CrossMark

Evidence of surgical outcomes fluctuates over time: results from a cumulative meta-analysis of laparoscopic versus open appendectomy for acute appendicitis

Tomohiko Ukai[1], Satoru Shikata[2*], Hiromu Takeda[3], Lauren Dawes[4], Yoshinori Noguchi[5], Takeo Nakayama[6] and Yousuke C. Takemura[2]

Abstract

Background: In surgical trials, complex variables such as equipment development and surgeons' learning curve are involved. The evidence obtained in these trials can thus fluctuate over time. We explored the stability of the evidence obtained during surgery by conducting a cumulative meta-analysis of randomized controlled trials for open and laparoscopic appendectomy.

Methods: We conducted a cumulative meta-analysis of randomized controlled trials comparing laparoscopic appendectomy with open appendectomy for acute appendicitis, a topic with the greatest number of trials in the gastroenterological surgical field. We searched the MEDLINE (PubMed), EMBASE, and CINAHL databases up to September 2014 and reviewed the bibliographies. Outcomes were the incidence of intra-abdominal abscess, incidence of wound infection, operative time, and length of hospital stay. We used the 95 % confidence interval (95 % CI) of effect size for the significance test.

Results: Sixty-four trials were included in this analysis. Of the 51 trials addressing intra-abdominal abscesses, our cumulative meta-analysis of trials published up to and including 2001 demonstrated statistical significance in favor of open appendectomy (cumulative odds ratio [OR] 2.35, 95 % CI 1.30–4.25). The effect size in favor of open procedures began to disappear after 2001, leading to an insignificant result with an overall cumulative OR of 1.32 (95 % CI 0.84–2.10) when laparoscopic appendectomy was compared with open appendectomy.

Conclusions: The evidence regarding treatment effectiveness changed over time, after treatment effectiveness became significant in trials comparing laparoscopic and open appendectomy. Observing only the 95 % confidence interval of effect size from a meta-analysis may not provide conclusive results.

Keywords: Cumulative meta-analysis, Randomized controlled trials, Laparoscopic appendectomy, Open appendectomy

* Correspondence: shikata-s@umin.org
[2]Department of Family Medicine, Mie University School of Medicine & Graduate School of Medicine, 2-174 Edobashi, Tsu, Mie 514-8507, Japan
Full list of author information is available at the end of the article

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 2 of 12

Background

Meta-analyses of randomized controlled trials (RCTs), which combine the evidence presented in individual research reports, are expected to produce the highest level of evidence and have accordingly become increasingly important in health care [1]. The concept of a cumulative meta-analysis, which is reanalyzed each time the results of a new trial are published, was introduced by Lau et al. in 1992 [2]. This technique was designed to enable determinations of both clinical efficacy and harm as well as the tracking of trials and planning of future trials [3].

The 1992 study by Lau et al. indicated that two very large clinical trials on the efficacy of streptokinase for acute myocardial infarction [4, 5] may have been unnecessary because, according to their cumulative meta-analysis, the treatment efficacy was already statistically significant before those two trials were conducted. Later, cumulative meta-analyses of other topics demonstrated that statistically significant results in meta-analyses can later disappear, especially when well-powered and well-designed trials with sufficient numbers of outcomes and patients appear. In the surgical field, it is quite possible that once a surgical intervention is established, evidence regarding its effectiveness can change over time because of the complexity of surgical trials, which involve advances in surgical equipment and techniques, progress in surgeons' learning curves as they develop novel skills, and variations in postoperative management, among other factors [6].

To identify changes in the evidence obtained in surgical trials over time, we selected trials comparing the clinical effectiveness of laparoscopic appendectomy and open appendectomy for acute appendicitis. We considered this topic suitable for the observation of chronological trends because, to the best of our knowledge, this topic is associated with the highest number of RCTs in the gastroenterological surgical field [7]. In light of the existing meta-analyses on this topic, including a Cochrane review [8–10], our purpose was to identify any changes in the evidence over time rather than the superiority of one procedure over the other.

We asked the following clinical question: might the evidence demonstrated by a meta-analysis of RCTs of surgical procedures change over time? To answer this question, we conducted a cumulative meta-analysis of RCTs that had compared laparoscopic appendectomy with open appendectomy.

Methods

Herein we conducted a cumulative meta-analysis of RCTs to ascertain chronological trends in the comparison of laparoscopic appendectomy and open appendectomy for acute appendicitis. We used the cumulative meta-analysis technique introduced by Lau et al. in 1992 [2]. In a cumulative meta-analysis, studies are added one at a time according to their date of publication, and the results are summarized as each new study is added.

Literature search

We systemically searched the MEDLINE (PubMed), EMBASE, and CINAHL databases for articles in all languages that described RCTs published between 1991, when laparoscopic appendectomy was initiated, and September 2014. In MEDLINE, we utilized the CRD/Cochrane Highly Sensitive Search Strategy [11] with the search terms "appendectomy" and "appendicitis." We performed the EMBASE search strategy to optimize sensitivity and specificity [12] with the terms "appendectomy" or "appendicitis." We searched the CINAHL database using a strategy in which terms with the best optimization of sensitivity and specificity [13] were combined with "appendectomy" or "appendicitis." Reference lists of the review articles and previously published meta-analyses were searched by hand. The search was last done on December 18, 2014.

Selection criteria for studies in this review

Our inclusion criteria were as follows: (1) prospective RCTs, (2) studies comparing laparoscopic surgery and open surgery for acute appendicitis, (3) studies with human adult participants, and (4) studies written in any language. We excluded studies with any of the following characteristics: (1) pediatric participants, (2) comparisons of diagnostic efficacy, and (3) assessment of the effectiveness of variations of standard laparoscopic techniques, such as the single trocar technique versus the standard technique.

Outcome measures

Outcomes included the incidence of intra-abdominal abscess, the incidence of wound infection, the operative time, and the length of hospital stay. We adopted these four outcome measures because these are most frequently measured in RCTs addressing this topic.

Assessment of study quality

We assessed the risk of bias with respect to adequate sequence generation, allocation concealment, blinding, incomplete outcome data addressed, and selective reporting. Two authors (TU and HT) assessed the studies that met the inclusion criteria (Table 1).

