Supplementary Table 1: PRISMA checklist.

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	2

Objectives	4	Provide an explicit statement of questions being addressed with reference to PICOS.	2
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	3
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present a full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	4
Study	9	State the process for selecting studies (i.e., screening,	4

selection		eligibility, included in a systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe the method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4
Risk of bias in individual studies	12	Describe methods used for assessing the risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5
Summary measures	13	State the principal summary measures (e.g., RR, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	5
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	6
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	6
Risk of bias within studies	19	Present data on the risk of bias of each study and, if available, any outcome level assessment (see item 12).	6
Results of	20	For all outcomes considered (benefits or harms), present, for	6–8

individual studies		each study: (a) simple summary data for each intervention group (b) effect estimates and CIs, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including CIs and measures of consistency.	6–8
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	6–8
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	6–8
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policymakers).	9
Limitations	25	Discuss limitations at the study and outcome level (e.g., risk of bias), and at the review level (e.g., incomplete retrieval of identified research, reporting bias).	10
Conclusions	26	Provide a general interpretation of the results in the context of	10

		other evidence and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); the role of funders for the systematic review.	11

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009;6:e1000097. doi: 10.1371/journal.pmed1000097. For more information, visit: www.prisma-statement.org.

CI: Confidence interval; PICOS: Participants, interventions, comparisons, outcomes, and study design; RR: Risk ratio.

Supplementary Table 2: Search strategies.

Search	**Query**	**Results**
PubMed		
#1	((((((bispectral index[Title/Abstract]) OR (bispectral index monitor[Title/Abstract])) OR (anesthesia depth[Title/Abstract])) OR (anesthetic depth[Title/Abstract])) OR (spectral	94,744

	entropy[Title/Abstract])) OR (depth of anesthesia[Title/Abstract])) OR (bis[Title/Abstract])	
#2	(((((((((postoperative outcome) OR (postoperative complication)) OR (complications)) OR (pain)) OR (death)) OR (mortality)) OR (cognitive)) OR (cognition)) OR (delirium)) OR (POCD)	6,528,169
#3	#1 AND #2	6385
#4	#3 AND "Randomized Controlled Trial"[pt]	682
EMBASE		
#1	"bispectral index":ab,ti OR "bispectral index monitor":ab,ti OR "anesthesia depth":ab,ti OR "anesthetic depth":ab,ti OR "spectral entropy":ab,ti OR "depth of anesthesia':ab,ti OR bis:ab,ti'bispectral index":ab,ti OR "bispectral index monitor":ab,ti OR "anesthesia depth":ab,ti OR "anesthetic depth":ab,ti OR "spectral entropy":ab,ti OR "depth of anesthesia":ab,ti OR bis:ab,ti	100,944
#2	"postoperative outcome":ab,ti OR "postoperative complication":ab,ti OR complications:ab,ti OR pain:ab,ti OR death:ab,ti OR mortality:ab,ti OR cognition:ab,ti OR cognitive:ab,ti	4,547,125

	OR delirium:ab,ti OR POCD:ab,ti	
#3	#1 AND #2	5090
#4	#3 AND "randomized controlled trial"/de	502
Cochrane Library		
#1	("bispectral index"):ti,ab,kw OR ("bispectral index monitor"):ti,ab,kw OR ("anesthesia depth"):ti,ab,kw OR ("anesthetic depth"):ti,ab,kw OR ("spectral entropy"):ti,ab,kw OR ("depth of anesthesia"):ti,ab,kw OR ("BIS"):ti,ab,kw	6403
#2	("postoperative outcome"):ti,ab,kw OR ("postoperative complication"):ti,ab,kw OR (complications):ti,ab,kw OR (pain):ti,ab,kw OR (death):ti,ab,kw OR (motality):ti,ab,kw OR (cognitive):ti,ab,kw OR (cognition):ti,ab,kw OR (delirium):ti,ab,kw OR (POCD):ti,ab,kw	516,467
#3	#1 AND #2 in trials	1810

Supplementary Table 3: Definitions of perioperative NCDs.

