Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Long-term health-related quality of life after iatrogenic bile duct injury compared with controls: an updated systematic review and meta-analysis

Arun Anand

arunanand093@gmail.com

All India Institute of Medical Sciences

Sushant Soren

All India Institute of Medical Sciences

Systematic Review

Keywords: bile duct injury, iatrogenic, cholecystectomy, quality of life, SF-36, hepaticojejunostomy, meta-analysis

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11829324/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose. Bile duct injury (BDI) is a serious complication of cholecystectomy, and the only meta-analysis of its long-term health-related quality of life (HRQoL) dates from 2013. We updated it, comparing patients with BDI against patients who had an uncomplicated cholecystectomy.

Methods. We searched MEDLINE, Embase, Scopus, LILACS, CENTRAL and SciELO on 15 November 2026 and included studies that measured HRQoL at least 12 months after injury with a validated instrument. The registered primary outcome was the mean difference in the SF-36 physical (PCS) and mental (MCS) component scores against concurrent controls (random-effects REML, Hartung–Knapp). A secondary analysis pooled standardised mean differences across all studies with usable controls. Certainty was graded with GRADE.

Results. Fourteen studies were included. Three (604 patients with BDI, 359 controls) allowed the primary analysis: the mean difference was − 6.7 points (95% CI − 15.4 to 2.1; I^2 79%) for PCS and − 2.6 points (− 5.7 to 0.4; I^2 18%) for MCS, both intervals including no difference. In the secondary analysis of seven studies (787 patients, 543 controls), standardised mean differences were − 0.43 (− 0.76 to − 0.09) for PCS and − 0.26 (− 0.44 to − 0.08) for MCS, and were smaller and not significant in the four studies rated good. Certainty was very low for both outcomes.

Conclusion. Long-term HRQoL after BDI is probably lower than after uncomplicated cholecystectomy, mainly in physical health, but the registered analysis is imprecise and its confidence intervals include no difference. Studies with matched concurrent controls and full reporting are needed.

PROSPERO CRD420261571389, registered 15 November 2026.

Introduction

Bile duct injury (BDI) remains one of the most serious complications of cholecystectomy, with a reported incidence of roughly 0.2% to 0.7% after laparoscopic cholecystectomy.[1] Modern reconstruction in specialist centres gives low mortality and good anastomotic patency, so attention has moved from survival to how patients feel years later.[2–4] Health-related quality of life (HRQoL) is now widely used for this purpose, most often through the 36-item Short Form Health Survey (SF-36) and the Gastrointestinal Quality of Life Index (GIQLI).

Published results are inconsistent. Early prospective and case-control series reported lasting impairment,[5, 6, 7] while others found no difference from patients who had an uncomplicated operation.[1, 8, 9] The only quantitative synthesis, by Landman and colleagues in 2013, searched to 2012.[11] A later systematic review looked only at non-operative management and concluded that results were conflicting, with major injury and non-surgical treatment as possible contributors to poorer HRQoL.[12] Since then, several cohort and case-control studies, including national cohorts, have appeared from Europe, Asia and North America.[13–16] A search of PubMed on 12 November 2026 found no later pooled analysis dedicated to this question.

We therefore updated the 2013 review. Our question was whether long-term HRQoL after iatrogenic BDI differs from that of patients who had an uncomplicated cholecystectomy, and how large any difference is.

Materials and methods

Protocol and registration

The protocol was registered in PROSPERO (CRD420261571389) on 15 November 2026, before screening, data extraction and analysis began. The review is reported according to PRISMA 2020,[17] PRISMA-S and MOOSE. Deviations from the registered protocol are listed at the end of this section.

Eligibility criteria

We included cohort, case-control and cross-sectional studies of patients of any age with iatrogenic BDI, managed by any strategy, in which HRQoL was measured with a validated instrument at least 12 months after injury or repair. Studies were eligible for the primary pooled analysis if they included a concurrent control group of patients who had an uncomplicated cholecystectomy. Studies without such controls, or that used other instruments, were eligible for narrative synthesis. We excluded non-iatrogenic biliary injury, assessment before 12 months, ad hoc questionnaires only, case reports, studies with fewer than 10 BDI patients, and conference abstracts without extractable data. Only English-language reports were included. Where several reports described one cohort, we used the most complete report and counted the cohort once.

Information sources and search

On 15 November 2026 we searched MEDLINE (PubMed), Embase, Scopus, LILACS (BVS portal), the Cochrane Central Register of Controlled Trials (CENTRAL) and SciELO, without date limits, together with ClinicalTrials.gov through Embase. The strategy combined terms for bile duct injury and stricture with terms for quality of life and named instruments (the PubMed string is given in Supplementary Appendix 1). Web of Science could not be searched with the available access. We also screened the reference lists of included studies and of Landman 2013 and Halle-Smith 2020, and searched forward citations.

Study selection

Records were managed in Rayyan. One reviewer (AA) screened titles and abstracts in blinded mode, and a second reviewer (SS) verified each decision. Full texts were assessed against the criteria in the same way, with disagreements settled by discussion. This differs from the registered plan of independent dual screening; no agreement statistic was calculated.

Data extraction

We extracted study design, setting, period, number of patients and controls, injury classification and management, follow-up, control definition and matching, instrument and scoring, mode of administration, and SF-36 results. The primary outcome was the PCS and MCS score at 12 months or more. Extraction was done by one reviewer (AA) and checked against the source papers by a second reviewer (SS). We did not contact study authors; analyses use only what the published reports contain.

