Appendix B

Table 3. Assessment of internal validity within systematic reviews

| Reference | Primary Outcomes | Secondary Outcomes | Method/Instrument Used to Assess Internal Validity | Overall Internal Validity Rating |
| --- | --- | --- | --- | --- |
| Bodri et al. 201119 | Recipient ongoing pregnancy rate per randomized donor | Number of retrieved oocytes, duration stimulation, total gonadotropin consumption, and OHSS) incidence per randomized oocyte donor.  | Using components of study design that are related to internal validity (Center for Reviews and Dissemination, 2001): randomization, allocation concealment, blinding, ITT, follow-up. | Quality of included trails was generally good |
| Groeneveld et al. 201111 | Safety: mortality, morbidity, bleeding and impaired coagulation, acute kidney injury, edema, hypoalbuminemia, pruritus, and anaphylactoid reactions | Not reported | Study quality and reliability for assessment of safety were judged according to several factors: randomized study design, size of patient population and statistical power to evaluate safety endpoints, colloid dose and demonstration of a dose-response relationship, adequacy of follow-up period, sensitivity of employed diagnostic methods for detecting complications, type of control fluid used as a comparator for safety, co-morbidities and severity of illness as indicators for the likelihood of observing complications, adherence to an a priori analysis plan, and multivariate analysis with adjustment for potential confounding factors.  | Not reported |
| Hockenhullet al. 201120 | Death, major adverse cardiac and cerebrovascular events, or other major adverse events  | Acute myocardial infarction, target vessel revascularization, target lesion revascularization, repeat treatment, and thrombosis | Used methods proposed by the Heart Collaborative Review Group and grading similar to that used in Villanueva, which considers the following: adequacy of randomization, allocation concealment, potential for selection bias, and adequacy of masking. | The quality of most of the included studies was rated as B due to lack of adequacy of allocation concealment and masking/blinding. According to the authors, the reporting of the use of intention-to-treat analysis was very good across all trials.  |
| Seitz et al. 201121 | Change in symptoms of agitation and psychosis in dementia as measured on the various dementia NPS scales, drop-up due to adverse events | Changes in total scores for dementia NPS scales, changes on the CGI scale, changes in cognitive impairment scores, and caregiver stress; falls, headache, gastrointestinal upset, worsening of dementia, anxiety, headache, bleeding extrapyramidal symptoms, and hyponatremia | Risk of bias assessment provided by the Cochrane Collaboration: sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data, selective outcome reporting, other sources of bias. | Not reported |
| Beauchamp et al. 20105 | Peak power and peak oxygen uptakemeasured during an incremental exercise test on cycle ergometer or treadmill; endurance time measured from a constant power test; functional exercise capacity measured by 6MWT or 12MWT; HRQoL measured by CRQ; anxiety and depression measured by HAD | Lactate threshold, isotime ventilation, heart rate, breathing frequency and symptoms | Jadad (0 to 5 scale): randomization, blinding, withdrawals; PEDro 10 point scale: blinding, randomization, withdrawals, comparability of baseline characteristics, data reporting  | Jadad – 2/5 (range 1–3) |
| Davis et al. 201022 | Clinical outcomes (e.g., scales that measure functional improvement), patient satisfaction, complications, length of hospital stay, and return to work | Not reported | Not reported | Not reported |
| Dibra et al. 201016 | All-cause death, recurrent myocardial infarction, reintervention, and stent-thrombosis | Not reported | Trials were evaluated for adequacy of allocation of concealment, intent-to-treat analysis, and blind assessment of outcomes. The authors reported using the criteria of Altman et al. and Juni et al. to assess adequacy of allocation of concealment. | The authors indicate that the main limitation was the absence of blinding of outcome assessors. |
| Dong et al. 201023 | Angiographic: Rates of TIMI grade 3 flow before and after PCI, Myocardial perfusion evaluated by cumulative ST-segment resolution in postprocedural electrocardiograms Clinical: 30-day and 8-month rates of mortality and reinfarction Safety: Major and minor bleeding complications according to the criteria of the TIMI trial | Not reported | Not reported | Not reported |
| Dong et al. 201024 | Angiographic: combined TIMI grade 2 and 3 on the initial angiogram; preprocedural and post-procedural TIMI grade 3 flow were also assessed; pre and post procedural myocardial perfusion evaluated by TMBG 3; Clinical: mortality at 30 day follow-up; incidence of reinfarction evaluated as another clinical outcome of interest Safety: major bleeding complications | Not reported | QUOROM guidelines for meta-analysis. Evaluated studies for the adequacy of allocation concealment, performance of the analysis according to the intention-to-treat principle, and blind assessment of the outcomes of interest. We used the criteria recommended by Altman and Schulz and Juni et al. to decide whether treatment allocation was adequately concealed. | No summary score used |
