Appendix I: GRADE profiles

GRADE and Modified GRADE approaches used in this guideline

Criteria	Standard GRADE approach	Modified GRADE approach	
	Effectiveness evidence (Developed by GRADE working group)	**Diagnostic test accuracy evidence (Modified by Schunemann et al, based on QUADAS checklist)**	**Prognostic evidence (Developed by Internal Clinical Guidelines team, based on QUIPS checklist)**
Starting point	• High if RCT design, Low if any other design	• High if RCT, cohort or cross sectional design, Low if any other design	• High if prospective study designs, Low if retrospective study designs
Risk of bias	• Appraised using relevant methodology checklists	• Appraised using relevant methodology checklists	• Appraised using relevant methodology checklists
Indirectness	• Indirect population, intervention, comparator, or outcome • Indirect comparison e.g. no head to head comparisons (no A to B comparison but A to C and C to B are available)	• Differences in population studied and population where recommendations will be applied • Differences in diagnostic expertise in the study and in those intended to use the test • Diagnostic tests are not directly compared (each test is only compared to the reference standard)	• Differences in the population, prognostic factors or outcomes of the included evidence compared with those for whom the recommendation is intended
Inconsistency	• Inconsistency of point estimates • Confidence intervals that don't overlap • Statistical measures of heterogeneity are high (e.g. I^2)	• Unexplained inconsistency in sensitivity, specificity or likelihood ratios	• Unexplained inconsistency in point estimates
Imprecision	• Few participants, few events (<300) and wide confidence intervals (where clinical action would differ if the upper or the lower boundary of the confidence interval represented the truth) • The optimal information size has not been met	• Wide confidence intervals for test accuracy	• Few participants, and wide confidence intervals (where clinical action would differ if the upper or the lower boundary of the confidence interval represented the truth)
Upgrading	• Observational evidence can be upgraded when no downgrading has taken place AND one or more of the following is satisfied: • There is evidence of a large effect • Plausible biases underestimate true effect • There is evidence of a dose-response gradient	• Not applicable	• Not applicable

I.1 Signs, symptoms and risk factors for gallstone disease (Question 1) GRADE profile

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	With gallstone disease	Without gallstone disease	Relative (95% CI)	Absolute		
Risk factors												
1 Wegge (1985)	Prospective cohort	serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	None	-	-	-[3]	-[3]	VERY LOW	CRITICAL

[1] Did not use all patients that were available. Only those presenting during office hours were included in the analysis
[2] Study does not report data from the multivariate analysis. Study was underpowered for the number of variables entered into the analysis
[3] Not reported

I.2 Diagnosing gallstone disease (Question 2) GRADE profile

I.2.1 Diagnosing gallbladder stones

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Sensitivity (95% CI)	Specificity (95% CI)		
Ultrasound compared to surgery											
1 Ahmed (2001)	Prospective cohort	serious[1]	no serious inconsistency	serious[2]	serious[3]	None	1869	1.00 (1.00, 1.00)	0.14 (0.11, 0.17)	VERY LOW	CRITICAL

[1] No description of inclusion/exclusion criteria and no definitions provided for the population term used. Unclear if researchers were blinded when interpreting the results of the test.
[2] unclear if the population selected is appropriate to answer the research question
[3] Lower bound of one of the confidence intervals was at or below 0.50

I.2.2 Diagnosing cholecystitis

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Sensitivity (95% CI)	Specificity (95% CI)		
MRCP compared to Surgery											
2 Hakansson (2000), Park (1998)	Prospective cohort	serious[1,2]	no serious inconsistency	no serious indirectness	serious[3]	none	70	0.89 (0.70, 0.96)	0.89 (0.50, 0.99)	LOW	CRITICAL
Ultrasound compared to surgery											
3	Prospective	serious[1]	serious[4]	no serious	serious[3]	none	100	0.71 (0.28,	0.88 (0.64,	VERY	CRITICAL