Effect measures and handling of missing data

The registered effect measure, and our primary measure, was the mean difference in SF-36 points (BDI minus control). Three studies (Moore 2004, Booij 2018 and Sturesson 2020) reported or allowed calculation of means and standard deviations for both groups, which meets the protocol minimum of three studies for pooling. Reported means and standard deviations were used directly (Moore 2004, Sturesson 2020), and the medians and interquartile ranges of Booij 2018 were converted with the method of Wan and colleagues.[18] To make use of the remaining studies we also calculated, as a secondary analysis, the standardised mean difference (SMD; Hedges g with the small-sample correction) for all seven studies with usable controls. Negative values mean worse HRQoL in the BDI group.

For the secondary analysis, the following steps were applied in order. (1) Reported or converted means and standard deviations were used as above. (2) For Koppatz 2021, medians were used as means and the standard deviation was imputed from the control interquartile range divided by 1.35. (3) For Sarmiento 2004 and Hogan 2009, which reported medians and exact P values only, g was derived from the P value, group sizes and direction of effect. (4) For Kharel 2026, which reported the eight SF-36 domain scores but not component scores, the PCS proxy was the mean of the four physical domain effect sizes and the MCS proxy the mean of the four mental domain effect sizes. Studies that could not be placed in any of these steps were described narratively. Studies with derived or imputed effects were labelled and examined in sensitivity analyses.

Risk of bias

Risk of bias was assessed with the Newcastle-Ottawa Scale for case-control studies,[19] adapted so that BDI was the exposure and SF-36 at 12 months or more was the outcome (maximum nine stars; 7 to 9 good, 4 to 6 fair, 0 to 3 poor). Hospital-based or volunteer controls received no star for control selection in any study. Differences in follow-up between groups and low response rates are not captured by the stars and were carried into the certainty assessment. The ratings were made by the first reviewer and verified by the second.

Synthesis

Mean differences were pooled with random-effects meta-analysis using restricted maximum likelihood and the Hartung–Knapp confidence interval (t distribution, k − 1 degrees of freedom, standard error not allowed below the fixed-effect value).[20] Heterogeneity was summarised with τ^2, I^2 and Cochran Q, and a 95% prediction interval was calculated for three or more studies.[21] A fixed-effect model was reported alongside. The secondary SMD analysis used the same methods, and SMDs were converted to approximate SF-36 points by multiplying by 10, because norm-based summary scores have a standard deviation of about 10. Sensitivity analyses of the primary analysis were exclusion of Booij 2018 (data converted from medians) and a fixed-effect model. Sensitivity analyses of the secondary analysis were leave-one-out, exclusion of the domain-proxy study, exclusion of the largest study, restriction to studies reporting or converting means and standard deviations, restriction to studies with matched controls, restriction to studies rated good (7 to 9 stars), and a fixed-effect model. Prespecified subgroup analyses by time since injury, injury severity, instrument and control type, and meta-regression, were not carried out because of the small number of studies. Publication bias was shown with funnel plots for description only, and no Egger test was done because fewer than ten

studies were available. Effect sizes were calculated with spreadsheet formulas. Pooling was done in Python 3 (NumPy and SciPy) with code written for REML estimation and the Hartung–Knapp adjustment, and the REML pooled estimates and between-study variances were reproduced with the PyMARE package.

Certainty of evidence

Certainty was rated for each outcome with GRADE, with observational evidence starting at low and downgrading for risk of bias, inconsistency, indirectness, imprecision and publication bias.[22]

Use of artificial intelligence

A large language model (Claude, Anthropic) was used to help write the statistical code and to draft the manuscript text. The authors checked the pooled estimates against an independent implementation (PyMARE), verified the extracted data and results against the source reports, edited the text, and take responsibility for the content.

Deviations from the protocol

(1) The registered mean difference was kept as the primary effect measure and was estimated from the three studies that allowed it; a standardised mean difference analysis of all seven studies was added as a secondary analysis, which was not described in the registered protocol. (2) In the secondary analysis, effect sizes for four studies were derived from exact P values, imputed standard deviations or domain scores. (3) Screening, data extraction and risk-of-bias rating were done by one reviewer and verified by a second, not independently by two. (4) Study authors were not contacted. (5) Web of Science was not searched. (6) Subgroup analyses and meta-regression were not feasible. (7) The Sturesson 2020 cohort includes the 86 patients and the controls of the national case-control study by Rystedt and Montgomery,[23] which was therefore not pooled separately. (8) The PROSPERO record was updated on 17 November 2026, after data extraction and analysis, to describe the review as conducted.

Results

Study selection

The searches identified 1,112 records, including 25 registry records retrieved through Embase. After removal of 476 duplicates, 636 records were screened and 595 were excluded on title and abstract. Of 41 reports sought for full-text assessment, two could not be retrieved, and two were added by citation searching or manual search, leaving 41 assessed. Twenty-five were excluded, 13 because they were conference abstracts or superseded by a full paper and 12 for other reasons, mainly no control group, no validated instrument or a different population. The 25 registry records were reviewed and none reported HRQoL results. Fourteen studies were included (Fig. 1): seven with usable data for pooling and seven for narrative synthesis, with two companion reports.

