Review: bona fide psychotherapies are similar in effectiveness


Question
Are all bona fide psychotherapies equally effective (Dodo bird effect)?

Data sources

Study selection
Studies were selected if they were published in any of the 6 journals between 1970 and 1995, contained data that could be used to calculate effect sizes for the dependent measures, and compared ≥2 bona fide psychotherapies. Bona fide treatment involved a therapist with at least a master’s degree who tailored a treatment based on psychological principles and the patient problem was one that would reasonably be treated by psychotherapy. Studies with treatments for controlling common or non-specific factors such as placebos, alternative therapies, or non-specific therapies were excluded.

Data extraction
Data were extracted on journal name, publication year, treatment description, number of participants receiving each treatment, and means and standard deviations (SDs) for each outcome measure. The effect sizes were calculated by dividing the difference in means for the 2 bona fide psychotherapies being compared by the pooled SDs, and an aggregated effect size across outcome measures was estimated.

Main results
114 studies were included. The results were consistent with the Dodo bird effect. For the base data set in which all comparisons with sufficient summary statistics were included (277 effects), the effects were homogeneously distributed around zero, indicating that there was insufficient evidence to reject the null hypothesis that bona fide psychotherapies varied in their effectiveness; the upper bound of the true difference among psychotherapies was about 0.18. The addition of studies that only reported summary statistics for statistically significant outcome measures did not appreciably alter the results, with the effects remaining homogeneously distributed around 0 (295 effects) (upper bound 0.19). The aggregate effect size for data sets that included final measurements only (182 effects) and those from which the dependency among 3 treatments in a single study was modelled (136 effects) were also homogeneously distributed around zero (upper bounds 0.21 and 0.21, respectively). Publication year and treatment similarity did not predict effect size: results which are consistent with the Dodo bird effect.

Conclusion
The Dodo bird effect that states that all bona fide psychotherapies are equally effective is supported by this meta-analysis; effect sizes are homogeneously distributed around zero, indicating that the best estimate of true differences among the psychotherapies is zero.

Commentary
The Dodo bird hypothesis implies that “one size fits all.” If true, clinicians could not claim the superiority of one treatment over another, and researchers would need to examine the common factors that underlie the general efficacy of therapies.

Although meta-analyses are powerful, they require careful interpretation, and readers would benefit from reading the two commentaries and author’s reply that are published with the review by Wampold et al.1 A fundamental issue is the range of studies on which the meta-analysis is based: (1) it includes studies of conditions which are clinically unrepresentative; (2) the evidence base as a whole contains relatively few contrasts among commonly practised treatments; and the meta-analysis reflects this; and (3) it includes many contrasts of treatments which are variants of the same broad school, such as behavioural and cognitive behavioural treatments. Such treatments have common technical (as well as paratheoretical) elements, and finding that both approaches are equivalent in outcome is less compelling than finding no differences between treatments with more disparate theoretical bases.

A major problem with the Dodo bird hypothesis is that although a research perspective may well alert us to a uniformity of outcomes, as we move to a clinical setting woods resolve into trees. How does the hypothesis apply here? On the one hand we need to accept the broad conclusions of research—that at least for certain conditions (such as depression) there is little compelling evidence for the superiority of one treatment over another. But should this lead us to dismiss the importance of “brand names” of treatment? Averaging the results of treatments over all clinical populations may obscure the fact that some patients may be particularly responsive or unresponsive to different interventions. It also seems that some techniques are particularly helpful for some conditions (exposure appears to be a robust treatment component for anxiety disorders). And finally, research is only just beginning to yield information about the effects of comorbidity, complexity, and severity on outcomes.

We know that psychological treatments are effective, but questions remain about the admixture of technical and pantheoretical elements which contribute to efficacy. On this the jury is still out and it is premature to say that “anything goes.”

Anthony Roth, BSc, Dip Clin Psych, PhD
University College London
London, UK