De Vargas (2006), Hakansson (2000), Park (1998)	cohort			indirectness				0.94)	0.97)	LOW	
MRI compared to surgery											
1 Altun (2007)	Prospective cohort	serious[2]	serious[5]	serious[6]	serious[3]	none	32	0.95 (0.71, 0.99)	0.69 (0.41, 0.88)	VERY LOW	CRITICAL
CT compared to surgery											
1 De Vargas (2006)	Prospective cohort	serious[2]	no serious inconsistency	no serious indirectness	serious[8]	none	12	0.95 (0.53, 1.00)	0.88 (0.27, 0.99)	LOW	CRITICAL

[1] In Hakansson (2000) there is possible selection bias as more than half of eligible participants were excluded because they presented outside of office hours. In Park (1998) it is unclear how participants were selected (random/consecutive/ self-selected).
[2] Patients were selected retrospectively, this could lead to selection bias. Unclear if the number included in the study was all the available cases or whether it is a sample of the available cases.
[3] Lower bound of one of the confidence intervals was at or below 0.50
[4] Individual study estimates vary widely
[5] Reference standard and index test were performed a month apart. This is an inappropriate interval.
[6] Study aims to differentiate between different types of cholecystitis.
[8] Lower bound of both of the confidence intervals were at or below 0.50

I.2.3 Diagnosing common bile duct stones

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations		Sensitivity (95% CI)	Specificity (95% CI)		
MRCP compared to ERCP											
8 Chan (1996), Regan (1996), Soto (2000), Griffin (2003), Kondo (2005), Stiris (2000), Sugiyama (1998), Holzknecht (1998)	Prospective cohort	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	470	0.83 (0.72, 0.91)	0.90 (0.83, 0.95)	MODERATE	CRITICAL
Ultrasound compared to ERCP											
5 Jovanovic (2011) Regan (1996), Rickes (2006), Sugiyama (1997), Sugiyama (1998)	Prospective cohort	serious[1]	serious[2]	no serious indirectness	no serious imprecision	none	383	0.70 (0.52, 0.83)	0.88 (0.63, 0.97)	LOW	CRITICAL
1 Karki (2013)	Prospective cohort	serious[1]	no serious inconsistency	no serious indirectness	very serious[9]	none	88	100	89	VERY LOW	CRITICAL
Endoscopic ultrasound compared to ERCP											
3 Kondo (2005), Polkowski (1999), Sugiyama (1997)	Prospective cohort	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	220	0.94 (0.87, 0.97)	0.94 (0.41, 1.00)	LOW	CRITICAL
CT cholangiography compared to ERCP											
4 Kondo (2005), Polkowski (1999), Soto (2000), Soto (1999)	Prospective cohort	serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	108	0.82 (0.67, 0.91)	0.84 (0.72, 0.92)	MODERATE	CRITICAL

CT compared to ERCP											
3 Soto (2000), Sugiyama (1997), Tseng (2008)	Prospective cohort	serious[5]	no serious inconsistency	no serious indirectness	no serious imprecision	yes[6]	51	0.76 (0.69, 0.81)	0.90 (0.66, 0.97)	MODERATE	CRITICAL
Predictive model ***(Model predicts CBDS by measuring alkaline phosphatase, total bilirubin, amylase, dilated common bile duct>8mm. Bile duct stones are predicted in patients with one or more of the four factors. An absence of bile duct stones is predicted in patients with none of the four factors)***											
1 Shiozawa (2005)	Prospective cohort	serious[7]	no serious inconsistency	no serious indirectness	no serious imprecision	none	513	0.98 (0.94, 1.00)	0.95 (0.92, 1.00)	MODERATE	CRITICAL
Predictive model ***(Model predicts CBDS by measuring gamma-glutamyl transferase, common bile duct diameter, and amylase. Values are entered into an equation and scores ≥0 predict CBDS, all other scores predict an absence of gallstones)***											
1 Barr (1999)	Retrospective cohort	serious[7]	no serious inconsistency	no serious indirectness	serious[4]	none	107	0.87 (0.60, 0.98)	0.71 (0.49, 0.89)	VERY LOW	CRITICAL
Predictors of common bile duct stones											
1 Alponat (1997)	Retrospective cohort	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[8]	none	192	**Aspartate aminotransferase** Adj OR= 2.9 (1.25, 5.88) **CBD diameter** Adj OR= 2.9 (2.85, 18.99), **Cholangitis** Adj OR= 5.30 (1.55, 71.79)		VERY LOW	CRITICAL

