



Literature Critique Criteria Tabular form for Randomized Clinical Trials

Criterion	Green	Yellow	Red	Comments
Randomization	Method of generation of an unpredictable randomization sequence clearly described (e.g., random number table, computer random number generator), including details of any restrictions (e.g., blocking, stratification)	Randomization is claimed, but method is not clearly	Not randomized	“Not randomized” includes allocation by chart number, date of birth, or other method which does not use an allocation list which is prepared by a random process generated by the investigators; however, minimization may be an acceptable alternative method of participant allocation
Concealment of allocation	Method of concealment of allocation list is adequately described	Concealment method is not clearly described	Not concealed	Concealment methods may include sequentially numbered opaque envelopes, allocation sequence kept in a central telephone location, etc.
Participant recruitment and eligibility	Clear designation of how participants were recruited (referral by primary care physician, self-	Recruitment or eligibility criteria vague or sketchy	Recruitment and eligibility criteria missing	Recruitment and eligibility criteria are applied before randomization; hence, they do not affect the



Criterion	Green	Yellow	Red	Comments
	referral, advertisement) and what was required for trial entry (clinical diagnosis, comorbid conditions, age, etc.)			internal validity of the study, but may limit its external validity; clear eligibility criteria are needed for the reader to decide if the results are applicable to a particular patient population
Blinding of patients and caregivers	Patients and caregivers are not aware of their treatment group until the end of the study	Patients or caregivers are likely to be aware of their treatment group before the study ends	Lack of blinding	Some interventions do not allow for blinding of patients or providers of care, and some degree of bias may be unavoidable
Blinding of assessors of outcome and of data analysts	Researchers who are measuring or assessing the outcome are unaware of the treatment group of the patient being assessed, and those who analyze the statistical results are also unaware	Blinding of assessors is possible, but not clearly described	Lack of blinding of either assessors or analysts	Blinding of outcome assessors and data analysts is feasible in many circumstances which do not permit blinding of patients and caregivers
Description of interventions	Both study and control interventions are described in sufficient detail to enable the reproduction of the intervention	Some aspects of the interventions are clear, but reasonable inferences may be made, as when the	Interventions are vaguely described, and the reader cannot make reasonable inferences about what	Judgment about the adequacy of the description of the interventions may require experience with the treatment



Criterion	Green	Yellow	Red	Comments
	in both arms of the study; time frame, intensity, frequency, and quantity of each intervention are reported	interventions are well standardized in general clinical practice	interventions were provided	modalities; e.g., for acupuncture, the needle types, depths of insertion, location, etc.; for physical therapy, the techniques and combinations of treatments
Information about care and intervention providers	Expertise, background, experience, and specific training are described (such variables as the learning curve involved in specialized surgical procedures, supervision of providers by providers, when appropriate)	The job titles of the providers are mentioned, but information about their training and experience is lacking	No information is given about who actually delivered the interventions being evaluated in the study	For non-pharmacologic interventions such as surgery or physiotherapy, it is useful to be told about the degree to which the care is done by people with specific skills and training which can influence the effectiveness of the interventions
Information about modes of delivery of interventions, especially when these interventions are non-pharmacological	Descriptions are given as to where the interventions are done (home, in a physiotherapy clinic, individually or in a group class), whether instructions are given in writing or face-to-face, whether the intervention is planned to be	Some information is provided concerning the ways in which the intervention was delivered, but some of the information is missing	Information is too sparse to enable the reader to know how to replicate the intervention, either for patient care or for planning additional research on the intervention	Interventions such as exercise which is done at home may have different effects from exercise done under supervision; standardized programs are not the same as those which are personalized or adapted to the circumstances of the individual



Criterion	Green	Yellow	Red	Comments
	tailored to the individual patient or standardized			patient
Participant follow-up	A flow diagram, accompanied by description in the text of the study, shows how many patients were recruited, were eligible, and enrolled in the study; after randomization, there is clear accounting for each group's attrition, the numbers of crossovers, the number completing the study, the number analyzed for each outcome, and reasons for attrition and exclusion from analysis	Some description of numbers of patients at each stage of the study, but lacking a flow diagram, or requiring effort on the part of the reader to determine the flow of patients through the stages of the study, with reasons for attrition or exclusion not described even though numbers are reported	Insufficient information to determine the flow of patients through the stages of the study	Especially important when there is significant attrition during the study, when there are crossovers from treatment groups initially assigned, or when patients are excluded from the analysis for reasons that are not apparent to the reader
Length of follow-up	Outcomes reported for more than one short-term measurement (once during and once at the end of the intervention period) and more than one long term	One short term and one long term outcome reported	Short term outcome only	



