**Evaluation Guidelines for Rating the Quality of an Intervention Study**

*This guide helps you critically appraise a published paper by assigning scores for each item of the critical appraisal checklist . Select the score where the description sounds most like what was done and reported in the study. In general, scores of 2 indicate strong research design, a 1 is suboptimal, and a 0 is poor research design.*

|  |  |  |
| --- | --- | --- |
| Question | | **Descriptors** |
| # | Score |  |
| 1 | 2 | The authors  -performed a thorough literature review indicating what is currently known about the problem being addressed and the intervention(s) being evaluated  - presented a critical, but unbiased, view of the current state of knowledge  - indicated how the current research question evolves from the current knowledge base  - established a clear research question(s) based on the above (PICOT: patients, intervention, comparison, outcomes evaluated and timeframes) |
| 1 | All of these above were not fulfilled, but some background literature and a clear rationale for the research question were provided; the research question was general, not specific |
| 0 | A foundation for the current research question was not developed; and the research question was clearly stated. |
|  | | **Study design** |
| 2. | 2 | Two or more contemporary (same point in time) groups of similar patients were compared. Crossover trials which include randomization/blinding of intervention order and complete wash-out of effects equally appropriate. |
| 1 | A comparitor group was present, but it was not clear that the groups were comparable. |
| 0 | No comparitor group was included. |
| 3. | 2 | Patients were evaluated prior to the intervention, and at one or more clinically relevant time points, following the intervention using the same evaluation process. |
| 1 | Patients were evaluated at more than one point in time (including case control studies); but the above criteria were not fulfilled |
| 0 | Patients were evaluated at only one point in time. |
| 4. | 2 | A standardized set of data were collected at specific pre-set intervals according to a preplanned study protocol.(Prospective cohort) |
| 1 | A set of prospective data were collected from patients and later retrieved e.g from a database. This data was collected across multiple intervals, but the actual data collection strategy was not determined specifically for this study (retrospective cohort) |
| 0 | Data were based on retrospective records/interpretations or recall of past events. |
| 5. | 2 | An appropriate randomization strategy was used to allocate patients to interventions and the specifics of randomization were described. |
| 1 | Randomization was used, but information describing the randomization process was not included or did not confirm a truly random process. |
| 0 | Randomization was not used |
| 6. | 2 | All relevant people involved in the trial including care providers, patients, study administrators etc. are not informed about which treatment has been assigned until immediately before the treatment is administered. This includes the explicit decision and protections required to keep allocation concealed. |
| 1 | Either treatment providers or patients are unaware of the treatment to be provided until immediately before the treatment is administered |
| 0 | Allocation to treatment is not blinded; providers and patients are aware of the allocation well before treatment administration |
| 7. | 2 | Patients were blinded from the knowledge about which intervention they received. A thorough description of blinding procedures or a post-hoc analysis to indicate blinding indicated that blinding was effective; or it was evident that patients would be unable to distinguish which intervention they received. |
| 1 | Blinding patients was not possible or it was unclear whether an effective blinding strategy was used. |
| 0 | Blinding was possible, but not utilized (includes all studies without comparison groups). |
| 8. | 2 | Treatment providers were blinded to the intervention they were administering and this blinding was substantiated either through audits or other post-hoc analyses indicated that the blinding procedure was effective. |
| 1 | Blinding was not possible or it was unclear whether an effective blinding strategy was used |
| 0 | Blinding was possible, but was not utilized |
| 9. | 2 | Outcome measures were administered by an evaluator who was blind to the treatment provided or the purpose of the study. Self-report measures can be considered blinded - if provided by an independent person. |
| 1 | Evaluators were not blinded, but were not involved in treatment of patients (were independent) or Self-report forms were administered by treatment provider. |
| 0 | Outcome measures were obtained by unblinded treatment providers who could influence the measurements |
|  | | **Subjects** |
| 10. | 2 | The authors explained their recruitment strategy and reported what recruitment rate was obtained from their target population. Recruitment and sampling procedures were applied equally across comparison groups. |
| 1 | The study sample appears representative of the population of clinical interest, but adequate information on sampling procedures or description of the reference population is not provided. |
| 0 | Sampling biases are evident; systematic differences occurred between the comparison groups; and/or selection procedures used make it impossible to determine what types of patients were included. |