Reference	Definitions

POD	
Chan *et al*[12]	POD was defined as acute fluctuating course of inattention and either disorganized thinking or an altered level of consciousness. The incidence of delirium in the hospital, as determined by the CAM.
Evered *et al*[16]	Delirium was assessed for 5 days postoperatively or until discharge, using the CAM or if patients were in the ICU, using the CAM-ICU.
Kunst *et al*[37]	Delirium was defined by at least one positive postoperative CAM test. In case of a positive CAM test result, the written results were double-checked for the correct diagnosis of delirium by a second member of the study team. The incidence of delirium during the first 3 days or 5 days after surgery.
Zhou *et al*[48]	The diagnosis of delirium required the following clinical symptoms: (1) an acute onset of cognitive changes with a fluctuating course, (2) inattention, together with either (3) disorganized thinking, or (4) an altered level of consciousness. The incidence of delirium during the first 5 days after surgery, as determined by the CAM.
DNR and postoperative NCDs	
An *et al*[8]	A neuropsychologic battery including seven tests with nine subscales was

	administered preoperatively and 5 days after surgery. A postoperative deficit was defined as a decrement to baseline score >1 SD on any test. Patients who experienced two or more deficits were deemed to have DNR.
Chan *et al*[12]	A battery of three neuropsychological tests was administered before and at 1 week and 3 months after surgery. DNR was defined by comparing with matched control patients who did not have surgery during the same period.
Farag *et al*[35]	The primary cognitive outcome measures consisted of the Processing Speed Index, Working Memory Index, and a Verbal Memory Index. NCDs were defined as decrements in performance that exceed those expected by chance alone in normal samples at the lower fifth percentile (i.e., a negative *Z*-score ≤1.64).
Hou *et al*[49]	A neuropsychological assessment was conducted at 1 day, 3 days, and 7 days after surgery using MoCA by an experienced psychiatrist. The MoCA included 16 items and 11 categories, and examines visuospatial and executive functions naming, memory, attention, language, abstraction, and orientation. DNR was defined as *Z*-score >1.96.
Jildenstål *et al*[36]	The MMT and the Cognitive Failure Questionnaire were used preoperatively

	and postoperatively to evaluate cognitive status. A MMT value <25 was regarded as DNR at postoperative day 1 and a value <16 was regarded as NCDs at 7 days and 1 month postoperatively.
Quan *et al*[44]	A battery of nine neuropsychological tests was administered at baseline (1 day before surgery) and at 7 days and 3 months after surgery. The SD for each test was computed from all the preoperative scores. An individual with postoperative performance deteriorated by ≥1 SDs on two or more tests was classified as having DNR and postoperative NCDs.
Valentin *et al*[45]	DNR and postoperative NCDs were defined by the occurrence of cognitive impairment in Telephone Interview for Cognitive Status and at least one of eight possible deficits of the other neuropsychologist tests.
Xu *et al*[51]	Cognitive function was assessed using the MMSE before operation and at 3 h after operation. DNR was defined as the patients with a score of ≤26.

CAM: Confusion assessment method; CAM-ICU: Confusion assessment method in the intensive care unit; DNR: Delayed neurocognitive recovery; ICU: Intensive care unit; MMSE: Mini-mental State Examination; MMT: Mini-mental test; MoCA: Montreal cognitive assessment; NCDs: Neurocognitive disorders; POD: Postoperative delirium; SD: Standard deviation.

Supplementary Table 4: Primary and secondary outcomes.

Outcomes	Number of studies	Deep anesthesia (no. or no./total)	Light anesthesia (no. or no./total)	Effect size (95% CI)	*P* value	I^2 (%)
Primary outcomes						
VAS pain scores at rest 0–1 h postoperatively	5	249	256	WMD = – 0.72 (– 1.25, –0.18)	0.009	33
POD up to 1 week postoperatively or until discharge	4	202/794	125/785	RR = 1.57 (1.28, 1.91)	<0.0001	0
Secondary outcomes: pain						
VAS scores at rest at 8 h postoperatively	3	100	100	WMD = – 1.16 (– 1.74, –0.57)	0.0001	0
VAS scores at rest at 24 h postoperatively	4	130	130	WMD = – 0.50 (– 0.94, –0.06)	0.03	52

Outcomes	Number of studies	Deep anesthesia (no. or no./total)	Light anesthesia (no. or no./total)	Effect size (95% CI)	*P* value	I^2 (%)
VAS scores on movement at 8 h postoperatively	3	100	100	WMD = – 1.25 (– 1.88, –0.61)	0.0001	0
VAS scores on movement at 24 h postoperatively	3	126	129	WMD = – 0.52 (– 1.14, 0.11)	0.11	55
Intraoperative sufentanil consumption (μg)	9	970	954	WMD = 4.39 (– 1.88, 10.65)	0.17	82
Postoperative rescue analgesia	3	24/112	48/113	RR = 0.46 (0.19, 1.07)	0.07	64
Persistent pain during 3–12 months postoperatively	2	226/3380	253/3369	RR = 0.89 (0.75, 1.06)	0.19	0
Secondary outcomes: cognitive function						
DNR during 1–7 days postoperatively	7	157/865	132/834	RR = 1.29 (0.69, 2.41)	0.42	75
NCDs during 1–3 months postoperatively	6	100/1042	77/1006	RR = 1.17 (0.76, 1.80)	0.47	34
MMSE scores on postoperative day 1	5	210	206	WMD = 0.79 (–0.70, 2.28)	0.30	98

