

**Hoffman BM, Papas RK, et al. Meta-Analysis of Psychological Interventions for Chronic Low Back Pain. Health Psychology 2007;26(1):1-9.**

Design: Meta-analysis of controlled clinical trials

PICOS:

- **Patient population:** adults with nonmalignant chronic low back pain lasting at least 3 months

- **Interventions:** a wide variety of psychosocial interventions, including counseling, psychotherapy, behavior therapy, biofeedback, cognitive therapy, relaxation techniques, meditation, gestalt therapy, and several other forms of psychological intervention
- **Comparison intervention:** waiting list, active control, and combined waiting list/active control
- **Outcomes:** pain intensity, emotional functioning, physical functioning, health-related quality of life, pain interference, pain-specific disability, global improvement, treatment satisfaction
- **Study types:** English language with randomized and quasi-randomized assignment (such as alternating assignment), provided that there was a comparison between groups and a pain outcome was reported

Search strategy and selection:

- Databases included MEDLINE, PsychINFO, EMBASE, CENTRAL, CINAHL through October 2004
- 952 abstracts were retrieved; 196 articles were reviewed, of which 39 met inclusion criteria for the meta-analysis
- Not all 39 articles had data suitable for extraction (e.g., some lacked sample sizes, means, or standard deviations); data were extracted from 34 articles detailing 31 separate research studies

Results:

- Meta-analysis was done using standard mean differences between intervention and control groups; this means that the effect size was defined as the difference between the mean psychological group outcome and the control group outcome, divided by the standard deviation of the pooled group outcome
- This number is known as Cohen's *d*, if it is 0.2, that is considered small, if it is 0.5, that is considered moderate, and if it is 0.8 or greater, that is considered a large difference between treatment groups
- Several kinds of comparison were made, since different studies made different kinds of intervention and control groups and used different outcomes
- One outcome was pain intensity; the pooled effect size depended on whether the comparison was made right after treatment, made at a later follow-up, and whether the control group was put on a waiting list or received some other form of active treatment

- When pain intensity was compared at the end of treatment, combined psychological and multidisciplinary treatment was superior to a waiting list; data from 7 studies with 382 participants gave a Cohen's  $d$  of 0.5, which is moderate; however, when the control group received an active treatment, using data from 5 studies with 308 participants,  $d$  was much smaller, only 0.06, which is statistically the same as zero
- When pain intensity was compared between combined psychological and multidisciplinary treatment and an active control at follow-up, data from 5 studies with 393 participants gave a  $d$  of only 0.16, which was also not statistically significant
- Cognitive-behavioral treatment (CBT) was superior to waiting list for pain intensity; 4 studies with 256 participants gave a  $d$  of 0.62 at the end of treatment
- CBT was not superior to self-regulatory treatment (biofeedback and relaxation training) on pain intensity or depression at the end of treatment; for depression, self-regulatory treatment was slightly better than CBT, with 3 studies and 182 participants yielding a  $d$  of 0.41
- While numerous other comparisons were made, most of them had low heterogeneity in the meta-analysis, suggesting that the estimated effect sizes were similar across studies

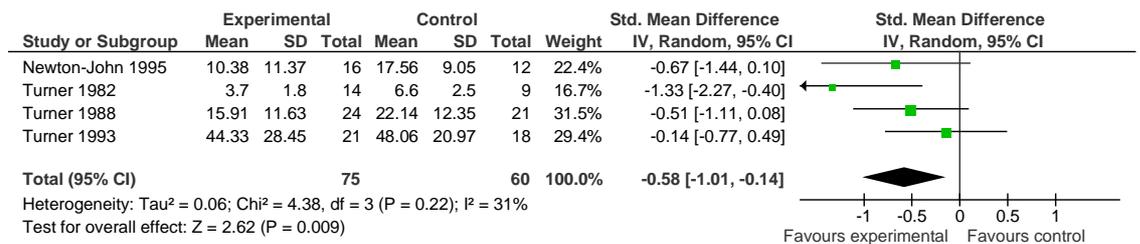
#### Authors' conclusions:

- Psychological interventions appear to be superior to wait-list controls for pain intensity and health-related quality of life; and for work-related disability
- There is less superiority of psychological over other active interventions
- Self-regulatory treatments such as biofeedback and relaxation training have fairly strong effects, and may outperform CBT on relieving depression
- The findings of effectiveness of psychological interventions are fairly robust, and should encourage confidence among clinicians and researchers