Data extraction

Binary data were extracted for the incidences of intra-abdominal abscess and wound infection, and continuous data were extracted for the operative time and length of hospital stay. Two authors (TU and HT) independently undertook this process, and disagreements were resolved

Table 1 Outcomes and assessment of risk of bias in the included studies

Author	Year	Outcomes				Risk of bias				
		Wound infection	Abscess	Operative time	Length of hospital stay	Adequate sequence generation?	Allocation concealment?	Blinding?	Incomplete outcome data addressed?	Free of selective reporting?
Al-Mulhim	2002	○	○	–	–	?	+	?	?	?
Attwood	1992	○	○	–	–	?	+	?	+	?
Barth	1999	○	–	–	–	?	?	?	?	?
Bauwens	1999	○	○	○	○	?	+	?	-	+
Bruwer	2000	○	○	○	○	+	+	?	+	?
Cox	1996	○	○	○	○	?	?	?	?	?
Eichen	1994	○	○	○	○	+	+	?	?	?
Frazee	1994	○	○	○	○	?	?	?	+	?
Gouder	2011	○	–	–	○	?	?	?	-	?
Hall Long	2000	○	○	–	○	?	+	?	?	?
Hansen	1996	○	○	–	–	+	+	?	-	?
Hart	1996	○	○	○	○	+	+	?	-	+
Hebebrand	1994	○	○	○	○	?	+	-	+	+
Heikkinen	1998	○	○	–	–	?	+	?	?	?
Hellberg	1999	○	○	–	–	+	+	?	-	?
Helmy	2001	–	○	–	–	?	?	?	?	?
Henle	1996	○	○	–	–	+	?	?	-	?
Huang	2001	○	○	○	○	?	+	+	?	?
Ignacio	2003	○	○	○	○	+	+	+	+	?
Jan	2011	○	○	○	–	+	?	?	+	?
Kaiser	2006	○	○	○	○	?	+	?	+	+
Kald	1999	○	○	–	–	+	+	?	-	+
Kaplan	2009	○	–	○	○	?	?	?	+	?
Karadayi	2003	○	○	○	–	+	?	?	-	+
Kargar	2010	○	–	○	○	+	+	?	?	?
Katkhouda	2005	○	○	–	–	+	+	+	-	+
Kazemier	1997	○	○	○	○	+	+	?	+	?
Kehagias	2009	○	○	–	–	+	+	?	?	?
Khalil	2011	○	–	○	○	+	?	?	-	?
Kocatas	2013	○	○	–	○	+	+	?	-	?
Kouhia	2010	○	○	–	–	?	+	?	+	+

Ukai et al. BMC Gastroenterology (2016) 16:37

Page 3 of 12

Table 1 Outcomes and assessment of risk of bias in the included studies *(Continued)*

Kum	1993	○	○	○	○	+	+	?	-	?
Laine	1997	○	○	○	○	?	?	?	+	?
Macarulla	1995	○	○	○	○	?	+	?	+	?
Martin	1995	○	○	○	○	?	?	?	+	?
Minne	1997	○	○	–	–	+	?	?	-	?
Moberg	2005	○	○	–	–	+	+	+	+	+
Moirangthem	2008	○	–	○	○	?	?	?	+	?
Mutter	1996	○	○	○	–	?	?	?	+	?
Navarra	2000	○	○	○	○	+	+	?	+	?
Nordentoft	2000	–	–	○	–	?	?	?	-	?
Olmi	2005	○	○	–	–	?	?	?	?	?
Ortega	1996	○	○	○	○	+	+	+	+	?
Özmen	1999	○	○	○	○	?	?	?	?	?
Pedersen	2001	○	○	–	–	+	+	?	?	+
Pozo	1996	○	○	○	○	?	?	?	+	?
Rashid	2013	○	–	○	○	+	?	?	+	?
Reiertsen	1997	○	○	○	○	?	?	-	-	?
Ricca	2007	○	○	○	–	+	+	+	+	?
Schippers	1997	○	○	○	–	?	+	?	+	?
Settmacher	1995	–	–	○	–	?	+	?	?	+
Sezeur	1997	○	○	○	○	?	+	?	-	?
Shirazi	2010	○	–	○	○	?	?	?	?	?
Stare	1998	○	○	○	–	?	+	-	+	?
Sun	1998	○	○	○	○	?	?	?	+	?
Tate	1993	○	○	○	○	+	+	?	+	?
Tzovaras	2010	○	–	–	–	+	+	?	+	?
Vallribera	2003	○	○	○	○	+	+	?	+	?
van Dalen	2003	–	–	○	–	?	+	?	+	?
Wei	2010	○	○	○	○	?	?	?	+	?
Williams	1996	○	○	○	○	?	?	?	-	?
Witten	1998	○	○	○	○	+	+	-	+	?
Yin	1996	○	–	–	○	+	+	+	+	?
Zhang	1998	○	○	○	○	+	?	?	?	?

+, low risk of bias; –, high risk of bias; ?, unclear risk of bias

through discussion. Participants who were converted intraoperatively from laparoscopic appendectomy to open appendectomy were included in the laparoscopic appendectomy arm on an intention-to-treat basis.

Statistical analysis

A meta-analysis was performed using Review Manager (RevMan) software, version 5.3.5, provided by the Cochrane Collaboration, Copenhagen, Denmark. Since a cumulative meta-analysis cannot be performed by RevMan, we used Comprehensive Meta Analysis software, version 3.3.070 (Biostat, Englewood NJ, USA). For the binary variables (i.e., the incidence of intra-abdominal abscess and wound infection), the statistical analyses were performed using the odds ratio (OR) of laparoscopic appendectomy to open appendectomy as the summary statistic. The OR point estimate was considered significant at the $p < 0.05$ level when the 95 % confidence interval (95 % CI) did not include the value 1. For the continuous variables (i.e., the operative time and the length of hospital stay), the statistical analyses were performed using the mean difference (MD) as the summary statistic, i.e., the time taken for an laparoscopic appendectomy subtracted by the time taken for an open appendectomy. The MD point estimate was considered

significant at the $p < 0.05$ level if the 95 % CI did not include the value 0.

We used both the fixed-effects model and the random-effects model according to the Mantel-Haenszel method [14] for the statistical analysis. The fixed-effects model assumes the homogeneity of the true treatment effect, whereas the random-effects model accepts between-study differences in the treatment effects. The confidence interval thus tends to be wider in the random-effects model when a certain level of treatment effect heterogeneity is observed. We performed both the fixed-effects model and random-effects model, and if their results were similar, the random-effects model was adopted.