[1] Methods of patient selection were unclear and selection bias could be present. Most studies did not state if the researchers/clinicians were blinded to the results of the reference standard when interpreting index test result. Studies employed different exclusion criteria
[2] Wide variation in individual studies estimates of sensitivity and specificity.
[3] Most studies did not state if the researchers/clinicians were blinded to the results of the reference standard when interpreting index test result. Studies employed different exclusion criteria.
[4] Lower bound of one of the confidence intervals was at or below 0.50
[5] Studies used different exclusion criteria
[6] Tseng (2008) divides participants into 3 separate groups who each undergo CT but with 3 different scanning parameters. Unclear why this was done as results are not presented separately for the 3 different groups
[7] Unclear if researchers were blinded to results of the reference standard when interpreting the results of the index tests.
[8] Wide confidence intervals for one or more predictive factor, where clinical action would differ if the upper or lower confidence interval were true.
[9] Insufficient data are provided to enable calculation of confidence intervals.

I.3 Risk factors for asymptomatic complications (Question 3) GRADE profile

Quality assessment							No of patients	Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**No of patients**	**Relative (95% CI)**	**Absolute**		
Risk factors											
1 Attili (1995)	Prospective cohort	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	118	Biliary colic No significant predictors Complications The low number of events meant that analysis was		VERY LOW	CRITICAL

								not possible Cholecystectomy Occurrence of biliary colic predicted cholecystectomy Death No associations between potential predictive factors and death were reported.		

[1] Study population may not be representative of the actual population: participants were invited to participate in screening which may have led to self selection of people with different symptoms and risk factors to those in the general population. Men and women were recruited in different recruitment rounds that took place in different years. Unclear why this approach was taken.
[2] Inappropriate statistical analysis: not all analyses are reported and those that are only report p values.

I.4 Managing asymptomatic gallbladder stones (Question 4a) GRADE profile

No evidence was found

I.5 Managing symptomatic gallbladder stones (Question 4b) GRADE profile

I.5.1 Laparoscopic cholecystectomy compared to laparoscopic cholecystectomy + intraoperative cholangiography

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LC	LC+IOC	Relative (95% CI)	Absolute		
Bile leak												
1 Soper (1992)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	0/59 (0%)	0/56 (0%)	-[2]	-[2]	MODERATE	CRITICAL
Bile duct injury												
3 Amott (2005), Khan (2011), Soper 1992)	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	2/306 (0.65%)	2/302 (0.66%)	RR 0.98 (0.17 to 5.59)	0 fewer per 1000 (from 5 fewer to 30 more)	LOW	CRITICAL
Length of stay (Better indicated by lower values)												
1 Soper (1992)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[5,6]	none	1	1	-	mean 0 higher (0 to 0 higher)	MODERATE	IMPORTANT
Missed common bile duct stones												
3 Amott (2005), Khan	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[4]	none	0/306 (0%)	1/302 (0.33%)	RR 0.56 (0.15 to 2.04)	1 fewer per 1000 (from 3 fewer to 3 more)	LOW	IMPORTANT

(2011), Soper 1992)												

[1] Study wasn't powered to detect any event in either arm
[2] Unable to analyse zero event data
[3] Inappropriate method of randomisation (month of birth).
[4] Low event rates
[5] Small sample size
[6] No measures of dispersion are reported

