INPLASY

# Remimazolam versus propofol and postoperative delirium during adult general anaesthesia: a systematic review and meta-analysis

INPLASY202680077

doi: 10.37766/inplasy2026.8.0077

Received: 20 August 2026

Published: 20 August 2026

**Corresponding author:**
Jingjing Zhang

zhangjing_jing2025@126.com

**Author Affiliation:**
Zhangjiakou First Hospital.

Cao, N; Liu, JY; Zhang, XY; Cai, ZC; Zhao, CX; Tang, CY; Ma, ZH; Zhang, Y; Zhang, JJ.

## ADMINISTRATIVE INFORMATION

**Support -** Observation on the efficacy of remimazolam in the cerebral oxygen metabolism and postoperative cognitive function of elderly patients undergoing thoracic surgery, No.2322089D.

**Review Stage at time of this submission -** Completed but not published.

**Conflicts of interest -** None declared.

**INPLASY registration number:** INPLASY202680077

**Amendments -** This protocol was registered with the International Platform of Registered Systematic Review and Meta-Analysis Protocols (INPLASY) on 20 August 2026 and was last updated on 20 August 2026.

## INTRODUCTION

**Review question / Objective** The PICOS question was defined as follows: adults having surgery under general anaesthesia; remimazolam for induction and/or maintenance; propofol as the comparator; POD as the main outcome, with haemodynamic, safety and recovery measures as secondary outcomes; and parallel-group randomised or observational comparative studies.

**Condition being studied** Adults underwent surgery with general anaesthesia.

## METHODS

**Search strategy** Five bibliographic sources were examined from their start dates through 30 October 2025: PubMed/MEDLINE, Embase, Web of Science Core Collection, Cochrane CENTRAL, and CNKI. Searches were not limited by language or publication year. Database-specific subject headings were combined with ordinary text terms describing remimazolam, propofol, delirium, postoperative cognitive outcomes, and related concepts. Reference sections of eligible articles and closely related reviews were examined for additional records not retrieved by the electronic searches.

**Participant or population** Participants are adults aged 18 years or older and underwent surgery with general anaesthesia.

**Intervention** Separate remimazolam and propofol groups, with the drugs used for anaesthetic induction, maintenance, or both.

**Comparator** Compare remimazolam and propofol-based general anaesthesia in adults.

**Study designs to be included** Randomised trial and observational comparative studies.

**Eligibility criteria** A study entered the primary review set when its participants were adults aged 18 years or older and underwent surgery with general anaesthesia. The study also had to contain separate remimazolam and propofol groups, with the drugs used for anaesthetic induction, maintenance, or both. POD had to be identified with a structured delirium assessment. Accepted approaches included CAM-based instruments, CAM-ICU, 3D-CAM, CAM-CR, DSM-based diagnostic criteria, and other clearly defined delirium tools reported by the original investigators. For numerical outcomes, usable data required the number of participants together with a mean and a measure of dispersion that could be converted to a standard deviation. and a randomised trial or observational comparative study with parallel treatment groups.

Studies conducted primarily under neuraxial or regional anaesthesia with intravenous sedation were excluded from the general-anaesthesia primary synthesis. An otherwise eligible study conducted under subarachnoid block with remimazolam or propofol sedation was retained only for a separate add-back sensitivity analysis.

We did not include reports without a concurrent comparison group, studies lacking an independent propofol arm, paediatric investigations, non-surgical sedation studies, duplicate reports without additional usable information, or studies in which the required outcome data could not be obtained. Reports with an undefined or non-reproducible delirium assessment were also excluded from the POD synthesis.

**Information sources** PubMed/MEDLINE, Embase, Web of Science Core Collection, Cochrane CENTRAL, and CNKI.

**Main outcome(s)** The principal endpoint was whether a participant developed POD during the early period after surgery. If a study gave several POD time points, we selected one outcome using this order: cumulative POD during days 1-3, cumulative POD during days 1-7, then the latest single assessment available within the first 7 days. For studies conducted in intensive care settings, an assessment period of ≤72 hours was considered clinically comparable to the postoperative-day 1–3 window.

When separate time-point counts were reported without clarification as to whether the same patients were represented at more than one assessment, the principal extraction used the most clinically appropriate cumulative estimate available from the report. Alternative plausible assumptions regarding unique-patient counts were examined in sensitivity analyses.

**Additional outcome(s)** Secondary binary outcomes were study-defined hypotension, bradycardia and PONV. Secondary continuous outcomes were extubation or airway-removal time and PACU stay duration. Vasopressor requirements were extracted when available but were summarised narratively when studies reported incompatible drug types, exposure definitions or dose metrics.