Outcomes	Number of studies	Deep anesthesia (no. or no./total)	Light anesthesia (no. or no./total)	Effect size (95% CI)	*P* value	I^2 (%)
MMSE scores during 3–5 days postoperatively	2	71	70	WMD = − 0.28 (− 2.16, 1.61)	0.77	68
Secondary outcomes: postoperative recovery						
Time to emergence (min)	6	463	465	WMD = 3.65 (1.94, 5.36)	<0.0001	90
Time to extubation (min)	6	217	219	WMD = 3.64 (1.39, 5.90)	0.002	89
Orientation recovery time (min)	3	91	89	WMD = 4.51 (1.61, 7.40)	0.002	88
Length of PACU stay (min)	7	3560	3852	WMD = 5.85 (2.30, 9.41)	0.001	83
Length of ICU stay (days)	2	492	492	WMD = − 0.00 (− 0.02, 0.02)	0.97	0
Length of hospital stay (days)	6	4194	4178	WMD = 1.00 (0.14, 1.86)	0.02	94
QoR-9 scores on postoperative day 1 (0–18)	2	513	514	WMD = − 0.56 (− 3.50, 2.38)	0.71	95

Outcomes	Number of studies	Deep anesthesia (no. or no./total)	Light anesthesia (no. or no./total)	Effect size (95% CI)	*P* value	I^2 (%)
90-day physical recovery scores (0–100)	3	687	681	WMD = − 1.49 (− 3.09, 0.10)	0.07	35
90-day mental recovery scores (0–100)	3	687	681	WMD = 1.44 (0.17, 2.71)	0.03	0
Secondary outcomes: complications and mortality						
Clinically significant hypotension	6	209/878	172/881	RR = 1.19 (0.90, 1.58)	0.23	64
PONV	5	635/3185	679/3470	RR = 0.67 (0.37, 1.20)	0.17	80
Any major complication	4	495/4028	425/4024	RR = 1.22 (0.85, 1.76)	0.28	75
Myocardial infarction	2	79/3515	81/3510	RR = 0.98 (0.72, 1.33)	0.87	0
Sepsis	2	223/3515	206/3510	RR = 1.08 (0.90, 1.30)	0.42	0
Stroke	2	33/3515	43/3510	RR = 0.76 (0.49, 1.20)	0.24	0
Wound infection	6	350/4120	342/4119	RR = 1.15 (0.84, 1.59)	0.38	56
Intraoperative awareness	4	0/3441	2/3428	RR = 0.34 (0.04, 3.20)	0.34	0
1-year cancer recurrence	3	234/3441	241/3433	RR = 0.97 (0.82, 1.15)	0.76	0

Outcomes	Number of studies	Deep anesthesia (no. or no./total)	Light anesthesia (no. or no./total)	Effect size (95% CI)	*P* value	I^2 (%)
Mortality within 30–90 days postoperatively	2	6/239	4/247	RR = 1.55 (0.44, 5.45)	0.50	0
1-year mortality	4	268/3802	240/3798	RR = 1.12 (0.95, 1.32)	0.19	0

CI: Confidence interval; DNR: Delayed neurocognitive recovery; ICU: Intensive care unit; MMSE: Mini-mental State Examination (0–30); PACU: Post-anesthesia care unit; POD: Postoperative delirium; PONV: Postoperative nausea and vomiting; QoR: Quality of recovery; RR: Risk ratio; VAS: Visual analogue scale (0–10); WMD: Weighted mean difference.

Supplementary Table 5: GRADE evidence profile of the main outcomes.