| Dubicka et al. 201018 | Depression and impairment scores, overall improvement, suicidality, and adverse events | NR | Used a method based on the authors of other systematic reviews on similar topics. Nine features of study quality were rated on a scale of 0 to 3, with a maximum score of 27. The 9 features included: method of randomization, intent-to-treat analysis, blinding of outcome assessors, blinding of patients, description of improvement, use of multiple outcome assessors, description of treatment dosage, use of manualized therapy, assessment of therapy adherence, and assessment of adherence to medication. | Mean quality score was 21, range 18 to 24. |
| Fuentes et al. 201025 | Pain measured by the VAS or numeric pain rating scale  | Not reported | 7 scales used: Delphi List, PEDro, Masstricht, Maastricht-Amsterdam List, Bizzini, van Tulder, and Jadad compiled in a set of 39 items. Categories included: patient selection, blinding, intervention, outcomes, statistics. | Not reported |
| Hong et al. 201026 | Microvascular events, mortality | Not reported | Quality of each selected trial evaluated by means of empirical evidence: randomization, allocation concealment, blinding. | Not reported |
| Hughes et al. 201027 | Disability rating | Recovery of unaided walking, recovery of walking with aid, discontinuation of ventilation, mortality, death or disability, treatment related fluctuation, and adverse events | Used methods described in the Cochrane Handbook, which considers sequence generation, allocation concealment, blinding, completeness of follow-up, freedom for selective reporting and other sources of bias. | Moderate quality evidence |
| Kalil et al. 20102 | Clinical cure | Microbiological eradication, methicillin-resistant, adverse events | Used the Jadad scale QUOROM guidelines to assess study quality. | Mean quality score according to the Jadad scale was 3.3, range 3 to 4. |
| Kesselheim et al. 20103 | Number or severity of seizures | NR | Used the Jadad scale to assess quality of RCTs and the Newcastle-Ottawa scale for assessing non-randomized trials. | Mean quality score 2.7, range 2 to 4 |
| Krenke et al. 201017 | Length of hospital stay and failure rate | Duration of chest tube drainage, duration of fever, duration of respiratory distress, and volume of pleural fluid drainage | The quality of studies was assessed for the following strategies: allocation concealment, blinding, intention-to-treat analysis, and completeness of follow-up Authors did not report using a specific assessment scale or checklist. | The two studies that made up the comparative treatment evidence base used methods to conceal allocation, did not report blinding, used an intent-to-treat analysis, and adequate follow-up.  |
| Lanitis et al. 201028 | Local recurrence, distant relapse, disease-free interval, severe postoperative complications, other outcomes (quality of life and cosmetic satisfaction) | Adverse events, complications | Since the studies used in the review were all non-randomized, the Newcastle-Ottawa Scale was used to assess study quality. Studies achieving 6 or more stars were considered to be of higher quality.  | 7 studies scored 6 or more stars on the modified Newcastle-Ottawa scale.  |
| Lee et al. 201029 | Number of patients that: experienced (1) full remission; (2) partial remission; (3) overall remission; (4) relapse; (5) treatment failure; (6) end stage renal disease; or (7) died.  | Number of patients who experienced a side effect | Used the Jadad scale to assess study quality. | Studies of induction therapy (MMF vs. CYC) Jadad scores ranged from 1 to 3; Studies of maintenance therapy (MMF vs. AZA) Jadad scores were 1 and 2; and for studies of high-dose vs. low-dose CYC scores were 2 and 2. |
| Liu et al. 201030 | Survival (during 1-year follow-up, during 3-year follow-up, and up to the end of the follow-up period) and recurrence (during 1-year follow-up, during 3-year follow-up, and up to the end of the follow-up period) | Not reported | No specific instrument reported. The authors indicate that study quality was measured according to the non-randomized controlled clinical trial quality evaluation standard. | Quality scores for the included studies ranged from 7 to 9. |
| Liu et al. 201031 | Operating time, complications, death rates, time of hospital stay, time to return to normal activities, and time to return to normal diet | Overall cost | Used the Jadad scale to assess study quality. | The mean Jadad score of the included studies was 4.25. The main limitations included: sample size, allocation concealment, and double blinding. |
| Liu et al. 201032 | Complications, death rates, survival rates, recurrence-free survival rates, and recurrence | Not reported  | Used the Jadad scale to assess study quality. | The mean Jadad score of the included studies was 3. The quality of the studies was limited in terms of sample size, allocation concealment, and double blinding. |
| Liu et al. 201033 | OS, RFS, pelvic control rate | Rates of local and distant recurrence, complications | Juni quality assessment criteria for RCTs: randomization method, blinded assessment of outcomes, allocation concealment, losses to follow-up, ITT. | Not reported |