We also tested for study homogeneity by calculating I^2. This value can be calculated as $I^2 = 100\ \% \times (Q - df)/Q$, where Q is Cochran's heterogeneity statistic and df is the degree of freedom [15]. An outcome with no events was considered a "zero cell" in the 2 × 2 table. Although correction is needed to pool the ORs of studies that include zero cells, this can influence the results and possibly introduce bias [16]. To conduct a bias-free meta-analysis, we used the Mantel-Haenszel model with a correction factor of 0.5 (0.5 was added to all cells in the 2 × 2 table when there was a zero cell).
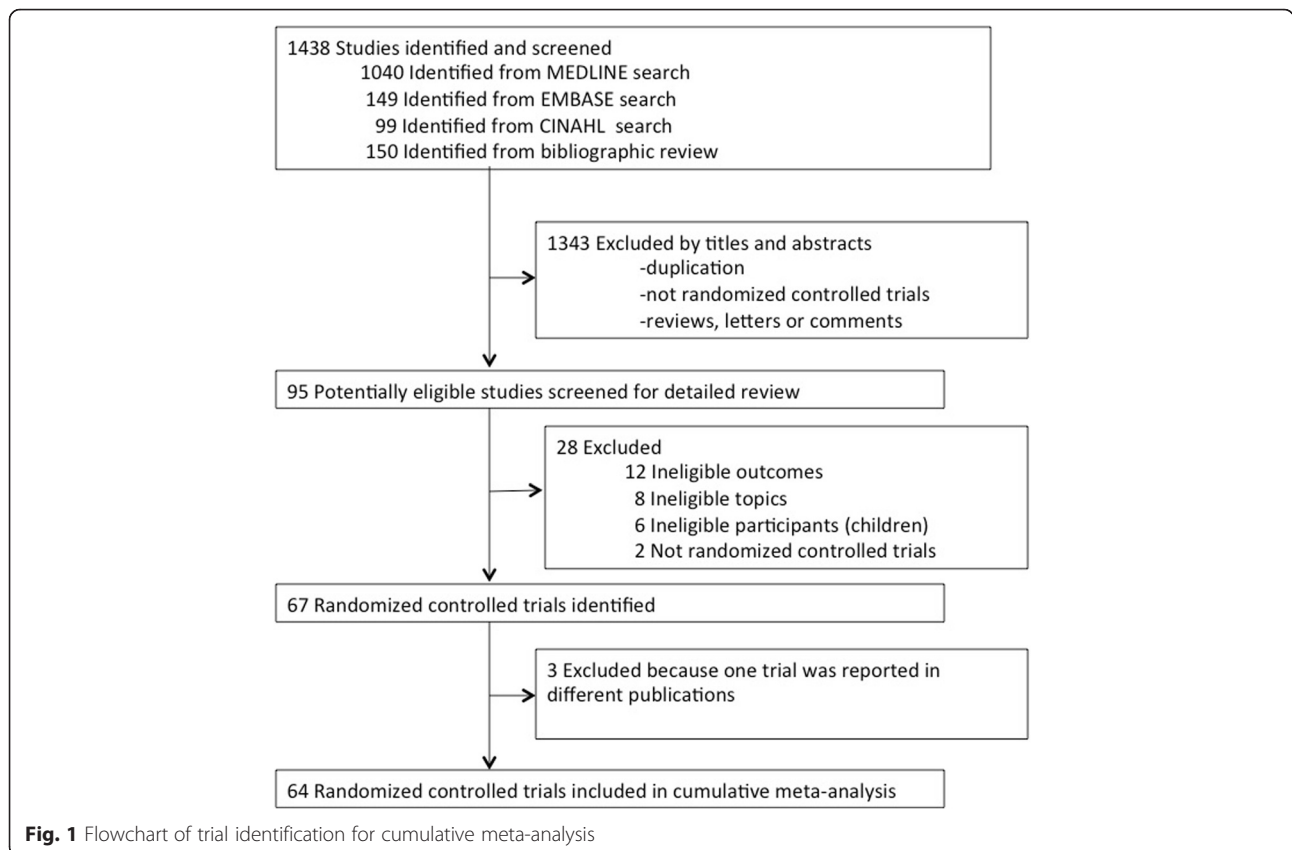

Fig. 1 Flowchart of trial identification for cumulative meta-analysis

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 6 of 12

Results

Our database and bibliography searches yielded 1,438 and 150 articles. After eliminating duplicate articles, we evaluated the titles and abstracts of these studies according to the inclusion and exclusion criteria, after which 95 articles remained (Fig. 1). After the full texts of these articles were read and the ineligible studies were excluded, 64 RCTs published from 1992 to 2012 were used for the data extraction (Table 1).

Intra-abdominal abscess

This outcome analysis included 51 relevant studies with a total of 6,512 participants (3,273 for laparoscopic appendectomy and 3,239 for open appendectomy) (Fig. 2). The total numbers of events were 61 in the laparoscopic appendectomy group (1.80 %) and 43 in the open

appendectomy group (1.30 %). The overall OR was 1.34 (95 % CI 0.92–1.94) in the fixed-effects model and 1.32 (95 % CI 0.84–2.10) in the random-effects model. The overall I^2 was 6 %. A visual inspection of the funnel plot for small-study effects did not show asymmetry (Fig. 3). A cumulative meta-analysis demonstrated that the CI narrowed until it identified the first significant difference in favor of open appendectomy in the trial published in 2001 (OR 2.35, 95 % CI 1.30–4.25). However, as more studies were added, the CI shifted to the left in favor of laparoscopic appendectomy. Finally, the CI included the value 1 in 2010, and there was no significant difference (Fig. 4).

Wound infection

Sixty studies and 7,462 participants (3,736 for laparoscopic appendectomy and 3,726 for open appendectomy)