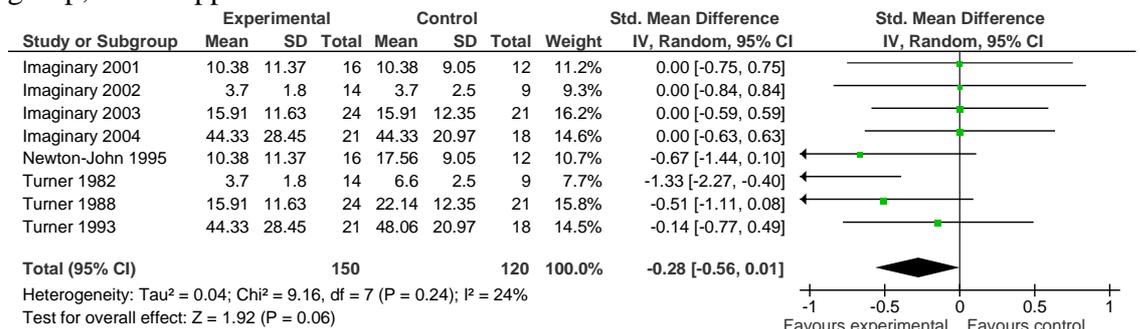
#### Comments:

- The research question is an intrinsically difficult one to examine by meta-analysis, because the psychological interventions are so variable in their approaches and applications that a statistically significant effect size cannot be applied to endorse any one therapeutic method
- One minor ambiguity in the study selection is that quasi-randomized studies (such as alternating sequence rather than true randomization) were considered eligible for inclusion into the meta-analysis; yet many studies were excluded for failure to randomize
- Unlike Cochrane meta-analyses, which show forest plots for each comparison, stating which studies were used in each comparison, this meta-analysis gives no information about which studies were used to compile the pooled effect sizes
  - The authors were contacted for more information, and supplied additional data for many of their analyses

- The authors used fail-safe N to estimate how many unpublished studies with effect sizes of zero would be needed to reduce the significance of the pooled results to a statistically non-significant value ( $p > 0.10$ )
- Fail-safe N is sometimes employed to deal with the so-called “file-drawer problem,” which postulates that many studies with null results end up being filed away rather than submitted for publication; if N is small, this suggests that it would not require a large number of additional null studies to create a meta-analysis with a non-significant pooled effect size
  - o If N is large, this suggests that the meta-analysis results are likely to be robust to the discovery and publication of unpublished null studies
- Fail-safe N has some problems which have caused it to fall into disfavor since it was put forth in 1979; its focus on p values and statistical significance rather than on clinically meaningful effect sizes is the major objection to its use (the 2008 Cochrane)
- Fail-safe N for the comparison of CBT vs. wait-list on pain intensity (the top line of Table 4), states that 4 studies were used to estimate a  $d$  of .62, and that fail-safe N was 4.07
- RevMan software was used to create a forest plot using data supplied by the authors on request, and is appended below; the value of  $d$  is very close to .62



- Not all of the numbers match exactly, but the overall effect estimate is close
- Fail-safe N of 4 was tested by imagining that four other studies of the same sample size and standard deviations had been done, but that in each study, the mean outcome of the control group was equal to that of the experimental group; that is appended below



- If these four imaginary studies had been combined with the published studies, the effect size ( $d$ ) would be only 0.28 instead of 0.58, and the 95% confidence interval (the Total at the bottom) would cross the null value of 0, making it “not statistically significant”
- The other fail-safe Ns in Tables 2, 3, and 4 have similar interpretations

- The fail-safe Ns are not much greater than the number of included studies
- This is difficult to interpret, since it suggests that the estimated pooled effect sizes may not be as robust as was hoped by the authors
- The latter finding, however, need not mean that the effect of CBT and other psychological interventions is weak, only that the means of estimating its robustness is not satisfactory (and has, in fact, been supplanted by newer methods of dealing with the file-drawer problem, such as maximum likelihood models)

Assessment: Adequate for good evidence that psychological interventions, especially CBT, are superior to no psychological intervention for chronic low back pain, and that self-regulatory interventions such as biofeedback and relaxation training may be equally effective