Characteristics of the studies

The seven pooled studies came from the USA (two), Ireland, the Netherlands, Finland, Sweden and India and included 787 patients with BDI and 543 controls (Table 1). Three of them (Moore, Booij and Sturesson) entered the primary analysis. Designs were cohort, case-control and cross-sectional. Management was surgical reconstruction in most patients, with endoscopic and percutaneous treatment in the Dutch and Swedish cohorts. Median follow-up ranged from about 4 to 13 years and differed between BDI and control groups in four studies. Control selection and matching varied from unmatched hospital controls to frequency matching on seven variables.

Table 1
Characteristics of the seven studies included in the meta-analysis.

Study	Country and setting	Design	BDI / controls (n)	BDI management	Follow-up after injury	Controls	Instrument
Moore 2004	USA, single centre	Cross-sectional survey	50 / 74	Not separated	BDI mean 62 months; controls 47	Uncomplicated cholecystectomy, same hospital, unmatched	SF-36
Sarmiento 2004	USA, tertiary centre	Cohort with matched controls	45 / 44	Surgical reconstruction (Hepp–Couinaud)	Mean 101 months (range 62–137)	Uncomplicated laparoscopic cholecystectomy, matched on year, sex and age group	SF-36
Hogan 2009	Ireland, tertiary centre	Case-control	62 / 62	Mixed referrals	BDI median 152 months (range 2 months–20 years); controls 27 months	Age- and sex-matched, uncomplicated cholecystectomy	SF-36 (telephone)
Booij 2018	Netherlands, tertiary centre	Cohort with controls	399 / 93	Endoscopic, radiological and surgical	BDI median 4.3 years; controls 10.0 years	Uncomplicated laparoscopic cholecystectomy, unmatched	SF-36, GIQLI
Koppatz 2021	Finland, tertiary centre	Retrospective cohort with matched controls	35 / 53	Major (Strasberg E) injury, mainly hepaticojejunostomy	Median 90 months	Matched on age, urgency and follow-up	SF-36, GIQLI
Sturesson 2020	Sweden, national	National cohort with matched controls	155 / 192	Hepaticojejunostomy (n = 69); drainage, suture or endoscopy (n = 86)	Hepaticojejunostomy median 99 months; other repair 64; controls 43	Frequency-matched 1:2 on age, sex, ASA, date, indication, urgency and technique	SF-36v2, norm-based
Kharel 2026	India, tertiary centre	Cross-sectional survey with matched controls	41 / 25	Roux-en-Y hepaticojejunostomy (delayed repair)	Median 60 months	Uncomplicated laparoscopic cholecystectomy at least 3 years earlier, matched on age and sex	SF-36 (domains only)

BDI, bile duct injury; GIQLI, Gastrointestinal Quality of Life Index. Follow-up is time since injury or repair unless stated.

Risk of bias

Scores on the Newcastle-Ottawa Scale ranged from 5 to 8 (Supplementary Table S3). Four studies were rated good (Sarmiento, Sturesson, Hogan and Koppatz) and three fair (Kharel, Moore and Booij). No study earned a star for control selection. Unequal follow-up between groups was a concern in Hogan, Booij, Moore and Sturesson, and control response was 51% in Koppatz and about 19% in Moore.

Primary analysis: physical component summary

Three studies contributed (604 patients with BDI and 359 controls; Table 2, Fig. 2). The BDI group scored lower in all three, with mean differences of − 11.0 (Moore), − 6.2 (Booij) and − 3.8 points (Sturesson). The pooled mean difference was − 6.7 points (95% CI − 15.4 to 2.1), so the interval included no difference. Heterogeneity was substantial (I^2 79%, τ^2 9.5; Cochran Q 9.6, 2 df, P = 0.008). The 95% prediction interval (− 53.7 to 40.3) is uninformative with three studies. A fixed-effect model gave − 5.6 points (95% CI − 7.1 to − 4.2). Excluding Booij, whose data were converted from medians, gave − 7.0 points (95% CI − 52.1 to 38.2) from two studies.

Table 2
Primary analysis: SF-36 component scores and mean differences in the three studies with usable means and standard deviations.

Study	BDI / controls (n)	PCS, BDI mean (SD)	PCS, control mean (SD)	PCS mean difference (95% CI)	MCS, BDI mean (SD)	MCS, control mean (SD)	MCS mean difference (95% CI)
Moore 2004	50 / 74	36.0 (11.0)	47.0 (12.0)	−11.0 (− 15.1 to − 6.9)	43.0 (14.0)	49.0 (11.0)	−6.0 (− 10.6 to − 1.4)
Booij 2018 †	399 / 93	45.8 (13.7)	51.9 (10.0)	−6.2 (− 8.6 to − 3.7)	50.4 (10.0)	52.5 (6.1)	−2.1 (− 3.7 to − 0.6)
Sturesson 2020	155 / 192	47.3 (10.3)	51.2 (9.3)	−3.8 (− 5.9 to − 1.7)	49.3 (11.0)	52.1 (10.7)	−2.8 (− 5.1 to − 0.5)
Pooled (REML, Hartung–Knapp)	604 / 359			−6.7 (− 15.4 to 2.1)			−2.6 (− 5.7 to 0.4)

PCS, physical component summary; MCS, mental component summary; MD, mean difference (BDI minus control) in SF-36 points. † Means and standard deviations converted from medians and interquartile ranges. Study-level intervals use the normal approximation; the pooled interval is the Hartung–Knapp interval.