Study or Subgroup	Laproscopic Appendectomy Events	Total	Open Appendectomy Events	Total	Weight	Odds Ratio M–H, Random, 95% CI	Year	Odds Ratio M–H, Random, 95% CI
Attwood 1992	0	30	0	32		Not estimable	1992	
Tate 1993	0	46	0	42		Not estimable	1993	
Kum 1993	0	52	0	57		Not estimable	1993	
Frazee 1994	1	38	0	37	2.0%	3.00 [0.12, 76.03]	1994	
Hebebrand 1994	0	34	0	23		Not estimable	1994	
Eichen 1994	0	97	1	96	2.0%	0.33 [0.01, 8.11]	1994	
Martin 1995	0	81	0	88		Not estimable	1995	
Macarulla 1995	1	106	0	104	2.0%	2.97 [0.12, 73.78]	1995	
Pozo 1996	1	20	0	24	1.9%	3.77 [0.15, 97.74]	1996	
Williams 1996	0	19	0	18		Not estimable	1996	
Mutter 1996	0	50	0	50		Not estimable	1996	
Ortega 1996	6	167	0	86	2.4%	6.96 [0.39, 125.07]	1996	
Henle 1996	1	87	2	82	3.4%	0.47 [0.04, 5.23]	1996	
Hansen 1996	0	86	0	72		Not estimable	1996	
Hart 1996	3	40	0	37	2.3%	7.00 [0.35, 140.26]	1996	
Cox 1996	0	33	0	31		Not estimable	1996	
Reiertsen 1997	0	42	0	42		Not estimable	1997	
Schippers 1997	2	20	0	20	2.1%	5.54 [0.25, 123.08]	1997	
Sezeur 1997	1	15	0	17	1.9%	3.62 [0.14, 95.78]	1997	
Minne 1997	2	27	1	23	3.3%	1.76 [0.15, 20.76]	1997	
Kazemier 1997	0	97	1	104	2.0%	0.35 [0.01, 8.79]	1997	
Laine 1997	0	25	0	25		Not estimable	1997	
Stare 1998	0	48	0	26		Not estimable	1998	
Sun 1998	0	50	0	100		Not estimable	1998	
Witten 1998	0	100	2	100	2.2%	0.20 [0.01, 4.14]	1998	
Zhang 1998	0	50	0	53		Not estimable	1998	
Heikkinen 1998	0	19	0	21		Not estimable	1998	
Ozmen 1999	0	35	0	35		Not estimable	1999	
Kald 1999	1	49	0	50	2.0%	3.12 [0.12, 78.55]	1999	
Hellberg 1999	5	244	3	256	8.8%	1.76 [0.42, 7.46]	1999	
Bauwens 1999	0	26	0	28		Not estimable	1999	
Navarra 2000	0	75	0	73		Not estimable	2000	
Hall Long 2000	4	93	1	105	4.1%	4.67 [0.51, 42.59]	2000	
Bruwer 2000	1	18	0	16	1.9%	2.83 [0.11, 74.46]	2000	
Pedersen 2001	13	282	3	301	11.0%	4.80 [1.35, 17.03]	2001	
Huang 2001	0	55	0	30		Not estimable	2001	
Helmy 2001	0	50	0	50		Not estimable	2001	
Al–Mulhim 2002	0	30	0	30		Not estimable	2002	
Vallribera 2003	0	35	1	30	2.0%	0.28 [0.01, 7.06]	2003	
Ignacio 2003	0	26	0	26		Not estimable	2003	
Karadayi 2003	0	30	0	30		Not estimable	2003	
Moberg 2005	0	81	1	82	2.0%	0.33 [0.01, 8.30]	2005	
Olmi 2005	2	150	3	138	5.9%	0.61 [0.10, 3.69]	2005	
Katkhouda 2005	6	113	4	134	10.6%	1.82 [0.50, 6.63]	2005	
Kaiser 2006	0	61	0	75		Not estimable	2006	
Ricca 2007	0	26	0	26		Not estimable	2007	
Kehagias 2009	2	46	1	48	3.4%	2.14 [0.19, 24.40]	2009	
Wei 2010	2	112	9	108	7.7%	0.20 [0.04, 0.95]	2010	
Kouhia 2010	2	47	1	52	3.4%	2.27 [0.20, 25.84]	2010	
Jan 2011	2	60	7	60	7.2%	0.26 [0.05, 1.31]	2011	
Kocatas 2013	1	50	1	46	2.6%	0.92 [0.06, 15.12]	2013	
Total (95% CI)		**3273**		**3239**	**100.0%**	**1.32 [0.84, 2.10]**		
Total events	59		42					
Heterogeneity: Tau² = 0.08; Chi² = 26.54, df = 25 (P = 0.38); I² = 6%								
Test for overall effect: Z = 1.20 (P = 0.23)								

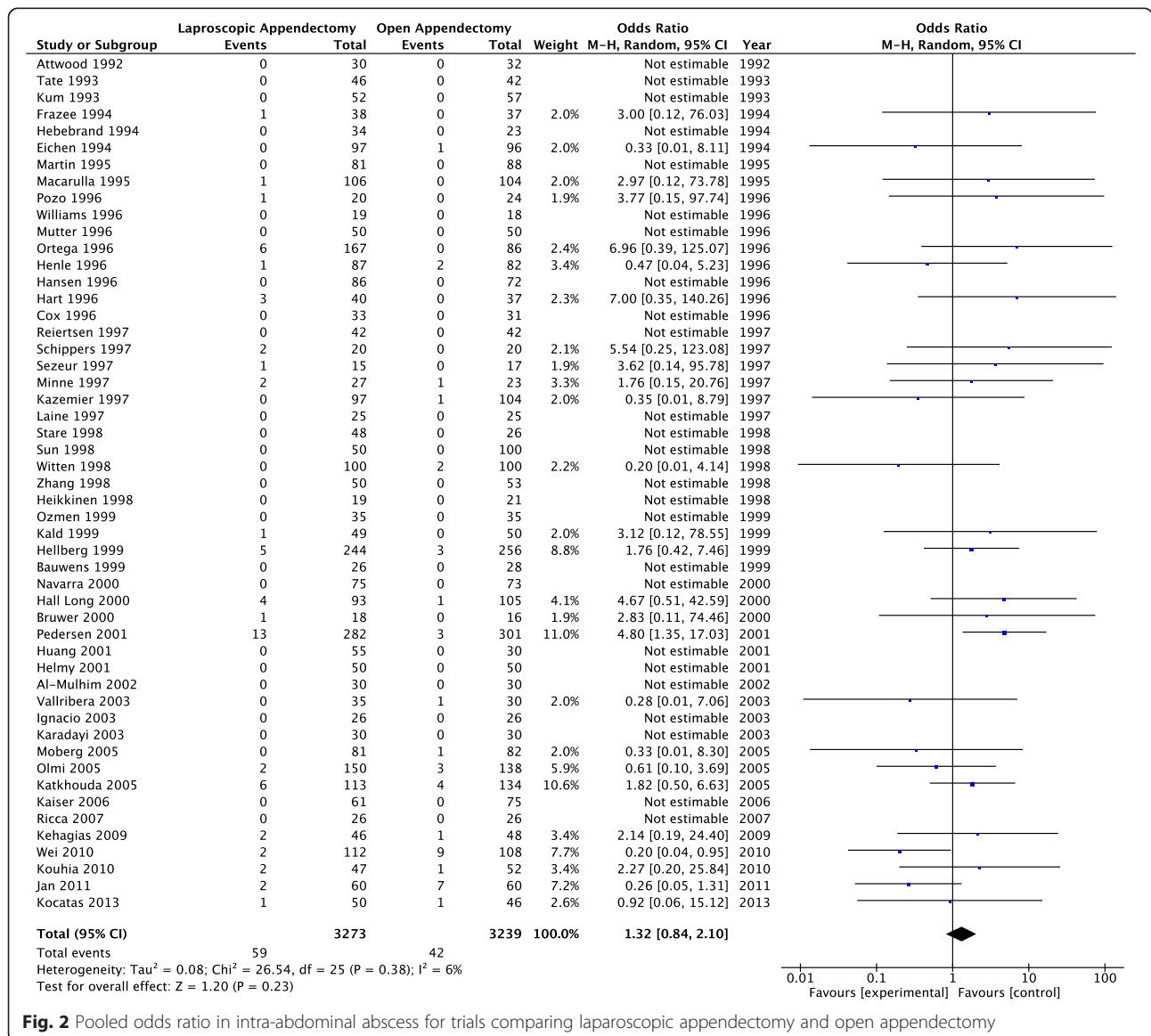0.01   0.1   1   10   100
Favours [experimental]   Favours [control]