Study-specific definitions and assessment windows for secondary outcomes were recorded before synthesis. Outcomes were quantitatively pooled only when at least three studies provided sufficiently comparable definitions and extractable group-level data. Where definitions or follow-up periods differed materially, results were evaluated using assessment-window sensitivity analyses or narrative synthesis rather than inferential comparisons of unweighted study averages.

**Quality assessment / Risk of bias analysis** For each randomised trial, we evaluated bias specifically for the POD result using RoB 2. The assessment covered sequence generation and allocation, deviations after treatment assignment, incomplete outcome information, the way POD was measured, and selective choice of the reported analysis.

Judgements were recorded in the three RoB 2 categories: low risk, some concerns, or high risk, both for individual domains and for the study-level POD result. Overall risk was low only when every domain was low. The label some concerns was used when at least one domain raised concern but none was high. A study was rated high when at least one domain was high or when several domain-level concerns together materially reduced confidence in the result.

Non-randomised comparative studies were examined with the Newcastle–Ottawa Scale. Points were assigned for participant selection, control of important confounding factors, and outcome ascertainment, giving a possible total of nine. We regarded totals of 7–9 as high, 4–6 as moderate, and 0–3 as low methodological quality. Two reviewers completed these evaluations separately, and a third person settled any remaining difference.

**Strategy of data synthesis** For the main pooled estimate, $\tau^2$ was estimated with the DerSimonian–Laird method and the confidence limits were obtained with the Hartung–Knapp procedure. Because the studies differed clinically and methodologically, conclusions were based mainly on the random-effects result.

Variation across study estimates was described with $I^2$ and $\tau^2$. We also calculated Cochran's Q and

treated p<0.10 as evidence warranting attention to between-study heterogeneity. When enough studies were available, we also calculated a 95% prediction interval to show the range expected for a comparable future study.

Binary secondary outcomes were synthesised as RRs, and continuous secondary outcomes were synthesised as MDs. The same random-effects framework was used for these analyses. Common-effect estimates were also presented where appropriate to demonstrate model sensitivity. No inferential p-values were calculated from simple, unweighted averages of study-specific event rates or means.

Post hoc subgroup analyses examined study design, study-level age category and surgical setting. Age groups were defined using the study-reported mean or median age as <65 years, 65–74 years or ≥75 years. Surgical categories included orthopaedic or spinal surgery, gastrointestinal surgery, thoracic surgery, cardiac procedures and neurovascular or other procedures. These analyses represented study-level comparisons and were not interpreted as individual-patient treatment interactions.

Between-subgroup differences were evaluated using interaction tests under common-effect and random-effects frameworks. Subgroups represented by a single study were displayed descriptively but were not interpreted as confirmed procedure-specific effects.

To determine whether the pooled result depended on a particular study, the random-effects analysis was recalculated repeatedly, leaving out one study on each run.

Possible small-study effects were examined visually with a funnel plot and, when enough effect estimates were available, numerically with Egger regression. Because such tests have low power with few studies, a non-significant Egger result was interpreted cautiously.

Tests were evaluated on a two-sided basis. For pooled effects and moderator analyses, statistical significance was set at p<0.05; the threshold for Cochran's Q was p<0.10.

**Subgroup analysis** Post hoc subgroup analyses examined study design, study-level age category and surgical setting. Age groups were defined using the study-reported mean or median age as <65 years, 65–74 years or ≥75 years. Surgical categories included orthopaedic or spinal surgery, gastrointestinal surgery, thoracic surgery, cardiac procedures and neurovascular or other procedures. These analyses represented study-level comparisons and were not interpreted as individual-patient treatment interactions.

Between-subgroup differences were evaluated using interaction tests under common-effect and random-effects frameworks. Subgroups represented by a single study were displayed descriptively but were not interpreted as confirmed procedure-specific effects.

**Sensitivity analysis** The following sensitivity analyses were conducted: (1) repeating the primary synthesis after adding the study conducted under subarachnoid block with intravenous sedation; (2) sequentially excluding each study in a leave-one-out influence analysis; (3) applying alternative assumptions to the postoperative-delirium event counts reported at separate assessment times in the Chen 2025 study; (4) comparing the random-effects and common-effect estimates; and (5) restricting PONV analyses according to the reported postoperative assessment window.

**Country(ies) involved** China, Korea, Japan.

**Keywords** Remimazolam; propofol; delirium; postoperative delirium; general anaesthesia; hypotension.

**Contributions of each author**

Author 1 - Na Cao.
Author 2 - Jingya Liu.
Author 3 - Xiaoyan Zhang.
Author 4 - Zechao Cai.
Author 5 - Caixia Zhao.
Author 6 - Chunying Tang.
Author 7 - Zhenhua Ma.
Author 8 - Yan Zhang.
Author 9 - Jingjing Zhang.