Primary analysis: mental component summary

Mean differences were − 6.0 (Moore), − 2.1 (Booij) and − 2.8 points (Sturesson). The pooled mean difference was − 2.6 points (95% CI − 5.7 to 0.4), again including no difference (Fig. 3). Heterogeneity was low (I^2 18%, τ^2 close to zero; Q 2.5, 2 df, P = 0.29) and the prediction interval was − 11.6 to 6.4. A fixed-effect model gave − 2.6 points (95% CI − 3.9 to − 1.4). With only two degrees of freedom the Hartung–Knapp interval is wide, which explains why the random-effects interval includes zero while the fixed-effect interval does not.

Secondary analysis: standardised mean differences in all seven studies

For PCS, six of the seven estimates were negative and one (Sarmiento) was a small positive value (Supplementary Table S4, Supplementary Fig. S3). The pooled SMD was − 0.43 (95% CI − 0.76 to − 0.09), about 4 points lower in the BDI group, with I^2 69% (τ^2 0.087; Q 19.4, P = 0.003) and a prediction interval of − 1.26 to 0.41. A fixed-effect model gave − 0.42 (− 0.54 to − 0.31). For MCS the pooled SMD was − 0.26 (95% CI − 0.44 to − 0.08), about 2.5 points lower, with I^2 35% (τ^2 0.000; Q 9.2, P = 0.16) and a prediction interval of − 0.45 to − 0.07 (Supplementary Fig. S4). A fixed-effect model gave − 0.26 (− 0.38 to − 0.14). The narrow MCS prediction interval partly reflects the small estimated between-study variance with only seven studies and should be read with caution.

Sensitivity analyses of the secondary analysis

Results are in Supplementary Table S5. For PCS, leave-one-out estimates ranged from − 0.51 to − 0.35, and the confidence interval just included zero when Booij was removed (− 0.42, − 0.85 to 0.01). Excluding Kharel gave − 0.37 (− 0.73 to − 0.01). Restricting the analysis to studies rated good, to studies with matched controls, or to the three studies reporting or converting means and standard deviations gave smaller or less certain estimates whose intervals included zero. For MCS, leave-one-out estimates ranged from − 0.28 to − 0.23; removing Sturesson or Booij moved the upper confidence limit to 0.02 and 0.00, while removing any other single study left the interval below zero. Estimates restricted to higher-quality or matched-control studies were again smaller and not significant.

Studies not pooled

Seven included studies could not enter the meta-analysis. Boerma and colleagues (82 patients) reported SF-36 medians in injured patients and matched healthy controls but without usable control numbers, and the same Amsterdam cohort is represented by Booij 2018.[5, 10, 15] Flores-Rangel and colleagues (46 patients, 51 controls) used the SF-12 and reported item-level scores only; patients with BDI scored lower on general health, physical functioning, role limitation due to physical problems and social functioning, with no difference in mental health.[14] Karvonen and colleagues used the 15D instrument and found no major differences from controls.[24] Jabłońska and colleagues compared hepaticojejunostomy (n = 22) with end-to-end repair (n = 21) using SF-36v2 without uninjured controls.[25] Dominguez-Rosado and colleagues,[26] Doğan and colleagues[27] and Otto and colleagues[28] compared patients with population norms or between groups of BDI patients, using SF-36 or the WHOQOL-BREF, without concurrent controls. These studies were consistent with the pooled finding of mainly physical impairment but gave no extra quantitative estimate.

Certainty of evidence

Certainty was very low for both outcomes (Table 3). The main reasons were the observational design, weak control selection and differing follow-up between groups; for PCS the high heterogeneity (I^2 79%); and, for both outcomes, very imprecise estimates from only three studies

whose confidence intervals include no difference. The secondary seven-study analysis added effect sizes derived from P values or imputed standard deviations and a domain-based proxy. Funnel plots (Supplementary Figs. S1 and S2) are not interpretable with seven studies.

Table 3
Certainty of the evidence (GRADE).

Outcome	Studies (participants)	Pooled effect	Certainty	Reasons for rating
SF-36 physical component	3 (604 BDI, 359 controls)	Mean difference −6.7 points (− 15.4 to 2.1)	Very low	Observational design; risk of bias (control selection, follow-up mismatch, converted data) − 1; inconsistency (I^2 79%) − 1; very serious imprecision (three studies, interval includes no difference and wide range of effects) − 2
SF-36 mental component	3 (604 BDI, 359 controls)	Mean difference −2.6 points (− 5.7 to 0.4)	Very low	Observational design; risk of bias − 1; serious imprecision (three studies, interval includes no difference) − 1

Observational evidence started at low certainty. Ratings apply to the primary mean-difference analysis; the secondary seven-study SMD analysis (physical − 0.43, mental − 0.26) was also rated very low because of risk of bias and derived effect sizes. Approximate points = SMD × 10.