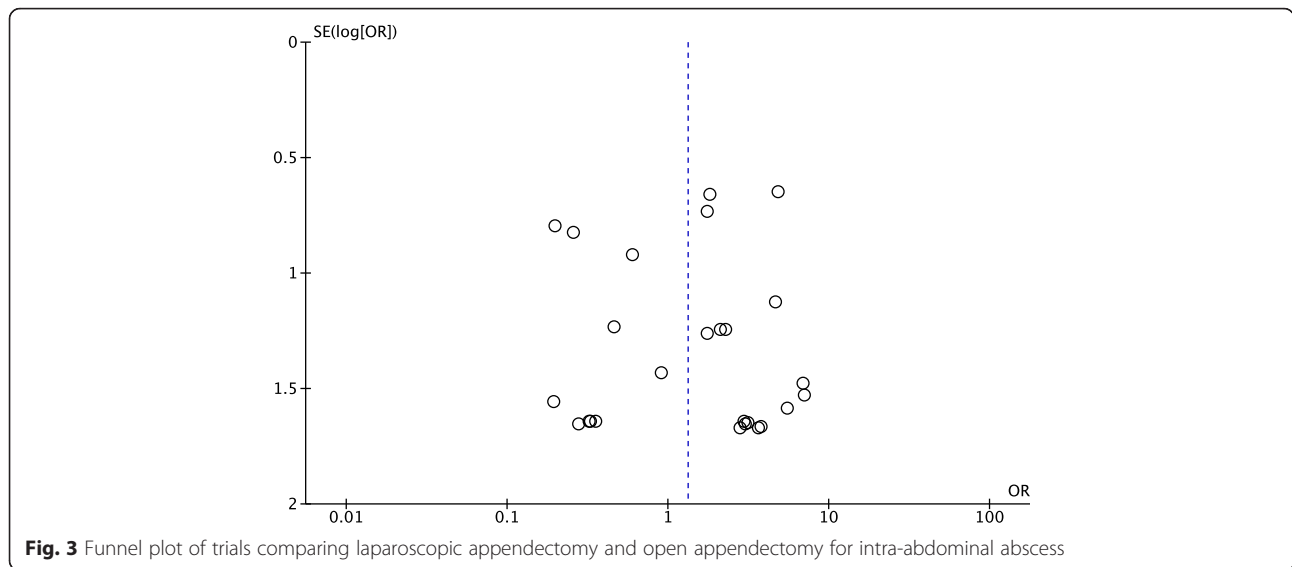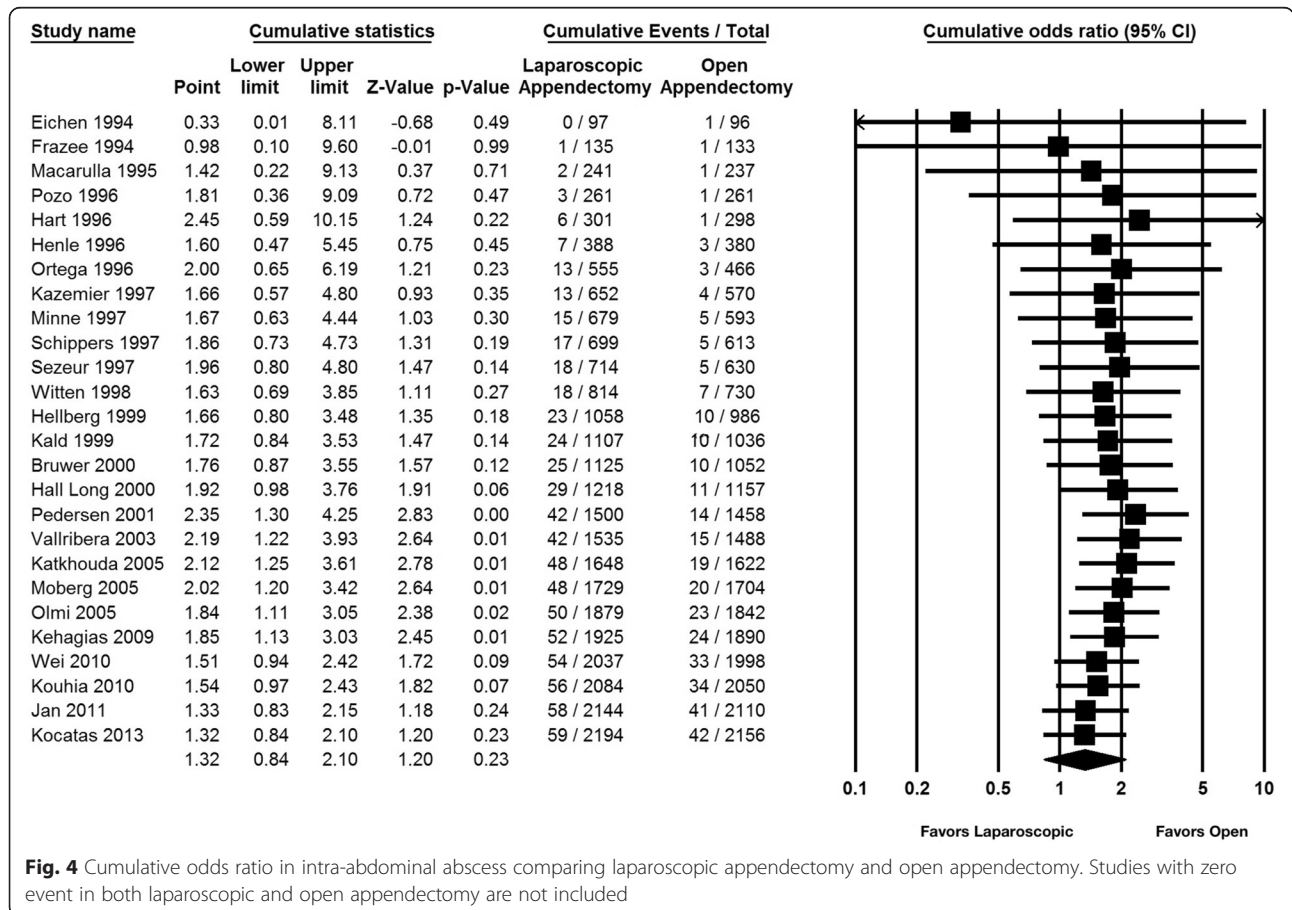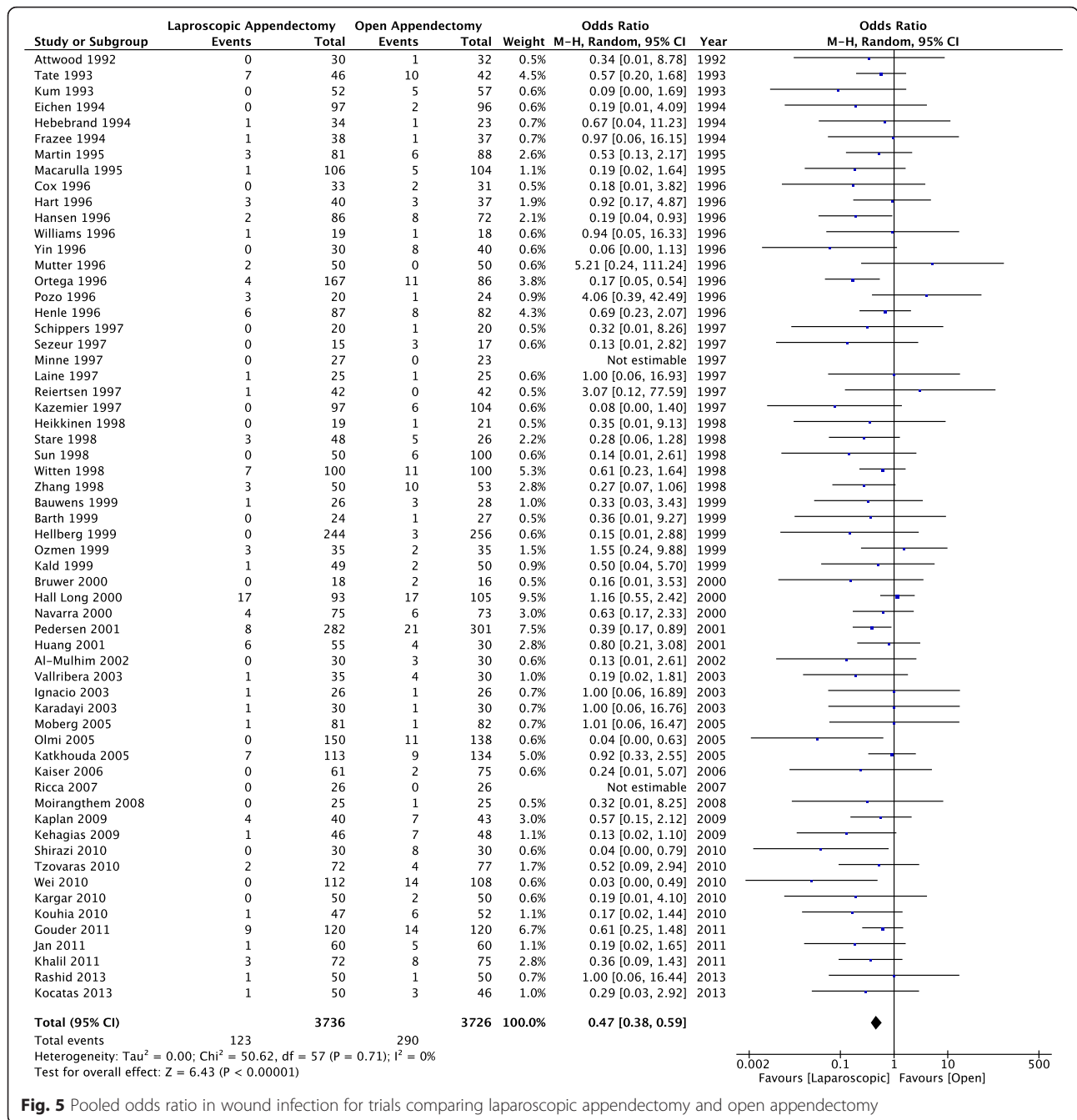**Fig. 2** Pooled odds ratio in intra-abdominal abscess for trials comparing laparoscopic appendectomy and open appendectomy