Criterion	Green	Yellow	Red	Comments
	measurement (e.g., several weeks and again several months after the intervention period)			
Baseline comparison	Tabular form clearly allows the reader to see the important variables at entry for each treatment group for potential known confounders (age, sex, symptom severity, symptom duration, number of previous interventions, etc.)	Partial description of baseline data, lacking tabular form, with some important variables not reported	Lack of description of baseline variables	Usually in Table I; p values are optional (since by definition all imbalances arose by chance), but it is useful if large chance imbalances are marked with an asterisk or other designation
Primary outcome	Clear designation of which outcome is regarded as the primary endpoint of the study, and at least one secondary outcome; there should be at least one symptom outcome and one functional outcome reported	Outcomes are reported for symptoms and for function, but it is not clear which was the primary outcome	Symptom outcomes are reported, but functional outcomes are not reported	It may be acceptable if a symptom (e.g., numerical pain score) is designated as primary, but a functional outcome is important as well
Analysis of	Intention to treat	As treated	Completers	Intention to treat



Criterion	Green	Yellow	Red	Comments
results	(patients analyzed in their original assigned treatment groups) is done for primary and secondary outcomes, with “as treated” outcomes reported when significant crossovers have occurred; sensitivity analysis is provided for “best case” and “worst case” scenarios for patients with missing data	analysis, with low attrition	only are analyzed	is expected to yield a conservative estimate of treatment effect, but preserves the randomization of the original allocation, and may give a more accurate estimate of the effectiveness of treatment in the real world
Rating scales for primary and secondary outcomes	The rating scale used to measure and assess outcomes has been previously published and has been tested in the population of interest, with measures of reliability (reproducibility) and validity (scores on the scale shown to be correlated with criteria known to be of clinical importance, and	A description of the rating scale has been published, but measures of reliability and validity are lacking	The scale has not been published and appears to have been developed by the authors of the study for application to the patients enrolled in the study	Unpublished outcome scales are likely to be biased and to yield an inflated estimate of the effectiveness of an intervention with respect to a clinical outcome of importance



Criterion	Green	Yellow	Red	Comments
	the studies showing validity and reliability are in the reference section			
Group comparisons between groups and not only within groups	Outcomes should be compared between groups in terms of between-group differences so that effect sizes with confidence intervals can be estimated	Between group comparisons are reported but confidence intervals for the differences are lacking	Only within-group effects are reported and these are used to support conclusions that there are differences between groups, or the authors report only p values for group comparisons	Between group differences cannot be inferred from within group effects alone, and these provide insufficient information to estimate how much one intervention differs from another
Adverse effects	Numbers of adverse events reported for all randomized participants both arms of the study, with separate data for each type of adverse event; participant withdrawals due to harms are reported for each arm; both absolute and relative risks of harm are compared for each arm; active and passive surveillance of harms are	Adverse events are reported, but presented as the total numbers of all events without separate data for each type of event; efforts at active surveillance not reported as such; when laboratory values are reported, only means or medians are reported	Generic statements such as “generally well tolerated” are used without numerical data, or adverse events are not reported	



Criterion	Green	Yellow	Red	Comments
	reported; for adverse effects having laboratory values, means, standard deviations, and extreme values are reported			
Attrition	Follow-up is close to complete (90% or more in each treatment arm) at the end of the study period	Follow-up is high (80-90%) at the end of the study period	Follow-up is less than 80% at the end of the study period	Attrition should be approximately equal in each treatment arm; differential attrition requires explanation supported by reliable data
Co-interventions (performance bias)	All interventions, including those in addition to the study intervention, are clearly reported and are the same in both groups	Co-interventions may have been equal, but this is not clearly stated	Co-interventions are likely to have been different in the treatment arms	Blinding of caregivers is expected to protect against performance bias
Presentation of outcome data	All outcomes which have numerical distributions are presented with actual numbers in tabular form, or in the text of the article, with means and standard deviations	Some outcomes presented with actual numbers in tables or the text, and some outcomes are presented with figures or graphs only, but the graphs are given with error bars which may allow the reader to infer the standard	All outcomes are presented in graphs and figures, without numerical tabulation, or with p values as the only numerical data, or graphs are given without error bars	It is not possible to extract numerical data by visual inspection of graphs and figures; actual numbers are needed; graphs are a supplement to, not a substitute for, numerical data. Error bars may allow the reader to infer the standard