I.5.2 Laparoscopic cholecystectomy compared to Cholecystostomy

No evidence was found

I.5.3 Laparoscopic cholecystectomy compared to conservative management

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**LC**	**Conservative management**	**Relative (95% CI)**	**Absolute**		
Disease progression												
1[1] Schmidt (2011)	randomised trials	no serious risk of bias	serious[2]	no serious indirectness[3]	serious[4]	none	4/99 (4%)	13/102 (12.7%)	RR 0.32 (0.11 to 0.93)	87 fewer per 1000 (from 9 fewer to 113 fewer)	LOW	CRITICAL
Additional intervention required (assessed with: Cholecystectomy)												
1[1] Schmidt (2011)	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	serious[4]	none		45/102 (44.1%) required cholecystectomy in the conservative management group			LOW	CRITICAL
Readmission (assessed with: Biliary pain)												
1[1] Schmidt (2011)	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	serious[4]	none	5/99 (5.1%)	19/102 (18.6%)	RR 0.33 (0.06 to 1.97)	125 fewer per 1000 (from 175 fewer to 181 more)	LOW	IMPORTANT
Length of stay - not reported												
1	-	-	-	-	-	none	-	-	-	-		IMPORTANT
Mortality												
1[1] Schmidt (2011)	randomised trials	no serious risk of bias	serious[2]	no serious indirectness	serious[4]	none	8/99 (8.1%)	4/102 (3.9%)	RR 2.20 (0.25 to 19.39)	47 more per 1000 (from 29 fewer to 721 more)	LOW	IMPORTANT

[1] A single study reported in 6 separate publications (Vetrhus 2002, 2003, 2004, 2005; Schmidt 2011a, 2011b)
[2] Lack of consistency between publications from this study- same outcomes are reported differently, and data don't always add up, therefore there is some ambiguity as to whether the correct outcome and correct numbers are used in this analysis.
[3] Some patients had open surgery. Not downgraded as the majority of patients did receive laparoscopic surgery.
[4] Small sample size and few event rates

I.5.4 Day case laparoscopic cholecystectomy compared to planned inpatient laparoscopic cholecystectomy

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Day LC	Overnight LC	Relative (95% CI)	Absolute		
Failed day case discharge												
3 Hollington (1999), Johansson (2006), Keulemans (1998)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	18/149 (12.1%)	18/49 (12.1%) of planned day cases required an inpatient admission -			MODERATE	CRITICAL
Readmission												
4 Barthelsson (2008), Hollington (1999), Johansson (2006), Keulemans (1998)	randomised trials	serious[2]	no serious inconsistency	no serious indirectness	serious[1]	none	3/145 (2.1%)	3/161 (1.9%)	RR 1.17 (0.27 to 5.05)	3 more per 1000 (from 14 fewer to 75 more)	LOW	CRITICAL
Length of stay												
3 Hollington (1999), Johansson (2006), Keulemans (1998)	randomised trials	serious[3]	serious[4]	no serious indirectness	no serious imprecision	none	**Hollington (1999)** 31/71 day case patients required prolonged hospitalisation of 2 days or more **Johansson (2006)** 48/52 day case patients were discharged within 4-6 hrs (4 patients were admitted), 42/48 inpatients were discharged on the first day after surgery, 6/48 inpatients were discharged on the second day after surgery **Keulemans (1998)** post surgical length of stay was Mean=7.2 SD= 0.8 hrs for the day case group and Mean =31 SD=3 for the inpatient group				LOW	IMPORTANT
Mortality - not reported												
0	-	-	-	-	-	none	-	-	-	-		IMPORTANT
Quality of life (measured with: Health Index ; Better indicated by higher values)												
3 Barthelsson (2008), Johansson (2006), Keulemans (1998)	randomised trials	serious[3]	no serious inconsistency	no serious indirectness	serious[5]	none	204	205	-	SMD 0.29 higher (0.42 lower to 1.01 higher)	LOW	IMPORTANT

[1] Low event rates
[2] Study/Studies does provide details about randomisation procedures
[3] Study/Studies does provide details about randomisation procedures
[4] Assessment of outcome is different across all studies
[5] Individual studies have small sample sizes

I.6 Managing Common bile duct stones (Question 4c)

I.6.1 ERCP + Laparoscopic cholecystectomy compared to ERCP alone.