| Macedo et al. 201034 | Pain, disability, global perceived effect, and return to work | Not reported | PEDro scale (0-10): masking, baseline comparability, allocation concealment, ITT, adequate follow-up | Median=6 (range 3 to 9) |
| Machadoet al. 201035 | Remission rate defined as scores<7 or 8 and <10 or 12 for the HAM-D and MADRS scales, respectively, and measured at 8-12 weeks of treatment.  | Not reported | Downs-Black, 27-item quality assessment checklist. Categories included: study design, sample selection, data presentation, statistical analysis and statistical power. | All studies reported quality above 80% |
| Meier et al. 20106 | Survival to hospital discharge,  | Return to spontaneous circulation, favorable neurologic outcome at discharge, and long-term outcome (survival at 1-year) | Used the Jadad scale to assess study quality | The mean Jadad score of the included studies was 4. The quality of the studies was limited in terms of lack of double blinding. |
| Milito et al. 201036 | Healing | Operative time and hospitalization, pain, analgesic requirements, blood loss, wound healing, convalescence period, postoperative continence impairment, anal stenosis and relapse, and cost effectiveness | No specific method reported; discuss the following aspects of quality of RCTs: allocation of concealment, blinding, mean outcome measures, statistical methods, and length of follow-up. | Not reported |
| Murphy et al. 201037 | VAS for pain, postoperative nausea and vomiting, pruritus | Not reported | Not reported | Not reported |
| Myers et al. 20109 | Pain measured using a validated scale | Not reported | No specific method reported; discuss the following aspects of quality of RCTs: publication status (full publication or meeting abstract), allocation of concealment, blinding, statement of statistical power or sample size calculation, intention-to-treat analysis, and statement of funding/sponsorship. | Not reported |
| Pan et al. 201038 | Mortality, recurrent myocardial infarction, repeat revascularization, and stent thrombosis | Not reported | Not reported | Not reported |
| Riemsa et al. 20108 | Overall survival, progression free survival, time to progression, response rate, complete response, response duration, stable disease, and quality of life | Adverse events | Used the Cochrane Collaboration quality assessment checklist.  | The authors state that overall all 4 studies included in analysis had a low risk of bias. |
| Sbruzzi et al. 201039 | Functional capacity (measured by peak oxygen uptake | Distance of 6-min walk test and muscle strength | Used the Jadad and PEDro scales and specifically considered concealment of allocation, intention to treat analysis, baseline comparability, blinding of outcome assessors, and description of losses and exclusions | The quality of the studies was poor, all studies received a Jadad score ≥ 3 and 5 studies received a PEDro score of ≥ 5 (out of 10).  |
| Sgourakiset al. 201040 | Patients requiring re-intervention, reflux, complications, procedural death, and overall survival. | Not reported | Used the Jadad scale to assess study quality. | The mean Jadad score of included studies was 2.7, range 1 to 4. |
| Simpsonet al. 201041 | Survival to hospital discharge (overall), survival to hospital discharge (response time < 5 minutes), survival to hospital discharge (response time > 5 minutes) | Not reported | Internal validity assessment performed using methodology recommended by Cochrane Collaboration: risk of bias across the following domains – sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data, selective outcome reporting, other potential threats to validity | Not reported |
| Squizzatoet al. 201012 | Visual acuity, neovascular complications, recurrent events, bleeding complications | Not reported | Jadad scale: randomization, blinding, follow-up | Not reported |
| Sunkara et al. 201042 | Ongoing pregnancy/live birth rate | Post-thaw blastocyst survival rate, clinical pregnancy rate, and miscarriage rate | Newcastle-Ottawa Quality Assessment Scales: selection of cases and controls, study group comparability, exposure to intervention and treatment outcome. | Not reported |
| Tamayo et al. 20107 | Response rates, remission rates, discontinuation rates due to adverse events, lack of efficacy, or discontinuation due to any cause, NNT or NNH | Not reported | Jadad scale | Not reported |
| Tamhaneet al. 201043 | Clinical: Death, stroke, TVR, reinfarction. Myocardial perfusion Angiographic: post procedural rates of TIMI grade 3 flow and TMGB  | Not reported | Allocation concealment, study design, ITT, blinding assessment of outcome measures | Did not use a quality score |
| Tang et al. 20104 | Success (complete healing or incomplete healing) and failure (uncertain healing or unsatisfactory healing)  | Not reported | Assessed quality based on the following factors: RCT, control, double-blinding, allocation of concealment, description of withdrawals and dropouts, sample size predetermined, intent-to-treat, operator experience reported, treatment procedures described, measurements standardized, and evaluation methods clearly described. Studies with low risk of bias graded A, moderate risk of bias B, and high risk of bias graded C. | RCTs rated as A-low risk of bias |
