Supplementary Material (5)
RISK OF BIAS ASSESMENT

1- ROBINS-I TOOL FOR NON-RANDOMISED STUDIES

Study: Tabanli et al., 2024

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	"Although patients were "randomly divided" into three groups, there is a statistically significant difference in age between groups (mean age in ACDF is 55.6 ± 7.9 years versus 41.0 ± 7.9 years in CDA and 46.5 ± 9.1 years in HS, p=0.039). The authors stated that they conducted sensitivity analyses that included age, baseline severity of symptoms, and other clinically relevant factors to determine whether any specific subgroup was disproportionately affecting the variability. Although these sensitivity analyses help mitigate the confounding issue, the imbalance in age remains a concern inherent to retrospective designs."	Moderate risk
Domain 2 (Selection of Participants)	Patients were selected retrospectively from a consecutive series that met strict inclusion and exclusion criteria. The study describes a systematic selection process minimizing selection bias.	Low risk
Domain 3 (Classification of Interventions)	The surgical interventions (ACDF, CDA, and HS) are clearly defined	Low risk
Domain 4 (Deviations from Intended Interventions)	All patients received the intended surgical procedures according to standardized protocols with no reported deviations or crossovers	Low risk
Domain 5 (Missing Data)	The study includes a complete dataset of 120 patients with available follow-up information. There is no evidence of substantial loss to follow-up	Low risk
Domain 6 (Measurement of Outcomes)	Outcomes, including clinical scores (NDI, VAS) and radiographic parameters (ROM, sagittal alignment), were measured using validated and standardized methods. Although blinding of assessors is not explicitly mentioned, independent evaluation helps minimize bias. Study stated that: "Clinical and radiological outcomes were independently evaluated by another surgeon who was not involved in surgeries"	Low risk
Domain 7 (Selection of the Reported Result)	The study reports all pre-specified outcomes across clinical and radiological domains without evidence of selective reporting	Low risk
Overall Risk		Moderate risk

Study: Hung et al., 2018

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	Important baseline differences existed between groups and were not fully controlled. For example, patients in the ACDF (fusion) group were significantly older on average than those in the other groups. The study did not report any statistical adjustment or matching for these differences. These uncontrolled confounders likely affect outcomes.	Serious risk
Domain 2 (Selection of Participants)	Retrospectively, participants were selected from those undergoing three-level surgery at a single hospital between 2012 and 2015, based on inclusion criteria (multilevel degenerative disease with radiculopathy/myelopathy) and exclusions (e.g. trauma, prior surgery). Importantly, selection into the study was not based on outcome – cases were included based on having received a qualifying intervention and meeting criteria, not on their results. However, because this was not a truly consecutive series (described as "retrospective case selection") and the authors do not clarify if any eligible patients were omitted (e.g. lost to follow-up), there is some uncertainty and this domain is judged as moderate risk.	Moderate risk
Domain 3 (Classification of Interventions)	Clear, pre-defined intervention groups without outcome-dependent misclassification	Low risk
Domain 4 (Deviations from Intended Interventions)	Patients received the intended surgery and were analyzed as treated. As an observational comparison, each patient simply underwent the surgical approach decided by the surgeon; there were no protocol deviations or cross-overs reported	Low risk
Domain 5 (Missing Data)	Complete follow-up with outcomes reported for all groups	Low risk
Domain 6 (Measurement of Outcomes)	Considering the unblinded assessment of clinical outcomes, we rate this domain as Serious risk. "Clinical characteristics in terms of the Japanese Orthopedic Association (JOA) score, Neck Disability Index (NDI), and visual analogue scale (VAS) scores were evaluated before and after surgery by independent assessors who were not blinded to the type of surgery"	Serious risk
Domain 7 (Selection of the Reported Result)	All key outcomes appear reported with no evidence of selective reporting	Low risk
Overall Risk		Serious risk