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ERCP+LC	ERCP alone	Relative (95% CI)	Absolute		
Quality of life - not reported												
0	-	-	-	-	-	none	-	-	-	-		CRITICAL
Recurrence/disease progression												
2 Boerma (2002), Lau (2006)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	7/138 (5.1%)	48/148 (32.4%)	RR 0.14 (0.02 to 0.96)	279 fewer per 1000 (from 13 fewer to 318 fewer)	MODERATE	CRITICAL
Additional intervention required (ERCP)												
2 Boerma (2002), Lau (2006)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	0/138 (0%)	22/148 (14.9%)	RR 0.05 (0.01 to 0.39)	141 fewer per 1000 (from 91 fewer to 147 fewer)	MODERATE	CRITICAL
Additional intervention required (Cholecystectomy)												
2 Boerma (2002), Lau (2006)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	0/138 (0%)	38/148 (25.7%)	RR 0.03 (0 to 0.2)	249 fewer per 1000 (from 205 fewer to 257 fewer)	MODERATE	CRITICAL
Mortality												
2 Boerma (2002), Lau (2006)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	11/138 (8%)	19/141 (13.5%)	RR 0.58 (0.29 to 1.15)	57 fewer per 1000 (from 96 fewer to 20 more)	MODERATE	IMPORTANT
Length of stay												
0	No evidence available					none	-	-	-	-		IMPORTANT

[1] Few participants

I.6.2 ERCP compared to conservative management

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	ERCP	Conservative mangement	Relative (95% CI)	Absolute		
Mortality												

8 Acosta (2006), Fan (1993), Folsch (1997), Hui (2002), Neoptolemos (1998), Nitsche (1995) Oria (2007), Vracko (2006)	randomised trials	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	none	25/515 (4.9%)	25/530 (4.7%)	RR 1.00 (0.54 to 1.86)	0 fewer per 1000 (from 22 fewer to 41 more)	VERY LOW	CRITICAL
Disease progression												
6 Acosta (2006), Fan (1993), Folsch (1997), Nitsche (1995), Hui (2002), Vracko (2006)	randomised trials	very serious[1]	serious[4]	no serious indirectness	no serious imprecision	none	74/403 (18.4%)	117/407 (28.7%)	RR 0.48 (0.2 to 1.14)	149 fewer per 1000 (from 230 fewer to 40 more)	VERY LOW	CRITICAL
Additional intervention required (ERCP)												
6 Acosta (2006), Fan (1993), Folsch (1997), Neoptolemos (1998), Nitsche (1995), Oria (2007)	randomised trials	very serious[2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	0/411 (0%)	78/407 (19.2%)	RR 0.05 (0.02 to 0.16)	182 fewer per 1000 (from 161 fewer to 188 fewer)	LOW	IMPORTAN
Additional intervention required (Cholecystectomy)												
4 Acosta (2006), Oria (2007), Vracko (2006), Zhou (2002).	randomised trials	serious[2]	serious[4]	no serious indirectness	no serious imprecision	none	105/153 (68.6%)	100/161 (62.1%)	RR 1.24 (0.57 to 2.72)	149 more per 1000 (from 267 fewer to 1000 more)	LOW	IMPORTANT
Length of stay												
0	No evidence available					none	-	-	-	-		IMPORTANT

[1] Unequal intervention and control groups: Hui (2002) found significant differences in baseline liver function tests between intervention and control groups
[2] Some studies fail to assess baseline characteristics between intervention and control groups.
[3] Some studies fail to report randomisation procedures
[4] Differences between studies regarding patient characteristics: Majority focus on patients with pancreatitis, Hui (2002) focuses on patients with cholangitis, Vracko (2006) focuses on patients with cholecystitis.
[5] Low event rates

I.6.3 Biliary stent compared to cleared duct

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Stents**	**Duct clearance**	**Relative (95% CI)**	**Absolute**		
Mortality												
1 Chopra (1996)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	4/43 (9.3%)	2/43 (4.7%)	RR 2.00 (0.39 to 10.35)	47 more per 1000 (from 28 fewer to 435 more)	LOW	CRITICAL
Disease progression												

