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-referral, advertisement)	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 internal validity of the study, but

Criterion	Green	Yellow	Red	Comments
	and what was required for trial entry (clinical diagnosis, comorbid conditions, age, etc.)			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
Blinding success	Participants are asked to guess which treatment they received, the percentage of correct guesses is recorded, and is compared to what is expected by chance	Participants are asked to guess which treatment they received, but there is no comparison with what is expected by chance	No mention of whether participants were asked to guess their treatment assignment	Useful to help reader assess how well the blinding worked, especially when there is reason to suspect that the physiologic effects of an intervention will be apparent;

Criterion	Green	Yellow	Red	Comments
				however, this should be applied cautiously, since unblinding by efficacy may occur, and is not a source of bias
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	One short term and one long term outcome reported	Short term outcome only	

Criterion	Green	Yellow	Red	Comments
	period) and more than one long term 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	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

Criterion	Green	Yellow	Red	Comments
	reported			
Analysis of results	Intention to treat (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	As treated analysis, with low attrition	Completers only are analyzed	Intention to treat 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
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	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
	and passive surveillance of harms are 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	All outcomes are presented in graphs and figures, without numerical tabulation, or with p values as the only numerical data	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
Sample size and precision	Sample size for the study is	Effect measure is reported with	Sample size is not discussed,	Success in recruiting and

Criterion	Green	Yellow	Red	Comments
of results	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)	appropriate confidence intervals; power is not reported, but can be calculated from the reported results	and power cannot be calculated from the reported results	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
Description of interventions	Both study and control interventions are described in sufficient detail to enable the reproduction of the intervention in both arms of the study; time frame, intensity, frequency, and quantity of each intervention are reported	Some aspects of the interventions are clear, but reasonable inferences may be made, as when the interventions are well standardized in general clinical practice	Interventions are vaguely described, and the reader cannot make reasonable inferences about what interventions were provided	Judgment about the adequacy of the description of the interventions may require experience with the treatment modalities; e.g., for acupuncture, the needle types, depths of insertion, location, etc.; for physical therapy, the techniques and combinations of treatments
Psychosocial variables	Baseline and follow-up descriptions of emotional and social functioning including scores on at least one validated scale	Psychosocial variables mentioned, but without details concerning diagnoses or measurements of function	Psychosocial variables lacking	Pertinent for many spine conditions, in which part of the prognosis and response to treatment are affected by these factors

Criterion	Green	Yellow	Red	Comments
	for pertinent diagnoses (e.g., Beck Depression Inventory, Profile of Mood States, SF-36 Mental Health and Role Emotional subscales, etc.)			
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
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
Protocol availability	There is an identifier of the trial protocol at clinicaltrials.gov or other public database, and	The protocol is available, but there appear to be changes in the outcome reporting	The protocol is not available, or the study appears to suggest that some of the	Clinicaltrials.gov is a useful database for the identification of primary and secondary

Criterion	Green	Yellow	Red	Comments
	the outcomes reported in the study are done in the way that was specified in the protocol	which are not identified at the public database; however, the published report does not appear to consist of data-driven analyses	outcome reporting was data-driven	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
Crossover trials	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)	Treatment effects are reported, but the authors omit mention of either the period effect or the carryover effect	Treatment effects are reported, but there is no description of carryover or period effects	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
For nonrandomized cohort studies with accurate measurement	The ratio of successful outcomes in the treated and control groups	The ratio of successful outcomes in the treated and control groups	The ratio of successful outcomes in the treated and control groups	Although residual confounding from unmeasured

Criterion	Green	Yellow	Red	Comments
of treatment and outcome, and adjustment for measured confounders, a large treatment effect is observed	is greater than 5	is greater than 2	is less than 2	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 the treatment effect	Patients in the treatment group are clearly sicker than patients in the control group, but still fare better in the outcomes of treatment	Patients in the treatment group have some prognostic indicators which are worse than the control group, and others may be better 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
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,	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	It is sufficient if the reference list includes articles which present the biomedical principles and cite preclinical studies

Criterion	Green	Yellow	Red	Comments
	and principles of pharmacology, biomechanics, etc.		knowledge	