Study: Wang et al., 2018

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	Although the study reports that demographic and baseline characteristics are comparable between groups, the retrospective design inherently limits full control over all potential confounders. Important prognostic factors (e.g., severity of degeneration, bone quality, comorbidities) may not have been completely measured or adjusted for.	Moderate risk
Domain 2 (Selection of Participants)	The study used a systematic, consecutive series of patients meeting strict inclusion and exclusion criteria over a defined period. This minimizes the risk that patients were selected based on outcome-related factors, even though the study is retrospective.	Low risk
Domain 3 (Classification of Interventions)	The three interventions (ACDF, CADR, and HS) are clearly defined. Device usage and operative techniques are well documented, reducing the likelihood of misclassification.	Low risk
Domain 4 (Deviations from Intended Interventions)	All patients received the planned surgical procedures following a standardized protocol, performed by the same surgical team. There is no evidence of deviations, crossovers, or additional interventions that would differentially affect outcomes between groups.	Low risk
Domain 5 (Missing Data)	The study reports a minimum follow up period of five years with mean follow up durations in the 83–84-month range across groups, indicating robust long-term data, there is no indication of excessive or differential loss to follow up that might bias the results	Low risk
Domain 6 (Measurement of Outcomes)	Outcomes, including clinical scores (NDI, VAS) and radiographic parameters (ROM, sagittal alignment), were measured using validated and standardized methods. the study mentioned that outcomes were collected by a specialized surgeon who performed the data collection and patient follow-up.	Moderate Risk
Domain 7 (Selection of the Reported Result)	The study appears to have reported all pre specified clinical and radiological outcomes along with complication rates.	Low risk
Overall Risk		Moderate Risk

Study: Xiong et al., 2018

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	As a retrospective study, treatment allocation was not randomized. Although baseline characteristics are reported as comparable, residual confounding remains a concern.	Moderate risk
Domain 2 (Selection of Participants)	The study included consecutive patients meeting strict inclusion/exclusion criteria. The selection appears systematic, and there is no indication that participants were chosen based on outcomes.	Low risk
Domain 3 (Classification of Interventions)	interventions are clearly defined. Classification was based on operative records and device usage, making misclassification unlikely.	Low risk
Domain 4 (Deviations from Intended Interventions)	All patients received the intended surgical procedures according to standardized protocols with no reported deviations or crossovers	Low risk
Domain 5 (Missing Data)	The reported follow-up rate is approximately 76%, meaning nearly one-quarter of the initially identified patients were lost to follow-up. In a retrospective study, this level of attrition may introduce bias if the losses are not random or differ between groups.	Moderate Risk
Domain 6 (Measurement of Outcomes)	Clinical outcomes were measured using validated scales (JOA, NDI, VAS) and radiological parameters were as part of outcome follow-up, no blinding reported	Serious Risk
Domain 7 (Selection of the Reported Result)	The study appears to have reported all the pre-specified clinical and radiological outcomes, as well as complication rates, with no evidence of selective reporting.	Low risk
Overall Risk		Serious Risk

Study: Jang et al., 2017

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	Because the study is retrospective, treatment allocation was not randomized. Although baseline demographic characteristics (such as age, sex, and BMI) appear balanced between the ACDF and HS groups, there is always a risk that unmeasured or inadequately controlled confounders	Moderate risk
Domain 2 (Selection of Participants)	The investigators retrospectively reviewed cases that met inclusion and exclusion criteria.	Moderate risk
Domain 3 (Classification of Interventions)	The study clearly defines the interventions: the ACDF group received 3 level fusion with a cervical plate, whereas the HS group underwent a combination of CDR with fusion. Intervention classification was based on operative records and device usage, which reduces the possibility of misclassification.	Low risk
Domain 4 (Deviations from Intended Interventions)	All patients received the intended surgical procedures according to a standardized protocol. Surgeries were performed by experienced surgeons using uniform techniques, and there is no report of crossovers or deviations that could affect the outcomes.	Low risk
Domain 5 (Missing Data)	the study provides follow up data for all included patients over the follow up period with no indication that loss to follow up was excessive or differential between groups.	Low risk
Domain 6 (Measurement of Outcomes)	Clinical outcomes were measured using validated and widely accepted instruments (VAS, NDI, Odom's criteria). Radiological outcomes (ROM, cervical lordosis, and ASD) but no mention of who assessed the outcomes; dependent or independent, blinded or not	Serious Risk
Domain 7 (Selection of the Reported Result)	The study appears to have reported all pre specified clinical and radiological outcomes as well as complication rates. There is no evidence of selective outcome reporting	Low risk
Overall Risk		Serious Risk