| 11. | 2 | Specific inclusion and exclusion criteria for the study were defined. |
| 1 | Some information on the type of patients included in the study is provided, but the information is insufficient to identify a specific clinical population. |
| 0 | No information on inclusion and exclusion criteria are stated; and limited patients descriptors are provided (e.g. 3 or less including age and gender). |
| 12. | 2 | Authors performed a sample size calculation to provide a minimum of 80% power on their primary outcome measure and recruited the prespecified number of subjects |
| 1 | The provided a satisfactory rationale for the number of subjects included in the study; or the sample exceeded 100 patients/per study arm. |
| 0 | The size of the sample was not justified. |
| 13. | 2 | 90% or more of the patients enrolled or eligible for study were evaluated for outcomes. |
| 1 | More than 70% of the patients eligible for study or enrolled were evaluated for outcomes. |
| 0 | Less than 70 percent of patients eligible for study or enrolled were evaluated. |
|  | | **Intervention** |
| 14. | 2 | The parameters of the treatments provided (provider/equipment, frequency, duration, application process, progression and other technical components) were sufficiently described that they could be replicated. The specific parameters used were based on published basic science or clinical evidence documenting that the specific treatment effects intended are achievable given the treatment parameters used. Treatment fidelity was considered (efforts were in place to make sure treatment was provided per study protocol (efficacy) or monitored (effectiveness). |
| 1 | A sound rationale and appropriate citations were provided for the treatment assignments; but there was an adequate detail about dosage for application. |
| 0 | Neither inadequate rationale or description of the interventions were provided OR the the intervention applied did not conform to present knowledge about what treatment parameters could potentially be effective. |
| 15. | 2 | Efforts were made to minimize sources of potential treatment provider bias as indicated by the type of study design (efficacy versus effectiveness). Blinding can minimize provider bias. Other methods for considering provider effects can include controlling or monitoring of providers actions or training. Examples include standardizing treatment protocols, equalizing attention to groups, selecting treatment providers without vested interests in a specific intervention, training treatment providers according to a standardized process or assuring a specific level of training when recruiting providers can be used to assure sufficient equipoise. For study designs focused on effectiveness it may be appropriate to look at the variation between providers-rather than control for it. |
| 1 | Minimal attention was directed either in methods, analysis, or discussion for the potential impact of treatment providers having differential outcomes, but no inherent opportunity for clinically relevant bias was apparent. |
| 0 | No attention was directed at the potential for treatment provider biases and the opportunity for bias is evident, given the nature in which interventions were applied. |
| 16. | 2 | An appropriate rationale was provided for the comparison intervention selected. Where no specific intervention has previously been demonstrated to be effective, placebo is an appropriate comparitor. Where a previously established effective intervention exists, this standard of care is an appropriate comparator. |
| 1 | A clear rationale for the comparison group was not established. |
| 0 | No comparison group was included. |
|  | | **Outcome** |
| 17. | 2 | A primary outcome measure which represents a clinically important outcome was selected and supported by evidence of appropriate psychometric properties (reliability, validity, responsiveness). |
| 1 | A relevant primary outcome measure was evident, but was insufficient in either its clinical relevance or its psychometric properties. |
| 0 | A primary outcome measure was not stated; or did not have validity for clinical or methodological reasons. |
| 18. | 2 | Appropriate secondary outcome measures were identified that augmented the perspective provided by the primary outcome measure, ensuring a comprehensive view of the mechanism of intervention action and associated outcomes was obtained; and these secondary outcome measures had sound psychometric properties. |
| 1 | Secondary or multiple outcome measures were collected, but were deficient either in terms of their relevance or methodological properties. |
| 0 | Appropriate secondary outcome measures were not collected. |
| 19. | 2 | Patients were followed at clinically relevant time points. This may include the to measure both the early response and longer-term outcomes. A clear rationale and/or discussion of the appropriateness of these follow-up periods was included. |
| 1 | At least one relevant follow-up evaluation was incorporated, but the study did include other important clinical time points or a rationale for the specific follow-up time. |
| 0 | The follow-up period was insufficient to establish the clinical utility of the intervention. |
|  | | **Analysis** |