Discussion

In this updated review, patients with bile duct injury scored lower than controls on long-term SF-36 in all three studies that allowed the registered mean-difference analysis, but the pooled differences were imprecise: −6.7 points (95% CI − 15.4 to 2.1) for the physical component and − 2.6 points (95% CI − 5.7 to 0.4) for the mental component. The intervals include no difference, so a true difference of zero cannot be excluded, but they also include differences large enough to matter. Pooling standardised differences across all seven studies gave smaller, statistically significant differences (0.43 and 0.26 standard deviations, roughly 4 and 2.5 points), but these rest on four studies with derived or imputed effect sizes and are less reliable. Minimal clinically important differences for the SF-36 summary scores have been reported over a wide range, from 3 to 21 points for the physical score in orthopaedic populations and about 5 for the mental score in one pain population,[29] so the primary physical estimate lies inside this range and the mental estimate is below the single figure reported. Certainty is very low, and the secondary estimates are not stable: they shrink and lose statistical significance in the better-controlled subsets and when single studies are removed.

Our findings agree in direction with the 2013 review, which reported a tendency to emotional detriment and an inconsistent physical effect,[11] although the larger and less consistent difference in our analysis is in the physical component. The 2020 review of non-operative management reached a conflicting conclusion and pointed to injury severity and non-surgical treatment as influences.[12] Within our included studies, the largest and best-matched national cohort found reduced HRQoL in patients who needed hepaticojejunostomy but not in those with less severe injury treated by drainage, suture or endoscopy, and better HRQoL when the injury was recognised and repaired at the index operation.[13, 23] Studies of major injury at a single tertiary centre found no difference,[1] while a small Indian study of delayed repair found large reductions across all eight domains.[16] These contrasts suggest that severity, timing of recognition and repair, and the setting of care may explain much of the heterogeneity, but our data cannot test this because subgroup analysis was not feasible.

Several features of the evidence weaken any firm conclusion. Controls were never selected from the community, response rates in controls were low in some studies, and follow-up was frequently much longer in injured patients than in controls, which could bias comparisons in either direction. Most studies measured HRQoL once, years after the event, so changes over time cannot be assessed. Part of the difference may also reflect the underlying gallstone disease and the complications of the original operation rather than the injury itself, which the available designs cannot separate.

Strengths and limitations

Strengths are a registered protocol, a broad search with citation tracking, full-text review of all potentially eligible reports, and explicit reporting of every conversion step. The limitations are substantial. The primary analysis rests on three studies, one of which required conversion from medians, so the Hartung–Knapp interval is wide and the prediction interval uninformative. The secondary analysis added four studies with effect sizes derived from P values, imputed standard deviations or domain scores, and the Kharel estimate is a proxy for the component scores. Treating medians as means, as for Koppatz, may bias the estimate. Authors were not contacted, so unpublished data could not be recovered. Screening, data extraction and risk-of-bias rating were done by one reviewer and verified by a second rather than done independently, and the English-language restriction and the absence of a Web of Science search may have missed studies. The Sturesson cohort overlaps with an earlier national study, which we counted once. With three studies in the primary analysis and seven in the secondary analysis, the between-study variance and the prediction interval are imprecise, and no test of publication bias was possible.

Implications for practice and research

The available evidence suggests that long-term HRQoL after bile duct injury may remain lower than after uncomplicated cholecystectomy, particularly physical function, although the size of the difference is uncertain; follow-up of injured patients should include HRQoL. Future studies should use matched concurrent controls from the same operative population, report means and standard deviations for the SF-36 component scores, record injury severity and timing of repair, and share patient-level data so that subgroup analysis becomes possible.

Conclusions

Long-term HRQoL after iatrogenic bile duct injury is probably lower than after uncomplicated cholecystectomy, mainly in physical health, but the pooled differences from the registered mean-difference analysis are imprecise and their confidence intervals include no difference. The certainty of this evidence is very low, and better-controlled studies with fuller reporting are needed before the size of the difference can be stated with confidence.

Declarations

Ethics approval

Not applicable. This systematic review used only published aggregate data and involved no human participants, so no ethics committee approval was required.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declare no competing interests.

Statement on sex

No individual-level data were analysed, and sex was not used as a variable.

Authors contributions

AA conceived the review, ran the searches, screened studies, extracted data, did the analysis and wrote the draft. SS verified the screening decisions, the extracted data, the risk-of-bias ratings and the final manuscript.

Supplementary material

Supplementary Appendix 1. PubMed search string (15 November 2026)

("Bile Duct Injuries"[Mesh] OR "bile duct injury" OR "bile duct injuries" OR "biliary injury" OR "iatrogenic bile duct injury" OR "biliary tract injury" OR "bile duct stricture" OR "biliary stricture") AND ("Quality of Life"[Mesh] OR "quality of life" OR "health-related quality of life" OR HRQoL OR HRQOL OR "SF-36" OR "short form 36" OR "SF-12" OR "short form 12" OR GIQLI OR "gastrointestinal quality of life" OR EQ-5D OR EuroQol OR 15D OR "health status" OR "patient-reported outcome*" OR "patient reported outcome*")

Records retrieved: PubMed 195; Embase 489; Scopus 387; LILACS/BVS 31; CENTRAL 3; SciELO 7; total 1,112. Strings for the other databases are available from the corresponding author.

Funding

No funding was received for this work.

Author Contribution

A.A. conceived the review, ran the searches, screened studies, extracted data, did the analysis and wrote the first draft. S.S. verified the screening decisions, the extracted data and the risk-of-bias ratings, and revised the manuscript. Both authors reviewed and approved the final manuscript.

Data and code availability

The extraction and analysis workbook and the analysis code are available from the corresponding author on reasonable request.