Study: Ji et al., 2017

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	The study is non randomized prospective. Patients were allocated to either hybrid surgery (HS) or 2 level ACDF based on clinical criteria. The groups were matched on age and sex but other important prognostic factors (e.g., severity of disc degeneration, baseline functional status, duration of symptoms) may not have been fully controlled so residual confounding remains a concern.	Moderate risk
Domain 2 (Selection of Participants)	The study prospectively enrolled the HS group and retrospectively collected the ACDF group from the same database using predefined inclusion/exclusion criteria and 40 consecutive patients were included in the study.	Low risk
Domain 3 (Classification of Interventions)	Interventions are clearly defined. Assignment was based on objective clinical and imaging criteria.	Low risk
Domain 4 (Deviations from Intended Interventions)	The study reports that all patients underwent the planned surgical procedures following a standardized protocol. There were no crossovers or deviations reported.	Low risk
Domain 5 (Missing Data)	Originally 40 patients were enrolled, with 35 patients completing the 5 year follow up (12.5% loss). The degree of missing data is small and unlikely to bias the results.	Low risk
Domain 6 (Measurement of Outcomes)	Clinical evaluation was performed by a nurse specializing in pain management who was blinded to the patients' treatment." radiographic ROM, fusion status) were assessed using validated instruments and standard radiological measures	Low Risk
Domain 7 (Selection of the Reported Result)	The study appears to have reported all pre specified clinical and radiological outcomes. There is no evidence of selective outcome reporting.	Low risk
Overall Risk		Moderate risk

Study: Grasso, 2015

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	Although the study uses clear inclusion criteria and prospectively enrolls the HS group, treatment allocation is not randomized, leaving residual confounding (e.g., differences in severity of degeneration) a concern.	Moderate risk
Domain 2 (Selection of Participants)	HS patients were enrolled prospectively, while control patients (ACDF and CDA) were identified retrospectively from the same surgical database. Although the inclusion/exclusion criteria were clearly defined, the non-random selection of controls may introduce some selection bias.	Moderate risk
Domain 3 (Classification of Interventions)	Interventions (ACDF, CDA, and HS) are clearly defined and accurately recorded	Low risk
Domain 4 (Deviations from Intended Interventions)	No deviations or crossovers occurred, and all patients received the intended treatment	Low risk
Domain 5 (Missing Data)	The study reports complete clinical and radiological evaluations at each follow-up time point for all enrolled patients with no mention of substantial loss to follow up.	Low risk
Domain 6 (Measurement of Outcomes)	No mention of who assessed patients' outcomes; dependent or independent, blinded or not	Serious Risk
Domain 7 (Selection of the Reported Result)	With all pre-specified outcomes clearly reported, there is no indication of selective reporting	Low risk
Overall Risk		Serious Risk

Study: Hey et al., 2013

Domain	Rationale for judgment	Judgment
Domain 1 (Confounding)	Patients were not randomized. Although the authors attempted to match patients on surgical levels (using a prospectively enrolled HS group and retrospectively selected controls for ACDF and ADR), important confounders (e.g. baseline severity, comorbidities, or subtle differences in patient selection between prospective and retrospective arms) may still exist.	Moderate risk
Domain 2 (Selection of Participants)	The HS group was enrolled prospectively while the controls (ACDF and ADR) were selected retrospectively from a hospital database. An independent party performed the matching; however, the differing methods of enrollment (prospective vs. retrospective) may introduce selection bias.	Moderate risk
Domain 3 (Classification of Interventions)	The interventions (HS, ACDF, and ADR) are clearly defined and described. There is no indication that misclassification of which intervention was performed occurred.	Low risk
Domain 4 (Deviations from Intended Interventions)	All procedures were performed by the same surgeon under a standardized protocol, and no deviations from the intended interventions were reported.	Low risk
Domain 5 (Missing Data)	The study sample is small (21) but appears to have complete follow-up data for all enrolled patients over the minimum 2-year period. There is no explicit report of missing outcome data.	Low risk
Domain 6 (Measurement of Outcomes)	Assessments by 2 independent spine surgeons not directly involved in patient management.	Low Risk
Domain 7 (Selection of the Reported Result)	The study appears to have reported all pre specified clinical and radiological outcomes. There is no evidence of selective outcome reporting	Low risk
Overall Risk		Moderate risk

2- ROB-2 TOOL FOR RANDOMISED CONTROLLED STUDIES (RCT)