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 7 of 12

Fig. 3 Funnel plot of trials comparing laparoscopic appendectomy and open appendectomy for intra-abdominal abscess

were included for this outcome (Fig. 5). The total numbers of events were 123 in the laparoscopic appendectomy group (3.29 %) and 290 in the open appendectomy group (7.78 %). The overall OR was 0.41 (95 % CI 0.33–0.51) in the fixed-effects model and 0.47 (95 % CI 0.38–0.59) in the random-effects model. The overall I^2 was 0 %. A visual inspection of the funnel plot showed slight asymmetry in small studies (Fig. 6). A cumulative meta-analysis showed that the significant difference was first observed in the seventh study in 1995, and that this

Study name	Cumulative statistics					Cumulative Events / Total		Cumulative odds ratio (95% CI)
	Point	Lower limit	Upper limit	Z-Value	p-Value	Laparoscopic Appendectomy	Open Appendectomy	
Eichen 1994	0.33	0.01	8.11	-0.68	0.49	0 / 97	1 / 96	
Frazee 1994	0.98	0.10	9.60	-0.01	0.99	1 / 135	1 / 133	
Macarulla 1995	1.42	0.22	9.13	0.37	0.71	2 / 241	1 / 237	
Pozo 1996	1.81	0.36	9.09	0.72	0.47	3 / 261	1 / 261	
Hart 1996	2.45	0.59	10.15	1.24	0.22	6 / 301	1 / 298	
Henle 1996	1.60	0.47	5.45	0.75	0.45	7 / 388	3 / 380	
Ortega 1996	2.00	0.65	6.19	1.21	0.23	13 / 555	3 / 466	
Kazemier 1997	1.66	0.57	4.80	0.93	0.35	13 / 652	4 / 570	
Minne 1997	1.67	0.63	4.44	1.03	0.30	15 / 679	5 / 593	
Schippers 1997	1.86	0.73	4.73	1.31	0.19	17 / 699	5 / 613	
Sezeur 1997	1.96	0.80	4.80	1.47	0.14	18 / 714	5 / 630	
Witten 1998	1.63	0.69	3.85	1.11	0.27	18 / 814	7 / 730	
Hellberg 1999	1.66	0.80	3.48	1.35	0.18	23 / 1058	10 / 986	
Kald 1999	1.72	0.84	3.53	1.47	0.14	24 / 1107	10 / 1036	
Bruwer 2000	1.76	0.87	3.55	1.57	0.12	25 / 1125	10 / 1052	
Hall Long 2000	1.92	0.98	3.76	1.91	0.06	29 / 1218	11 / 1157	
Pedersen 2001	2.35	1.30	4.25	2.83	0.00	42 / 1500	14 / 1458	
Vallribera 2003	2.19	1.22	3.93	2.64	0.01	42 / 1535	15 / 1488	
Katkhouda 2005	2.12	1.25	3.61	2.78	0.01	48 / 1648	19 / 1622	
Moberg 2005	2.02	1.20	3.42	2.64	0.01	48 / 1729	20 / 1704	
Olmi 2005	1.84	1.11	3.05	2.38	0.02	50 / 1879	23 / 1842	
Kehagias 2009	1.85	1.13	3.03	2.45	0.01	52 / 1925	24 / 1890	
Wei 2010	1.51	0.94	2.42	1.72	0.09	54 / 2037	33 / 1998	
Kouhia 2010	1.54	0.97	2.43	1.82	0.07	56 / 2084	34 / 2050	
Jan 2011	1.33	0.83	2.15	1.18	0.24	58 / 2144	41 / 2110	
Kocatas 2013	1.32	0.84	2.10	1.20	0.23	59 / 2194	42 / 2156	
	1.32	0.84	2.10	1.20	0.23			