1 Chopra (1996)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	10/43 (23.3%)	8/43 (18.6%)	RR 1.25 (0.55 to 2.86)	47 more per 1000 (from 84 fewer to 346 more)	LOW	CRITICAL
Additional intervention required (ERCP)												
1 Chopra (1996)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	13/39 (33.3%)	33/43 (76.7%)	RR 0.43 (0.27 to 0.7)	437 fewer per 1000 (from 230 fewer to 560 fewer)	LOW	IMPORTANT
Additional intervention (Cholecystectomy)												
1 Chopra (1996)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	5/39 (12.8%)	3/43 (7%)	RR 1.84 (0.47 to 7.19)	59 more per 1000 (from 37 fewer to 432 more)	LOW	IMPORTANT
Length of stay - not reported												
0	-	-	-	-	-	none	-	-	-	-		IMPORTANT

[1] No assessment of baseline characteristics. Unclear if there were important differences between groups that could confound the results
[2] Small sample size and few events

I.6.4 Day case ERCP compared to planned inpatient ERCP

No evidence was found

I.6.5 ERCP clearance of bile duct with laparoscopic cholecystectomy compared to surgical clearance of bile duct with laparoscopic cholecystectomy

I.6.5.1 Network comparisons

Quality assessment							Network meta analysis results						
									In comparison to surgical BDE		Median rank (95% Ci)	Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Option	Direct estimate	Indirect estimate	Probably best			
Length of stay													
8 Bansal (2010), Cuschieri (1999), ElGeidie (2011a), Elgeidie (2011b), Hong (2006), Noble	randomised trials	serious[2,3]	serious[4,5,6,7,8]	no serious indirectness	serious[9]	none	Surgical BDE	-	-	0.530	1 (1,3)	VERY LOW	CRITICAL
							Pre-op ERCP	0.42 (-3.62, 4.46)	0.81 (-0.53, 2.15)	0.067	3 (1,4)		
							Intra-op ERCP	0.56 (-0.38, 1.50)	0.17 (-1.20, 1.55)	0.369	2 (1,4)		
							Post-op ERCP	2.50	2.50	0.035	4 (1,4)		

(2009), Rhodes (1998), Rogers (2010)								(1.11, 3.89)	(-0.39, 5.41)				
Missed stones													
5 Elgeidie (2011a), Hong (2006), Nathanson (2005), Noble (2009), Sgourakis (2002)	randomised trials	serious[2,11]	serious[4,5,7]	no serious indirectness	serious[9]	none	Surgical BDE	-	-	0.026	2 (1,4)	VERY LOW	IMPORTANT
							Pre-op ERCP	3.53 (0.31, 39.90)	5.13 (0.40, 76.22)	0.011	4 (2,4)		
							Intra-op ERCP	0.28 (0.05, 1.70)	0.10 (0.00, 1.41)	0.892	1 (1,3)		
							Post-op ERCP	1.86 (0.16, 21.32)	2.36 (0.06, 168.30)	0.071	3 (1,4)		
Failed procedure													
9 Bansal (2010), Cuschieri (1999), ElGeidie (2011a), ElGeidie (2011b), Hong (2006), Noble (2009), Rhodes (1998), Rogers (2010), Sgourakis (2002)	randomised trials	serious[2,3,11]	serious[4,5,6,7,8]	no serious indirectness	serious[9]	none	Surgical BDE	-	-	0.164	2 (1,3)	VERY LOW	IMPORTANT
							Pre-op ERCP	4.62 (1.08, 19.72)	5.23 (1.51, 24.28)	0.001	4 (3,4)		
							Intra-op ERCP	0.68 (0.33, 1.39)	0.76 (0.19, 3.14)	0.476	2 (1,3)		
							Post-op ERCP	1.00 (0.36, 2.75)	0.99 (0.08, 11.72)	0.358	2 (1,4)		
Conversion to open surgery													
9 Bansal (2010), Cuschieri (1999), ElGeidie (2011a), ElGeidie (2011b), Hong (2006), Nathanson (2005), Noble (2009), Rhodes (1998), Sgourakis (2002).	randomised trials	serious[2,3,11]	serious[4,5,6,7,8]	no serious indirectness	serious[9]	none	Surgical BDE	-	-	0.027	3 (1,4)	VERY LOW	IMPORTANT
							Pre-op ERCP	0.84 (0.28, 2.47)	0.86 (0.32, 3.13)	0.104	3 (1,4)		
							Intra-op ERCP	0.71 (0.34, 1.48)	0.65 (0.17, 2.41)	0.249	2 (1,4)		
							Post-op ERCP	0.58 (0.07, 4.86)	0.34 (0.01, 6.23)	0.620	1 (1,4)		

