Supplement 1: Interpretation of Modified Coleman Criteria

|  |  |
| --- | --- |
| Inclusion criteria | Enrollment rate was interpreted as the percentage of patients evaluated for inclusion and enrolled in the randomized study.n |
| Power | “Methods described” was interpreted as any explanation for how a given power level was determined, such as discussion of historical incidence rates. |
| Blinding | Studies which were blinded only to outcome assessors (i.e. radiologists) were designated single-blind, as patients would be aware of treatment and potentially introduce bias in reporting symptoms or seeking care. Double-blind studies were those studies blinded to both outcome assessors and patients. One triple-blind study (Selby 2015) specifically reported blinding a steering committee, as well. |
| Similarity in treatment | Similar or no co-interventions was interpreted as studies reporting which treatments were permitted in addition to the study arm; those studies which did not report were assumed to be heterogeneous. |
| Group comparability | Partially comparable was interpreted as any statistically significant differences (p>0.05) or reported heterogeneity between the study arm populations. |
| Outcome assessment | If the outcome assessment was based on predetermined or confirmatory imaging, it was interpreted as “independent investigator” (4 points); if solely based on clinical symptoms, it was interpreted as “recruited patients” (6 points). |
| Clinical effect measurement | Studies which reported effect size typically used odds ratio (OR) or relative risk (RR). An OR or RR of 0.2 (or 5.0) was interpreted as an effect size of 80%. As with other categories, failure to report explicitly received no points. |