References

1. Koppatz H, Sallinen V, Mäkisalo H, Nordin A. Outcomes and quality of life after major bile duct injury in long-term follow-up. Surg Endosc. 2021;35:2879–88.
2. Sicklick JK, Camp MS, Lillemoe KD, et al. Surgical management of bile duct injuries sustained during laparoscopic cholecystectomy: perioperative results in 200 patients. Ann Surg. 2005;241:786–92.
3. de Reuver PR, Grossmann I, Busch OR, et al. Referral pattern and timing of repair are risk factors for complications after reconstructive surgery for bile duct injury. Ann Surg. 2007;245:763–70.
4. Perera MT, Silva MA, Hegab B, et al. Specialist early and immediate repair of post-laparoscopic cholecystectomy bile duct injuries is associated with an improved long-term outcome. Ann Surg. 2011;253:553–60.
5. Boerma D, Rauws EA, Keulemans YC, et al. Impaired quality of life 5 years after bile duct injury during laparoscopic cholecystectomy: a prospective analysis. Ann Surg. 2001;234:750–7.
6. Melton GB, Lillemoe KD, Cameron JL, et al. Major bile duct injuries associated with laparoscopic cholecystectomy: effect of surgical repair on quality of life. Ann Surg. 2002;235:888–95.
7. Moore DE, Feurer ID, Holzman MD, et al. Long-term detrimental effect of bile duct injury on health-related quality of life. Arch Surg. 2004;139:476–81.
8. Sarmiento JM, Farnell MB, Nagorney DM, et al. Quality-of-life assessment of surgical reconstruction after laparoscopic cholecystectomy-induced bile duct injuries: what happens at 5 years and beyond? Arch Surg. 2004;139:483–8.
9. Hogan AM, Hoti E, Winter DC, et al. Quality of life after iatrogenic bile duct injury: a case control study. Ann Surg. 2009;249:292–5.
10. de Reuver PR, Sprangers MA, Rauws EA, et al. Impact of bile duct injury after laparoscopic cholecystectomy on quality of life: a longitudinal study after multidisciplinary treatment. Endoscopy. 2008;40:637–43.
11. Landman MP, Feurer ID, Moore DE, et al. The long-term effect of bile duct injuries on health-related quality of life: a meta-analysis. HPB (Oxford). 2013;15:252–9.
12. Halle-Smith JM, Hodson J, Stevens L, et al. Does non-operative management of iatrogenic bile duct injury result in impaired quality of life? A systematic review. Surgeon. 2020;18:113–21.
13. Sturesson C, Hemmingsson O, Månsson C, et al. Quality-of-life after bile duct injury repaired by hepaticojejunostomy: a national cohort study. Scand J Gastroenterol. 2020;55:1087–92.
14. Flores-Rangel GA, Chapa-Azuela O, Rosales AJ, Roca-Vasquez C, Böhm-González ST. Quality of life in patients with background of iatrogenic bile duct injury. World J Surg. 2018;42:2987–91.
15. Booij KAC, de Reuver PR, van Dieren S, et al. Long-term impact of bile duct injury on morbidity, mortality, quality of life, and work related limitations. Ann Surg. 2018;268:143–50.
16. Kharel J, Yadav TD, Gupta R, Singh H. Long-term outcomes and quality of life after repair of benign biliary stricture of post-cholecystectomy bile duct injury. Indian J Gastroenterol. 2026;45:864–71.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
18. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.
19. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa: Ottawa Hospital Research Institute; 2000.
20. IntHout J, Ioannidis JPA, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol. 2014;14:25.
21. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.
22. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924–6.

23. Rystedt JM, Montgomery AK. Quality-of-life after bile duct injury: intraoperative detection is crucial. A national case-control study. HPB (Oxford). 2016;18:1010–6.
24. Karvonen J, Gronroos JM, Makitalo L, Koivisto M, Salminen P. Quality of life after iatrogenic bile duct injury: a case control study. Minim Invasive Ther Allied Technol. 2013;22:177–80.
25. Jabłońska B, Olakowski M, Lampe P, et al. Quality-of-life assessment in the treatment of iatrogenic bile duct injuries: hepaticojejunostomy versus end-to-end biliary reconstructions. ANZ J Surg. 2012;82:923–7.
26. Dominguez-Rosado I, Mercado MA, Kauffman C, et al. Quality of life in bile duct injury: 1-, 5-, and 10-year outcomes after surgical repair. J Gastrointest Surg. 2014;18:2089–94.
27. Doğan C, Borazan E, Yılmaz L, Balık AA. How much is the long-term quality of life impaired in cholecystectomy-related biliary tract injury? Turk J Surg. 2023;39:34–42.
28. Otto W, Sierdzinski J, Smaga J, et al. Long-term effects and quality of life following definitive bile duct reconstruction. Med (Baltim). 2018;97:e12684.
29. Keurentjes JC, Van Tol FR, Fiocco M, et al. Minimal clinically important differences in health-related quality of life after total hip or knee replacement: a systematic review. Bone Joint Res. 2012;1:71–7.