[1] Not used
[2] Inappropriate randomisation: Hong (2006) uses patient identifying numbers
[3] Unclear randomisation procedures used by Rhodes (1998)
[4] Differences in inclusion criteria and patient comorbidities/symptoms- most studies indicate than no specific exclusion criteria were used, where as other studies impose selective exclusion criteria.
[5] ElGeidie (2011a) excludes patients with acute cholangitis, gallstone pancreatitis,ASA grades IV-V
[6] Elgeidie (2011b) excludes patients with cholangitis, pancreatitis, ASA IV-V
[7] Noble (2009) includes high risk patients (over 70 years of age, over 60 with a co morbidity or over 50 with a BMI >40)

[8] Rogers (2010) Excludes patients with ASA status >II, suppurative cholangitis, severe pancreatitis.
[9] Wide credibility intervals for rankings within the network.
[11] Unclear randomisation procedures used by Sgourakis (2002)

I.6.5.2 Pairwise comparisons

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	LC+ERCP	LC+BDE	Relative (95% CI)	Absolute		
More than one ERCP required												
4 Bansal (2010), Cuschieri (1999), Nathanson (2005), Rhodes (1998)	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	Pre op ERCP: 9/165 (5%) required >1 ERCP Intra op ERCP: not reported Post op ERCP: 18/85 (21%) required ERCP				LOW	CRITICAL
Mortality												
3 Cuschieri (1999), Noble (2009), Sgourakis (2002)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[3]	none	5/225 (2.2%)	11/213 (5.2%)	RR 0.43 (0.15 to 1.23)	29 fewer per 1000 (from 44 fewer to 12 more)	MODERATE	CRITICAL

I.7 Timing of intervention (Question 5)

I.7.1 Early compared to delayed laparoscopic cholecystectomy for acute cholecystitis.

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Early LC	Delayed LC	Relative (95% CI)	Absolute		
Readmission due to symptoms												
4 Johansson (2003), Lai (1998), Lo (1998), Macafee (2009)	randomised trials	no serious risk of bias	no serious inconsistency	serious[1]	serious[2]	none	0/208 (0%)	36/191 (18.8%)	RR 0.05 (0.01 to 0.2)	179 fewer per 1000 (from 151 fewer to 187 fewer)	LOW	CRITICAL
Readmission due to surgical complications - not reported												
0	-	-	-	-	-	none	-	-	-	-		CRITICAL
Length of stay (Better indicated by lower values)												
6 Johansson (2003), Kolla (2004), Lai (1998), Lo (1998), Macafee (2009), Yadav (2009)	randomised trials	no serious risk of bias	no serious inconsistency	serious[3]	no serious imprecision	none	207	263	-	MD 3.29 lower (4.67 to 1.9 lower)	MODERATE	IMPORTANT

1 Gul (2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[6,7]	none	30	30	mean early= 4.77 days mean delayed= 10.10 days	LOW	IMPORTANT
Mortality											
4 Johansson (2003), Kolla (2004), Lai (1998), Lo (1998)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[4]	none	0/213 (0%)	0/205 (0%)	-[5]	-	MODERATE	IMPORTANT
Quality of life – measured by mean VAS scores											
1 Gul (2013)	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[7]	none	30	30	-	**1 hour** MD 0.57 higher (0.21 to 0.93 higher) **12 hours** MD 3.17 higher (2.41 to 3.93 higher) **24 hours** MD 0.33 higher (0.10 lower to 0.76 higher) **48 hours** MD 0.19 higher (0.08 lower to 0.46 higher)	MODERATE	IMPORTANT