Study: Kang et al., 2013

Domain	Signalling question	Response	Description	Judgment
Domain 1 (Randomization process)	1.1. was the allocation sequence random?	No	Twenty-four patients were divided into the study group (hybrid constructs) or control (3-level ACDF) group using randomization based on hospital number: patients from an odd-numbered hospital comprised the study group and patients from even-numbered hospitals comprised the control group. Although the authors attempted to randomize, using hospital numbers is a quasi-random method which can be predictable if the hospital number pattern is known	Some Concerns
	1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?	No Information	The article does not provide explicit details about the method used to conceal the allocation sequence.	
	1.3 Did baseline differences between intervention groups suggest a problem with the randomization process?	No	No significant differences in baseline characteristics were found between the 2 groups. The similarity in baseline characteristics suggests that—even if the method was not ideal—the groups were comparable at the start of the trial.	
Domain 2 (Deviation from intended interventions)	2.1. Were participants aware of their assigned intervention during the trial?	Yes	Due to the nature of surgical procedures, it is inherent that patients know which type of surgery they receive. What matters is that any potential bias from participant awareness is minimized by other measures (such as blinded outcome assessment). The procedures were performed by three surgeons using different surgical techniques. As a result, the surgeons (and likely the clinical staff involved in the care) were aware of which intervention was being performed. The	Low risk
	2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?	Yes		

			key is whether this knowledge led to deviations from the intended protocol. In this trial, the procedures were standardized, and no evidence suggests that the awareness resulted in systematic deviations.	
	2.3. If Y/PY/NI to 2.1 or 2.2: Were there deviations from the intended intervention that arose because of the experimental context?	No	Both intervention groups followed a standardized surgical technique and postoperative follow up. The study does not report any deviations (such as additional unplanned treatments or modifications) that occurred because of the knowledge of group assignment	
	2.4 If Y/PY to 2.3: Were these deviations likely to have affected the outcome?	Not applicable		
	2.5. If Y/PY/NI to 2.4: Were these deviations from intended intervention balanced between groups?	Not applicable		
	2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?	Yes	The authors analyzed outcomes using standard statistical methods (repeated-measures ANOVA and Student's t tests) in an intention-to-treat framework, comparing the groups as originally allocated.	
	2.7 If N/PN/NI to 2.6: Was there potential for a substantial impact (on the result) of the failure to analyze participants in the group to which they were randomized?	Not applicable		
Domain 3 (Missing outcome data)	3.1 Were data for this outcome available for all, or nearly all, participants randomized?	yes	The study included 24 patients (12 in each group) and reports that all patients were followed clinically and radiologically for at least 24 months	Low risk

Domain 4 (Measurement of the outcome)	4.1 Was the method of measuring the outcome inappropriate?	No	The study employed established and validated outcome measures: Clinical outcomes: The Neck Disability Index (NDI) and Visual Analog Scale (VAS) are widely used and validated in cervical spine research. Radiological outcomes: The Cobb method was used for measuring range of motion (ROM), which is a standard and reproducible technique."	Low risk
	4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?	No	Both intervention groups (hybrid constructs and 3-level ACDF) underwent the same schedule of preoperative and postoperative evaluations, and the same measurement techniques were used for all patients.	
	4.3 Were outcome assessors aware of the intervention received by study participants?	No	The study reports that clinical and radiological outcomes were performed by an experienced resident who was blinded to the patients' treatment assignment.	
Domain 5 (Selection of the reported result)	5.1 Were the data that produced this result analyzed in accordance with a pre-specified analysis plan that was finalized before unblinded outcome data were available for analysis?	Probably Yes	The study protocol (as described in the methods) outlined clinical outcomes (NDI, VAS) and radiological outcomes (ROM, adjacent segment motion), all of which were reported in the results. However, the paper does not explicitly mention a publicly available or registered protocol detailing a pre-specified analysis plan. Although a fully detailed, published protocol is not provided, the consistency between the methods described and the outcomes reported suggests that the analyses were conducted as planned. The absence of evidence for post hoc changes or ad hoc selection is reassuring, even though the lack of an external protocol could raise a minor uncertainty	Low risk

	5.2. Multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?	Probably No	The study assessed outcomes at multiple time points (1, 3, 6, 12, and 24 months) for both clinical (NDI, VAS) and radiological parameters (C2–C7 range of motion, adjacent segment motion). The authors report the results for all these time points and provide comparisons between the two groups at each interval. There is no evidence that only the most favorable or statistically significant time points were selectively reported.	
	5.3. Multiple eligible analyses of the data?	Probably No	The article includes complete tables for clinical outcomes (NDI, VAS) and detailed radiological data, along with comparisons between the intervention groups. The methods indicate that all the outcomes planned for evaluation were reported. There is no evidence of missing outcomes or unexplained changes in the outcome reporting.	
Overall Risk				Some concerns