**Questions for Critical Appraisal of Primary Effectiveness Study**

**What is the paper about?**

1. What clinical question did the paper address?
2. What type of study was done?
	1. Primary research (experiment, RCT, other controlled clinical trial, cohort study, other)? (If yes, this set of questions will help!)
	2. Secondary research (systematic review, meta-analysis)? (If yes, use the other set of questions!)
3. Was the study design appropriate for providing evidence about effectiveness?
4. Did the study meet expected standards of ethics/governance?

**What intervention is being studied?**

1. What problem does this intervention address?
2. What is the theory of change? How does this intervention address the problem?
3. What are the core and non-core components of the intervention?
4. Who did the research and what, if any, are their conflicts of interest?

**What methods were used for the study?**

1. **Participants**
2. Whom is the study about?
3. How were participants recruited?
4. Who was included in, and who was excluded from, the study?
5. Were the participants studied in “real-life” circumstances?
6. Is the sample representative of the population to which it is designed to generalize?
7. **Measures**
	1. Was the design of the study sensible?
	2. What outcome(s) were measured, and how?
	3. Do the outcomes make sense given the theory of change?
	4. Were the measures of high quality?
	5. Could the results be because of repeated testing of participants? Self-report?
8. **Comparison**
9. If a “randomized trial,” was randomization truly random?
10. If a nonequivalent comparison design, does the study do one or more of the following to alleviate doubt about a selectivity bias and provide a persuasive case for assuming that the groups being compared are really comparable?
	1. Does it present data comparing the groups on client attributes such as age, ethnicity, socioeconomic status, diagnosis, degree of impairment or well-being, pretest scores, and so on?
	2. Does it provide information showing that the facilities being compared are similar in regard to things like practitioner characteristics, caseload size, and so on?
	3. Does it provide logical grounds for deeming the threat of a selectivity bias to be far-fetched, such as by using:
		1. An overflow design?
		2. Multiple pretests?
		3. Switching replications?
		4. A nonequivalent dependent variable?
11. Does the study use multivariate data analysis procedures to control statistically for possible differences between groups that might explain away differences in outcome?
12. Does the study adequately handle potential problems in the following areas?
13. Measurement bias and obtrusive observation? (Was assessment of outcome “blind?”)
14. Treatment diffusion? (Did treatment effects spill over to any of the members of the comparison group?)
15. Treatment fidelity?
	* + 1. What was done to control/monitor implementation of the intervention?
16. Differential attrition?
17. Practitioner equivalence?
18. Research reactivity:
	* + 1. Placebo effects?
			2. Compensatory equalization?
			3. Compensatory rivalry?
			4. Resentful demoralization?
19. Was the study continued for long enough and was follow up complete enough to make the results credible?
20. Were there events during the study that influenced the results?
21. **Statistical Analysis**
22. Was the study large enough (sample size) to make the results credible? Was a power analysis presented to justify the sample size?
23. Have the authors set the scene correctly?
	1. Does the study use multivariate data analysis procedures to control statistically for possible differences between groups that might explain away differences in outcome?
	2. Did group differences evolve as a result of differential subject attrition (drop-out)?
24. Did the experimental group include only very high-risk participants, such that behaviors would appear to improve on their own by ‘regressing’ to the mean (i.e. natural improvement)?
25. What sort of data have they got, and have they used appropriate statistical tests?
26. If the statistical tests in the paper are obscure, did the authors explain clearly why they chose to use them?
27. Paired data, tails, and outliers
	1. Were paired tests performed on paired data?
	2. Was a two-tailed test performed whenever the effect of an intervention could conceivably be a negative one?
	3. Were outliers analyzed with common sense?
28. Probability and confidence
	1. Have “*p*-values” been calculated and interpreted appropriately?
	2. Have confidence intervals been calculated and do the authors’ conclusions reflect them?
29. Have the authors expressed their results in terms of effect size or reported the necessary data to calculate effect size?
30. Have the authors expressed their results in terms of the likely harm or benefit which an individual patient can expect?

**What implications can be drawn from the study?**

1. Did the study include or exclude people with diagnoses or other characteristics like the client(s) pertaining to your PICO question?
2. Does your community or agency have the resources necessary to implement the intervention studied?
3. How would you go about developing competence for delivering this intervention with fidelity?
4. If the findings were negative, to what extent can this be explained by implementation failure and/or inadequate optimization of the intervention?
5. If the findings carried across different subgroups, to what extent have the authors explained this within their theory of change?
6. What further research do the authors believe is needed, and is this justified?

**References**

Greenhalgh, T. (2010) *How to read a paper: The basics of evidence-based medicine* (4th edition). Hoboken, NJ: BMJ Books.

Olson, J.R. (2010) Choosing effective youth-focused prevention strategies: A practical guide for applied family professionals. *Family Relations: Interdisciplinary Journal of Applied Family Studies, 59*, 207-220.

Rubin, A. (2008). *Practitioner’s guide to using research for evidence-based practice.* Hoboken, NJ: John Wiley & Sons.