| 20. | 2 | Statistical tests were performed to establish treatment effectiveness. In each case the statistical test was an appropriate option for the numerical properties of the outcome measures. The authors documented important elements on the statistical tests (e.g software used, that statistical assumptions underlying tests were met, Alpha levels). |
| 1 | Statistical effects of treatment effectiveness were used, but it was unclear what specific analyses were performed or if data fulfilled the assumptions of the statistical tests used; or there was incomplete presentation of the findings of these tests. |
| 0 | Statistical results for treatment effectiveness were not presented or those selected were not appropriate to the research question or data collected. |
| 21. | 2 | Adequate power was established. A justified sample with significant statistical differences or narrow confidence intervals can indicate this. If there was a lack of statistical difference for treatment effectiveness, a post-hoc power analysis was conducted and identified that the study was appropriately powered. |
| 1 | The sample size was substantial (> 100/arm), but post-hoc power analyses were not conducted in response to nonsignificant results. |
| 0 | The sample size was small and post-hoc power analyses were not conducted in response to nonsignificant results. |
| 22. | 2 | The authors supplemented the reporting of statistical significance with information about the size of the treatment effects observed. This could be indicated by the inclusion of confidence intervals around the mean differences between treatment groups with reference to the clinical importance of these differences; and effect sizes, number-needed-to-treat; relation to clinically important differences or other statistical methods that can be directly related to clinical significance. |
| 1 | The relative differential treatment effects were quantified (means and confidence intervals), but the clinical significance of the treatment effect size is not specifically addressed. |
| 0 | There was no reporting of the differential treatment effect size between treatment groups (other than critical values or p values from statistical tests ). |
| 23. | 2 | 1) complete data collection was achieved on all subjects or  2) the rate of missing data (and whether it was random) was reported and where missing data occurred in more than 10% of cases, an appropriate imputation strategy was used. |
| 1 | Methods around missingness were not reported (none reported and a lack of imputation strategy), but given the experimental design missing data was not likely an issue (e.g. immediate pretest post-test or complete cases selected from a database etc.) |
| 0 | Missing data is a relevant concern and the protocol for handling missing data was not adequately described. Imputation was used and it is not clear what percentage of data was imputed or the strategies used; or more than 20 % of the data was imputed. |
| 24. | 2 | The study reports whether any adverse events occurred. Adverse events, costs and implementation criteria are considered when interpreting results. Practical issues such as specific training or equipment required to achieve the effects described in the study are specified. |
| 1 | Risks and benefits are considered, but without complete reporting of above criteria. |
| 0 | No reporting of adverse events, costs or practical issues. |
|  | | **Recommendations** |
| 25. | 2 | Specific conclusions and clinical recommendations are made for each of the specific objectives of the study. Specific recommendations are appropriate given the study findings with respect to risk-benefit and study limitations. Recommendations do not do any of the following 1. ignore observed results 2. overstate the generalizability or clinical impact of the study findings or 3. state that the treatment is ineffective - when there was insufficient power to establish this was the case. |
| 1 | Conclusions and clinical recommendations are incomplete, vague, or generalize to situations beyond those studied. |
| 0 | Conclusions or clinical recommendations were not founded on the results of the study or contradict findings of the study. |

© Joy C MacDermid 2013

Guidelines for Multiple Reviewers

1. Use the accompanying data collection sheets to extract content information from the study and your ratings of research design rigour on the 25 items above.
2. Raters should perform calibration reviews together until they feel comfortable that they understand the criteria, and have calibrated items where it may be necessary to further refine the criteria specific to the area being evaluated.
3. Raters can then independently review manuscripts and come together to compare scores. A process should be agreed upon for determining the final consensus score for each item.
   1. A potential Consensus Process Policy for Design Rigour Items is stated below (but reviewers can choose to use external reviewer for any disagreement).
      1. Reviewers will review their scores for all 25 items
      2. Differences of 2 points on the score for any item must be adjudicated so that they are minimized by consensus to a difference of 1 point or less for any given item.
      3. If the primary reviewers cannot agree to within 1 point, secondary reviewers will be used.
      4. Differences of 1 point will be adjudicated and an attempt made for reviewers to assign a score by consensus; if a consensus cannot be reached, then the lower score will be assigned.