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
VAS pain scores at rest at 0–1 h postoperatively												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
5	Randomized trials	Serious*	Not serious	Not serious	Not serious	None	249	256	-	WMD **0.72 lower** (from 1.25 lower to 0.18 lower), *P*=0.009	⊕⊕⊕◯ Moderate	IMPORTANT
Incidence of POD												
4	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	202/794 (25.4%)	125/785 (15.9%)	**RR 1.57** (1.28–1.91)	**91 more per 1000** (from 45 more to 145 more), *P*<0.0001	⊕⊕⊕⊕ High	CRITICAL
DNR during 1–7 days postoperatively												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
7	Randomized trials	Serious†	Very serious‡	Not serious	Not serious	None	157/865 (18.2%)	132/834 (15.8%)	**RR 1.29** (0.69–2.41)	46 more per 1000 (from 49 fewer to 223 more), *P*=0.42	⊕◯◯◯ Very low	CRITICAL
NCDs during 1–3 months postoperatively												
6	Randomized trials	Serious§	Not serious	Not serious	Not serious	None	100/1042 (9.6%)	77/1006 (7.7%)	**RR 1.17** (0.76–1.80)	13 more per 1000 (from 18 fewer to 61 more), *P*=0.47	⊕⊕⊕◯ Moderate	CRITICAL
Time to extubation (min)												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
6	Randomized trials	Very serious‖	Very serious¶	Not serious	Not serious	None	217	219	-	WMD **3.64 higher** (from 1.39 higher to 5.9 higher), *P*=0.002	⊕◯◯◯ Very low	IMPORTANT
Length of PACU stay (min)												
7	Randomized trials	Very serious**	Very serious††	Not serious	Not serious	None	3560	3852	-	WMD **5.85 higher** (from 2.3 higher to 9.41 higher), *P*=0.001	⊕◯◯◯ Very low	IMPORTANT
Length of hospital stay (days)												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
6	Randomized trials	Not serious	Very serious‡‡	Not serious	Not serious	None	4194	4178	-	WMD **1 higher** (from 0.14 higher to 1.86 higher), *P*=0.02	⊕⊕◯◯ Low	IMPORTANT
Clinically significant hypotension												
6	Randomized trials	Serious§§	Serious|||	Not serious	Not serious	None	209/878 (23.8%)	172/881 (19.5%)	**RR 1.19** (0.90–1.58)	37 more per 1000 (from 20 fewer to 113 more), *P*=0.23	⊕⊕◯◯ Low	IMPORTANT
Incidence of PONV												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
5	Randomized trials	Serious¶¶	Very serious***	Not serious	Not serious	None	635/3185 (19.9%)	679/3470 (19.6%)	**RR 0.67** (0.37–1.20)	65 fewer per 1000 (from 123 fewer to 39 more), *P*=0.17	⊕◯◯◯ Very low	CRITICAL
1-year cancer recurrence												
3	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	234/3441 (6.8%)	241/3433 (7.0%)	**RR 0.97** (0.82–1.15)	2 fewer per 1000 (from 13 fewer to 11 more), *P*=0.76	⊕⊕⊕⊕ High	IMPORTANT
Any major complication												

Certainty assessment							No of patients		Effect			
No of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Deep	Light	Relative (95% CI)	Absolute (95% CI)	Certainty	Importance
4	Randomized trials	Not serious	Very serious†††	Not serious	Not serious	None	495/4028 (12.3%)	425/4024 (10.6%)	**RR 1.22** (0.85–1.76)	23 more per 1000 (from 16 fewer to 80 more), *P*=0.28	⊕⊕◯◯ Low	CRITICAL
1-year mortality												
4	Randomized trials	Not serious	Not serious	Not serious	Not serious	None	268/3802 (7.0%)	240/3798 (6.3%)	**RR 1.12** (0.95–1.32)	8 more per 1000 (from 3 fewer to 20 more), *P*=0.19	⊕⊕⊕⊕ High	CRITICAL

CI: Confidence interval; DNR: Delayed neurocognitive recovery; GRADE: Grading of Recommendations Assessment, Development and Evaluation; NCDs: Neurocognitive disorders; PACU: Post-anesthesia care unit; POD: Postoperative delirium; PONV: Postoperative nausea and vomiting; RR: Risk ratio; VAS: Visual analogue scale (0–10); WMD: Weighted mean difference.

*Three trials were at unclear risk of bias. †Four trials were at unclear risk of bias.‡Heterogeneity: I^2 = 75%. §Two trials were at

unclear risk of bias. ||One trial was at unclear risk of bias. ¶Heterogeneity: I^2 = 89%. **One trial was at unclear risk of bias. ††Heterogeneity: I^2 = 83%. ‡‡Heterogeneity: I^2 = 94%. §§Two trials were at unclear risk of bias. ||||Heterogeneity: I^2 = 74%. ¶¶Four trials were at unclear risk of bias. ***Heterogeneity: I^2 = 80%. †††Heterogeneity: I^2 = 75%.