Favors Laparoscopic   Favors Open

Fig. 4 Cumulative odds ratio in intra-abdominal abscess comparing laparoscopic appendectomy and open appendectomy. Studies with zero event in both laparoscopic and open appendectomy are not included

Study or Subgroup	Laproscopic Appendectomy Events	Total	Open Appendectomy Events	Total	Weight	Odds Ratio M–H, Random, 95% CI	Year
Attwood 1992	0	30	1	32	0.5%	0.34 [0.01, 8.78]	1992
Tate 1993	7	46	10	42	4.5%	0.57 [0.20, 1.68]	1993
Kum 1993	0	52	5	57	0.6%	0.09 [0.00, 1.69]	1993
Eichen 1994	0	97	2	96	0.6%	0.19 [0.01, 4.09]	1994
Hebebrand 1994	1	34	1	23	0.7%	0.67 [0.04, 11.23]	1994
Frazee 1994	1	38	1	37	0.7%	0.97 [0.06, 16.15]	1994
Martin 1995	3	81	6	88	2.6%	0.53 [0.13, 2.17]	1995
Macarulla 1995	1	106	5	104	1.1%	0.19 [0.02, 1.64]	1995
Cox 1996	0	33	2	31	0.5%	0.18 [0.01, 3.82]	1996
Hart 1996	3	40	3	37	1.9%	0.92 [0.17, 4.87]	1996
Hansen 1996	2	86	8	72	2.1%	0.19 [0.04, 0.93]	1996
Williams 1996	1	19	1	18	0.6%	0.94 [0.05, 16.33]	1996
Yin 1996	0	30	8	40	0.6%	0.06 [0.00, 1.13]	1996
Mutter 1996	2	50	0	50	0.6%	5.21 [0.24, 111.24]	1996
Ortega 1996	4	167	11	86	3.8%	0.17 [0.05, 0.54]	1996
Pozo 1996	3	20	1	24	0.9%	4.06 [0.39, 42.49]	1996
Henle 1996	6	87	8	82	4.3%	0.69 [0.23, 2.07]	1996
Schippers 1997	0	20	1	20	0.5%	0.32 [0.01, 8.26]	1997
Sezeur 1997	0	15	3	17	0.6%	0.13 [0.01, 2.82]	1997
Minne 1997	0	27	0	23		Not estimable	1997
Laine 1997	1	25	1	25	0.6%	1.00 [0.06, 16.93]	1997
Reiertsen 1997	1	42	0	42	0.5%	3.07 [0.12, 77.59]	1997
Kazemier 1997	0	97	6	104	0.6%	0.08 [0.00, 1.40]	1997
Heikkinen 1998	0	19	1	21	0.5%	0.35 [0.01, 9.13]	1998
Stare 1998	3	48	5	26	2.2%	0.28 [0.06, 1.28]	1998
Sun 1998	0	50	6	100	0.6%	0.14 [0.01, 2.61]	1998
Witten 1998	7	100	11	100	5.3%	0.61 [0.23, 1.64]	1998
Zhang 1998	3	50	10	53	2.8%	0.27 [0.07, 1.06]	1998
Bauwens 1999	1	26	3	28	1.0%	0.33 [0.03, 3.43]	1999
Barth 1999	0	24	1	27	0.5%	0.36 [0.01, 9.27]	1999
Hellberg 1999	0	244	3	256	0.6%	0.15 [0.01, 2.88]	1999
Ozmen 1999	3	35	2	35	1.5%	1.55 [0.24, 9.88]	1999
Kald 1999	1	49	2	50	0.9%	0.50 [0.04, 5.70]	1999
Bruwer 2000	0	18	2	16	0.5%	0.16 [0.01, 3.53]	2000
Hall Long 2000	17	93	17	105	9.5%	1.16 [0.55, 2.42]	2000
Navarra 2000	4	75	6	73	3.0%	0.63 [0.17, 2.33]	2000
Pedersen 2001	8	282	21	301	7.5%	0.39 [0.17, 0.89]	2001
Huang 2001	6	55	4	30	2.8%	0.80 [0.21, 3.08]	2001
Al–Mulhim 2002	0	30	3	30	0.6%	0.13 [0.01, 2.61]	2002
Vallribera 2003	1	35	4	30	1.0%	0.19 [0.02, 1.81]	2003
Ignacio 2003	1	26	1	26	0.7%	1.00 [0.06, 16.89]	2003
Karadayi 2003	1	30	1	30	0.7%	1.00 [0.06, 16.76]	2003
Moberg 2005	1	81	1	82	0.7%	1.01 [0.06, 16.47]	2005
Olmi 2005	0	150	11	138	0.6%	0.04 [0.00, 0.63]	2005
Katkhouda 2005	7	113	9	134	5.0%	0.92 [0.33, 2.55]	2005
Kaiser 2006	0	61	2	75	0.6%	0.24 [0.01, 5.07]	2006
Ricca 2007	0	26	0	26		Not estimable	2007
Moirangthem 2008	0	25	1	25	0.5%	0.32 [0.01, 8.25]	2008
Kaplan 2009	4	40	7	43	3.0%	0.57 [0.15, 2.12]	2009
Kehagias 2009	1	46	7	48	1.1%	0.13 [0.02, 1.10]	2009
Shirazi 2010	0	30	8	30	0.6%	0.04 [0.00, 0.79]	2010
Tzovaras 2010	2	72	4	77	1.7%	0.52 [0.09, 2.94]	2010
Wei 2010	0	112	14	108	0.6%	0.03 [0.00, 0.49]	2010
Kargar 2010	0	50	2	50	0.6%	0.19 [0.01, 4.10]	2010
Kouhia 2010	1	47	6	52	1.1%	0.17 [0.02, 1.44]	2010
Gouder 2011	9	120	14	120	6.7%	0.61 [0.25, 1.48]	2011
Jan 2011	1	60	5	60	1.1%	0.19 [0.02, 1.65]	2011
Khalil 2011	3	72	8	75	2.8%	0.36 [0.09, 1.43]	2011
Rashid 2013	1	50	1	50	0.7%	1.00 [0.06, 16.44]	2013
Kocatas 2013	1	50	3	46	1.0%	0.29 [0.03, 2.92]	2013
Total (95% CI)		**3736**		**3726**	**100.0%**	**0.47 [0.38, 0.59]**	
Total events	123		290				