Figures

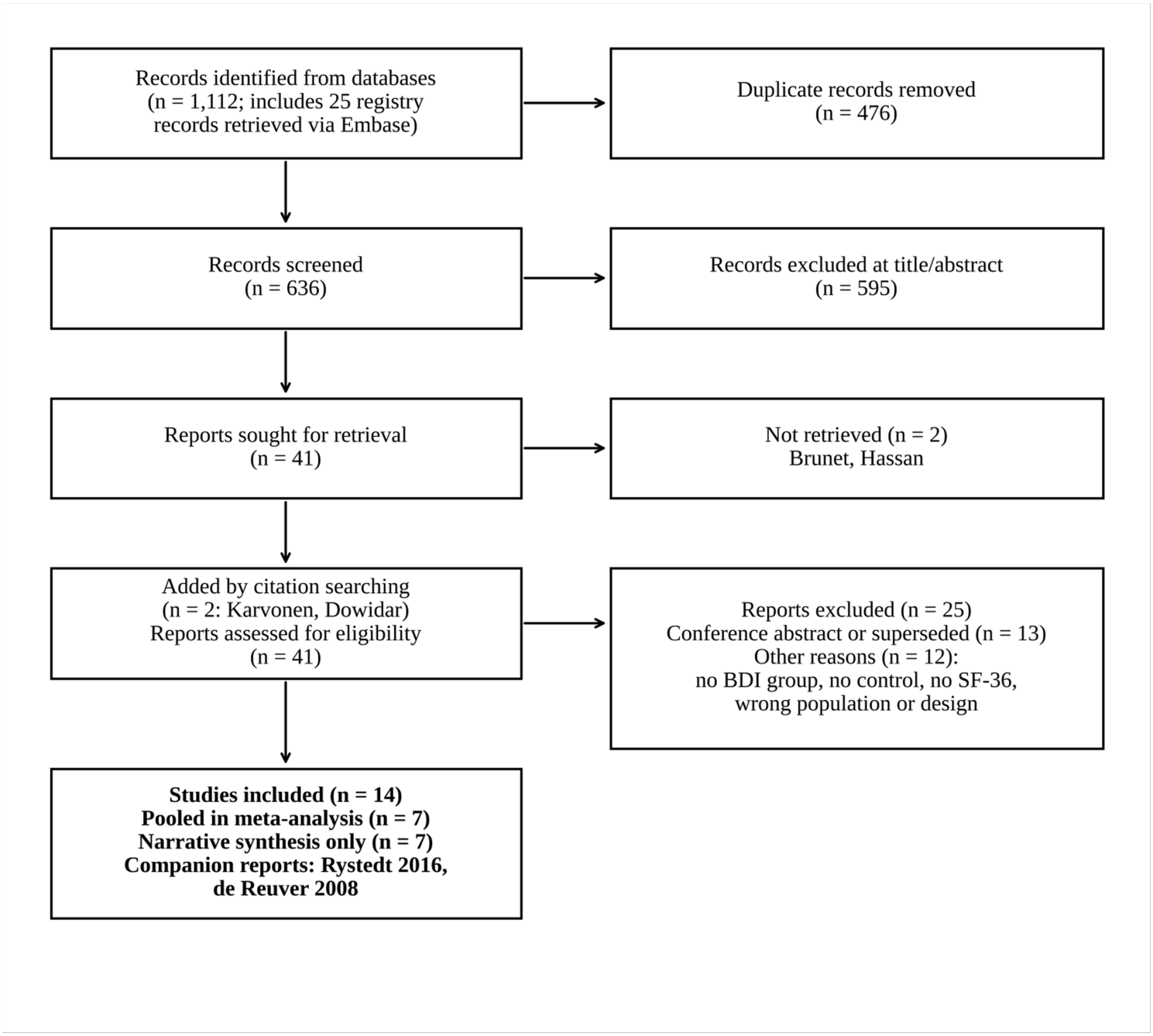

Figure 1

PRISMA 2020 flow diagram

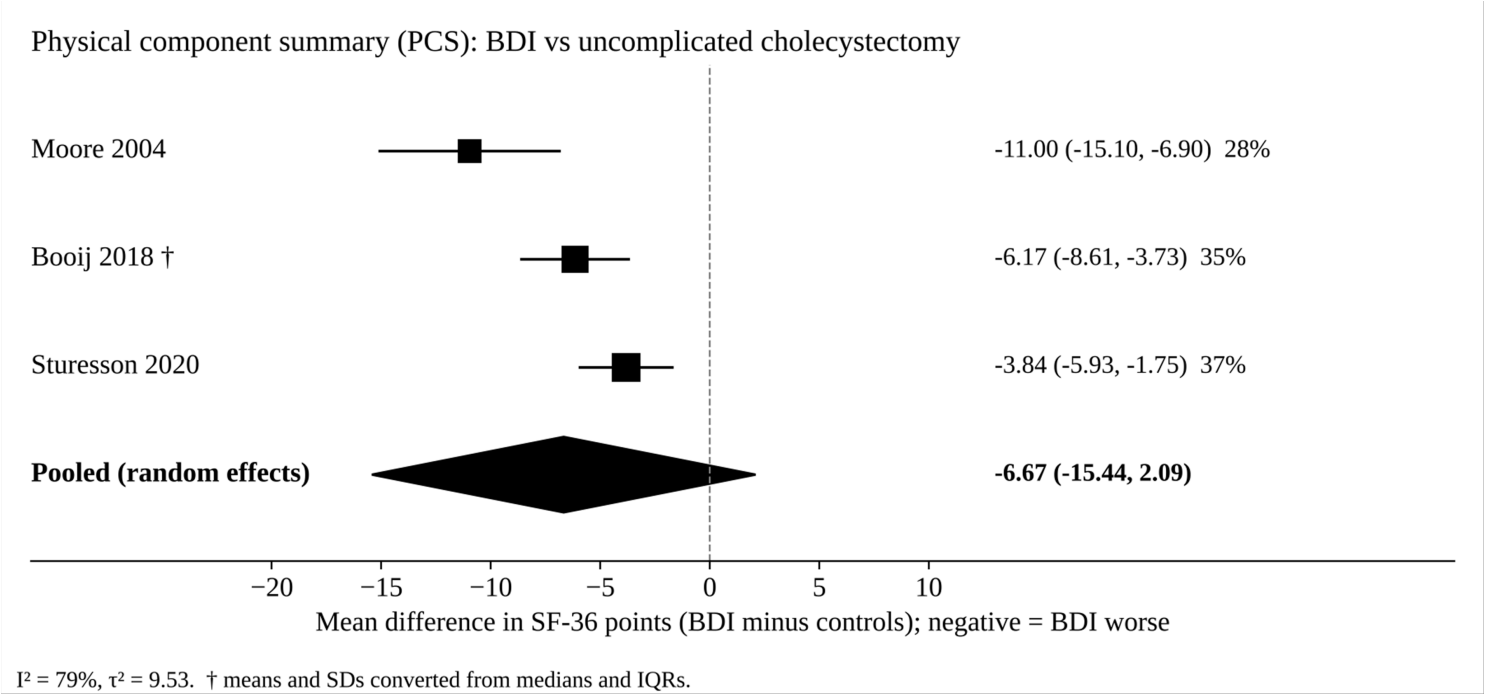

Figure 2

Primary analysis: forest plot of the mean difference in the SF-36 physical component summary, BDI versus uncomplicated cholecystectomy. Negative values favour controls. † Means and standard deviations converted from medians and interquartile ranges