Criterion	Green	Yellow	Red	Comments
		deviations		deviations, but this places an additional burden on the reader
Sample size and precision of results	Sample size for the study is explained, with the effect size of interest, the type I and type II error, and anticipation of attrition; effect size is given with estimate of statistical uncertainty (e.g., 95% confidence intervals)	Effect measure is reported with appropriate confidence intervals; power is not reported, but can be calculated from the reported results	Sample size is not discussed, and power cannot be calculated from the reported results	Success in recruiting and retaining desired sample size may depend on circumstances beyond the control of the researchers; this is more important for “negative” studies whose interpretation requires knowing whether they were adequately powered to detect a treatment effect
Dose-response relationships	When different doses of a drug are administered, there is data showing the response rates for each dose level of the drug, with adverse and therapeutic responses reported for each dose	Dose-response relationships are reported for therapeutic responses but not for adverse effects	Dose-response relationships are not reported	Small numbers may preclude reporting precise dose-response relationships, but when there are sufficient numbers of participants at each dose level, this is essential information
Dose titration and rescue medication	Details of dose titration (starting dose, rate of increase,	Some dosing information is given, but titration and	Dosing and rescue medication not explained	Flexible and fixed dose studies may show different



Criterion	Green	Yellow	Red	Comments
	maximum dose), fixed vs. flexible dosing, rescue medications are reported	rescue medications not clearly specified	or poorly defined	dose-response relationships, depending on whether the highest dose is a consequence of the randomization or a consequence of patient response to the starting dose
For studies with enriched enrolment and randomized withdrawal designs, the drug taper at the beginning of the double-blind is slow enough to prevent loss of blinding through inadvertent production of withdrawal symptoms	Tapering over at least one week	Tapering over less than one week	No tapering	The period of time required for tapering of a drug which has been fully titrated during the pre-randomization phase of the trial may depend on the specifics of the drug being studied, but one week is a minimum acceptable taper
Sponsorship and funding	Source of funding is identified, and competing interests (stock ownership, royalties, etc.) of authors are declared, when present; the authors have control of all the study data	Funding source identified, but unclear declaration concerning competing interests; the authors have control of all the study data	Sponsor not identified, no declaration concerning competing interests; the authors do not have control of all the study data, but some of the data is controlled by another party	Major journals routinely require declarations for conflicts of interest; however, current disclosure practices are likely to be less than completely transparent
Selective	There is an	The protocol is	The protocol	Clinicaltrials.gov



Criterion	Green	Yellow	Red	Comments
outcome reporting (protocol availability)	identifier of the trial protocol at clinicaltrials.gov or other public database, and the primary and secondary outcomes reported in the study are done in the way that was specified in the protocol	available, but there appear to be changes in the outcome reporting which are not identified at the public database; however, the published report does not appear to consist of data-driven analyses	is not available, or the study appears to suggest that some of the outcome reporting was data-driven, which creates a risk of bias arising from selective outcome reporting	is a useful database for the identification of primary and secondary outcomes, but the method of data analysis is often not included in the protocol
Baseline symptoms	For all treatment groups, baseline levels were sufficiently high to enable the trial to measure a difference between pre-treatment and post-treatment levels	Baseline levels likely to be too low to enable the trial to demonstrate a difference between pre-treatment and post-treatment levels	Baseline levels unclear or not reported	If there is an insufficient level of pain or disability at the beginning of the study, it may not be possible to measure a 30% or 50% difference between pre-treatment and post-treatment levels of the symptom
Credibility of reported effect sizes	Treatment differences between groups are within the bounds of credibility, when considered in the context of usually reported effect sizes	Treatment differences are outside the generally expected range of what is usually reported for similar interventions (for example, there is nearly complete success in the	Treatment differences are too large to be credible considering what is known about the usual clinical course of the condition and what is reported by all other studies of similar	Occasionally, a paper may inadvertently report a standard error as if it were a standard deviation, creating an impression that the two treatment groups are separated by several standard deviations when