[1] Macafee (2009) also includes patients with biliary colic. Unclear what proportion of the study sample this represents. This study represents approximately 18% of the population included in this outcome.
[2] Few events
[3] Length of stay was reported inconsistently by individual studies, therefore data reported have been converted into means and standard deviations
[4] Zero events were observed in both arms of the trial in all studies. Studies were underpowered to detect differences in mortality.
[5] Not estimable due to zero events in both arms
[6] Insufficient data are provide to estimate confidence intervals,
[7] Small sample size

I.7.2 Early compared to delayed laparoscopic cholecystectomy after ERCP for common bile duct stones

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**ERCP followed by Early LC**	**ERCP followed by Delayed LC**	**Relative (95% CI)**	**Absolute**		
Readmission due to symptoms - not reported												
0	No evidence available	-	-	-	-	none	-	-	-	-		CRITICAL
Readmission due to surgical complications												
0	No evidence available					none	-	-	-	-		CRITICAL

Length of stay (Better indicated by lower values)												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	serious[1]	none	47	47	-	MD 0.27 higher (0.88 lower to 1.42 higher)	MODERATE	IMPORTANT
Mortality												
1	randomised trials	no serious risk of bias	no serious inconsistency	no serious indirectness	very serious[1,2]	none	0/86 (0%)	0/87 (0%)	-[3]	-	LOW	IMPORTANT
Quality of life - not reported												
0	No evidence available	-	-	-	-	none	-	-	-	-		IMPORTANT

[1] Small sample size
[2] Zero events- studies were not adequately powered to detect differences in mortality.
[3] Effect size cannot be estimated due to zero event data in both arms

I.8 Patient information (Question 6)

Quality assessment					Results	Quality	Importance
No of studies	Design	Limitations	Transferability	Other			
Diet							
3 Blay, 2005; Blay, 2006; Young, 2001	Mixed[1]	very serious[2]	no serious issues	no other considerations	Patients and carers requested additional information on diet and fluids.	VERY LOW	CRITICAL
Wounds							
3 Blay, 2005; Blay 2006; Young, 2001	Mixed[1]	very serious[3]	no serious issues	no other considerations	Patients had questions on how their wounds should be cared for. Patients and carers requested additional information on wound care.	VERY LOW	CRITICAL
Pain management							
2 Blay, 2005; Young 2001	Mixed[1]	very serious[2]	no serious issues	no other considerations	Patients and carers requested additional information on pain management.	VERY LOW	CRITICAL
Resuming activity							
2 Blay, 2005; Blay 2006	Mixed[1]	very serious[3]	no serious issues	no other considerations	65% had not been told how long it would take to return to normal activities 2/23 requested information on activity 6% requested information on post operative activity	VERY LOW	CRITICAL
Memory							
1 Barthelsson, 2003	Mixed[1]	very serious[2]	no serious issues	no other considerations	Several respondents had no memory of the information given to them by the surgeon on discharge from hospital.	VERY LOW	CRITICAL
General							
1 Young, 2001	Mixed[1]	very serious[2]	no serious issues	no other considerations	100% of day case patients and carers had sufficient discharge information, compared to 44% of overnight patients and	VERY LOW	CRITICAL

					55.6% of overnight carers who thought they had sufficient discharge information.		
Information seeking							
1 Tamahanka, 2009	Mixed[1]	very serious[2]	serious[3]	no other considerations	31% of patients with internet access used it to acquire additional information about their operations and 58% used internet search engines to acquire additional information	VERY LOW	CRITICAL

[1] Modified GRADE approach used, where qualitative evidence was identified from a range of study designs (e.g. RCT, Prospective cohort, qualitative). Thus studies are grouped based on outcome type rather than study design as is used in the standard GRADE approach. Evidence from the various study designs was assessed using qualitative checklists as this was considered the most appropriate for the outcomes included in this review.
[2] Values/assumptions/theory underpinning the purpose of the study are not discussed. Lack of defensible/rigorous design/methodology. Absence of 'rich' findings.
[3] Study includes people waiting for hernia repair as well as those with gallstone disease. Separate analyses are not reporte