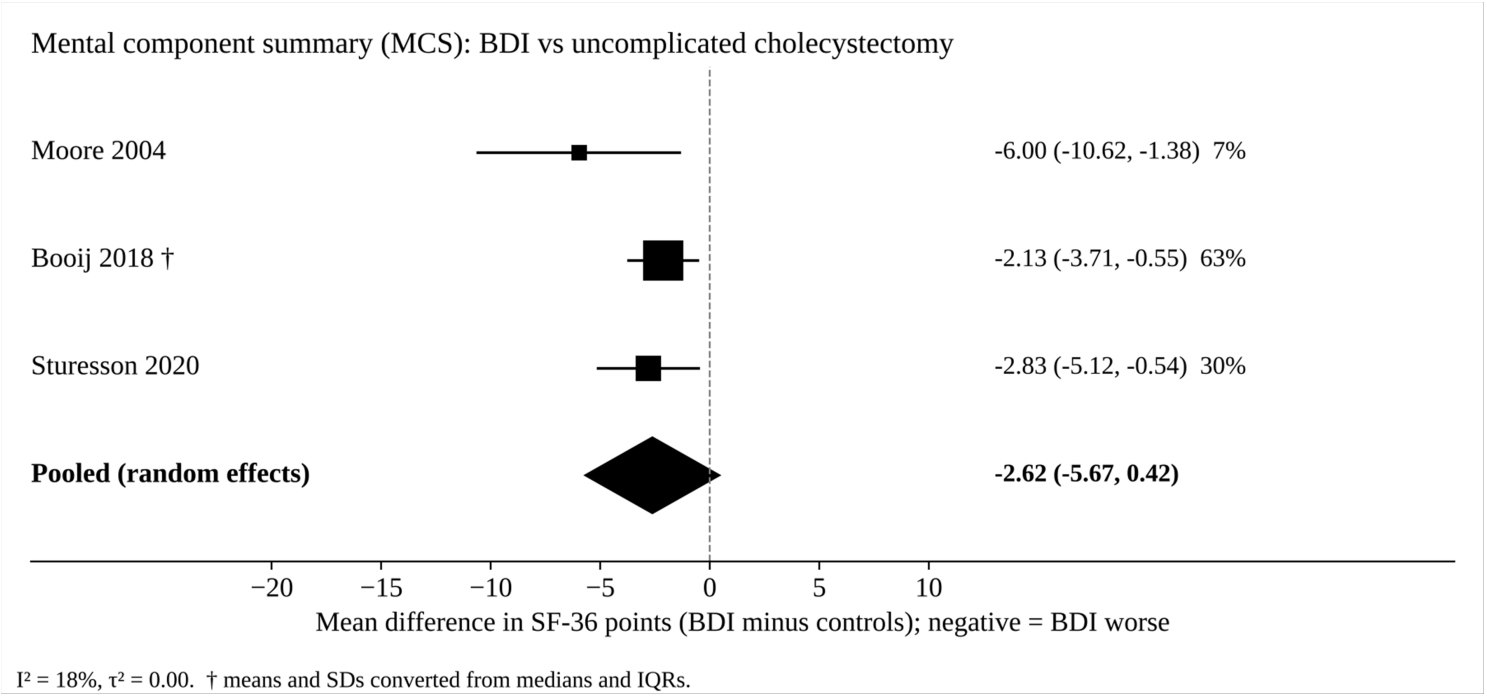

Figure 3

Primary analysis: forest plot of the mean difference in the SF-36 mental component summary, BDI versus uncomplicated cholecystectomy. Negative values favour controls. † Means and standard deviations converted from medians and interquartile ranges

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx
- SupplementaryFiguresTIFF.zip