GM, Svien H, Jensen J, Lundgreen R (2021): Group therapy for mood disorders: A meta-analysis. *Psychotherapy research : journal of the Society for Psychotherapy Research* 31: 369-85.
6. CochraneCollaboration (2008): Cochrane Handbook for Systematic Reviews of Interventions. Accessed 24/06/2011.
7. Rosenthal JA (1996): Qualitative Descriptors of Strength of Association and Effect Size. *Journal of Social Service Research* 21: 37-59.
8. GRADEpro (2008): [Computer program]. Jan Brozek, Andrew Oxman, Holger Schünemann. *Version 32 for Windows*