| Testa et al. 201044 | Combined rate of MAE, defined as cumulative risk of all cause death and nonfatal acute myocardial infarction, TVR and TLR, rate of stent thrombosis | Not reported | Cochrane Collaboration Newcastle-Ottawa scale for assessing quality of cohort study: sequence generation, allocation concealment, blinding, selective reporting, and incomplete data. | Registries overall low quality RCTs overall good quality |
| Valachiset al. 201045 | OS | Number of local recurrences, true and elsewhere breast recurrences, axillary recurrences, supraclavicular recurrences, and distant recurrences | No specific method reported; discuss the following aspects of quality: method of randomization, allocation concealment, intent-to-treat analysis, and patient withdrawal | The authors reported that 2 trials described the model of randomization and the model of allocation of concealment. |
| Vasiliadiset al. 201015 | Lysholm score, Tegner score, Modified Cincinnati score, VAS, Meyers score, Stanmore scores, SF-36 scores, repair tissue evaluation and histological assessment, complications, post-operative clinical improvement | Not reported | Cochrane Handbook for Systematic Reviews of Interventions: selective reporting, baseline comparability. | Overall the quality of evidence can be rated as average to low |
| Vermeulanet al. 201010 | Bacterial load and wound healing | Adverse events, cost, and length of hospital stay | No specific method reported | Authors indicate that study quality was limited, mainly because of lack of concealment of randomization, use of quasi-randomization procedures, not reporting if performed an intent-to-treat analysis, not reporting if outcome assessor were blinded, not using independent outcome assessors, and not reporting funding source. |
| Xie et al. 201046 | Survival, recurrence, DFS, additional treatment, complications, hospitalization, patients’ attitudes toward treatment options, cost analysis | Not reported | Not reported | Not reported |
| Yang et al. 20101 | Remission rate of pain and incidence of opioid-related adverse effects | Quality of life | Used a quality checklist adapted from MOOSE standard, which includes the following 6 measures: prospective study design, group comparability on confounding factors, blinding of outcome assessors, length of follow-up, relation between outcome and exposure appropriately measured, and used appropriate statistical analysis | Not reported |
| Agarwal et al. 201047 | 30 day all-cause mortality | Functional status, reinterventions, pacemaker insertion, ventricular arrhythmias, cardiac dimensions, mitral regurgitation, systolic anterior motion of mitral valve, length of hospital stay, and exercise tolerance | Not reported | Not reported |
| Avouac et al. 201048 | Pain and functional status | Not reported | Used the Jadad scale to assess study quality. | Jadad scores ranged from 1 to 5. |
| Chua et al. 201014 | Improved DFS | Overall survival, morbidity, mortality, pathological response, pattern of recurrence | Not reported | Not reported |
| Devaiah et al. 201049 | Relapse of subjective vertigo at follow-up | A negative Dix-Hallpike maneuver | Not reported  | Not reported |
| Lovemanet al. 201013 | Time to disease progression, progression-free survival, response rate, response duration, overall survival, symptom control, health-related quality of life, cost-effectiveness, adverse effects | Not reported | Criteria recommended by the Center for Reviews and Dissemination (CRD): randomization, allocation concealment, baseline characteristics, eligibility, blinding assessors, blinding care provider, patient blinding, reporting outcomes, ITT, withdrawals explained.  | Not reported |
| Valachiset al. 201050 | OS, time to tumor progression, proportion of patients with complete or partial response after treatment (objective response), and proportion of patients with an objective response or stable disease lasting ≥ 24 weeks (clinical benefit) | Adverse events | Recorded the following methodological quality items: mode of randomization, allocation concealment, subject withdrawals, blinding, if interim analysis was planned or performed, and if intent-to-treat analysis performed | No overall rating reported; 3 of 4 trials were double blind, 1 study reported mode of randomization and methods for ensuring allocation concealment, and no study was stopped early because of statistically significant differences in an interim analysis |

AKI Acute kidney injury

AZA Azathioprine

CGI Clinical global impression

CRD Center for reviews and dissemination

CRQ Chronic Respiratory Questionnaire

CYC Cyclophosphamide

DFS Disease-free survival

HAD Hospital Anxiety and Depression scale

HAM-D Hamilton Depression rating scale

HRQoL Health-related Quality of life

ITT Intention-to-treat

MAE Major adverse events

NNH Number needed to harm

NNT Number needed to treat

NPS Neuropsychiatric symptoms

NRS Numeric pain rating scale

OHSS Ovarian hyperstimulation syndrome

OS Overall survival

PEDro Physiotherapy evidence base database

RFS Relapse-free survival

TIMI Thrombolysis in Myocardial Infarction

TLR Target lesion revascularization

TMBG TIMI myocardial blush grade

TVR Target vessel revascularization

VAS Visual analog scale