Criterion	Green	Yellow	Red	Comments
		<p>experimental group and nearly complete failure in the control group for a condition which tends to improve over the course of time and where most studies show more modest treatment effects</p>	<p>interventions for similar conditions; for example, the p value for the effect size is so large as to be for all practical purposes impossible</p>	<p>other studies report that treatment groups differ by one standard deviation or less; if a p value can be calculated and is found to be astronomically low, the results are so highly suspect as to be considered invalid</p>
<p>Crossover trials</p>	<p>Authors report the duration of each treatment period, the duration of the washout period, and report on treatment effects, period effects, and carryover effects (if observed). The half-life of the intervention should be mentioned and documented so that washout periods can be justified.</p>	<p>Treatment effects are reported, but the authors omit mention of either the period effect or the carryover effect, and the half-life of the intervention is not documented</p>	<p>Treatment effects are reported, but there is no description of carryover or period effects</p>	<p>Crossover trials may be affected not only by the effects of the study treatments, but also by the order in which treatments are given (period effects) and by persistence of the first treatment during the second treatment administration. If a carryover effect is reported, the direction of bias is likely to underestimate the treatment effect of the experimental intervention. The intervention should not entail</p>



Criterion	Green	Yellow	Red	Comments
				training or learning because an effective treatment should not have a short half-life and a crossover design is not suitable for such interventions
For nonrandomized cohort studies with accurate measurement of treatment and outcome, and adjustment for measured confounders, a large treatment effect is observed	The ratio of successful outcomes in the treated and control groups is greater than 5	The ratio of successful outcomes in the treated and control groups is greater than 2	The ratio of successful outcomes in the treated and control groups is less than 2	Although residual confounding from unmeasured confounders may introduce bias into the treatment effect, the magnitude of this bias is generally bounded, rarely exceeding 5
For nonrandomized cohort studies, there is a clear dose-response gradient, especially if there is a rapid response to treatment	Several different levels of dose are reported, with a clear trend in the response rate	Several different levels of dose are reported, with a plausible but equivocal dose-response gradient	Dose-response gradients are unreported, or there is no relationship between different doses and different responses	Dose-response gradients are accepted as one element of a causal relationship in observational epidemiology
For nonrandomized studies, adjustment for plausible confounders are expected to increase confidence in	Patients in the treatment group are clearly sicker than patients in the control group, but still fare better in the outcomes of	Patients in the treatment group have some prognostic indicators which are worse than the control group,	Plausible confounders either clearly favor the treatment group, or tend to favor the treatment group	The direction of expected confounding is always an important consideration in the interpretation of observational studies



Criterion	Green	Yellow	Red	Comments
the treatment effect	treatment	and others may be better than the control group		
Medical and biological plausibility and coherency	Principles of action of the intervention are clearly mentioned and are consistent with the pathophysiology of the condition, preclinical data from in vitro, cadaver, or animal studies, and principles of pharmacology, biomechanics, etc.	Principles of action of the intervention may be consistent with general biomedical principles, but the proposed biological action of the intervention is not discussed	Principles of action are not clear, preclinical studies from animal studies have not been done, or action of the intervention is not consistent with general biomedical knowledge	It is sufficient if the reference list includes articles which present the biomedical principles and cite preclinical studies
Adherence to treatment	The authors report the numbers of patients allocated to an intervention, and also report the numbers of patients who adhered to the allocated treatment for each group at each followup period	Numbers of patients allocated to intervention is reported, and adherence to treatment is reported for the study as a whole, but not for each followup period	Adherence data are not provided for either treatment group	Adherence is an important part of treatment effectiveness, and has a direct bearing on whether it is likely to be acceptable in a real world setting. If adherence is low, this can provide useful evidence against making a recommendation for its use. When an intervention requires a high degree of patient participation, data on



Criterion	Green	Yellow	Red	Comments
				adherence are necessary for interpretation of the study.

TBI-2019