Heterogeneity: Tau² = 0.00; Chi² = 50.62, df = 57 (P = 0.71); I² = 0%
Test for overall effect: Z = 6.43 (P < 0.00001)

Fig. 5 Pooled odds ratio in wound infection for trials comparing laparoscopic appendectomy and open appendectomy

trend did not change substantially with subsequent studies (Fig. 7).

Operative time

There were 43 studies with a total of 4,202 participants (2,135 for laparoscopic appendectomy and 2,067 for open appendectomy) that compared the operative time between laparoscopic appendectomy and open appendectomy (Fig. 4). The average operative times were 57.3 min for laparoscopic appendectomy and 47.0 min for open appendectomy. The overall MD was 4.4 min

longer for laparoscopic appendectomy (95 % CI 3.5–5.3) in the fixed-effects model and 10.1 min (95 % CI 5.9–14.3) in the random-effects model. The I^2 value was 95 %. Significant heterogeneity was found, and thus a cumulative meta-analysis was not performed.

Length of hospital stay

Thirty-nine studies with a total of 4,240 participants (2,165 for laparoscopic appendectomy and 2,153 for open appendectomy) were included for this outcome (Fig. 5). The average length of hospital stay was 3.21 days

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 9 of 12

Fig. 6 Funnel plot of trials comparing laparoscopic appendectomy and open appendectomy for wound infection. OR, odds ratio; SE, standard error
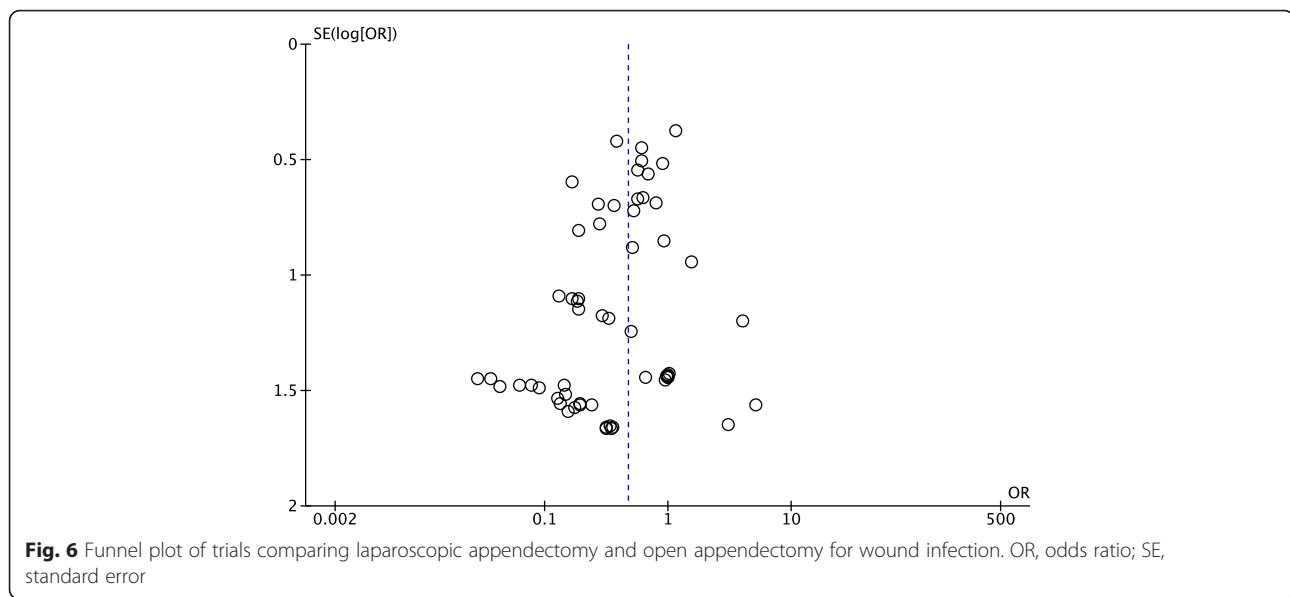
for laparoscopic appendectomy and 4.40 days for open appendectomy. The MD of the length of hospital stay was 1.08 days shorter for laparoscopic appendectomy (95 % CI 1.01–1.75) in the fixed-effects model and 1.12 days (95 % CI 0.77–1.47) for the random-effects model. The I^2 value was 95 %. Because significant heterogeneity was found, a cumulative meta-analysis was not performed.

Discussion

Our cumulative meta-analysis of the comparison of laparoscopic appendectomy and open appendectomy for acute appendicitis demonstrated that the evidence provided by the meta-analysis of surgical RCTs can change over time. Intra-abdominal abscesses were significantly more frequent in the laparoscopic appendectomy group during the period from 2001 to 2009, but this significance disappeared as more trials accumulated. Our present findings visually demonstrated how evidence changes over time in the surgical field. Although other outcome measures did not exhibit the same transition as intra-abdominal abscess, all of the outcome measures demonstrated similar trends in favor of laparoscopic appendectomy.

Fluctuation of evidence

When there is evidence concerning the effectiveness of a medical intervention, one can reasonably conclude that no further research is needed on the topic. However, previous studies have shown that the results of meta-analyses are underused, and many RCTs are conducted even after significant evidence has been demonstrated through a meta-analysis [2, 17]. Some researchers have contended that it is unethical and a waste of resources

to randomize participants in unnecessary trials, and they emphasized the importance of avoiding redundant trials. In the meta-analysis from a Cochrane review published in 2010 [10], intra-abdominal abscess was significantly more frequent in laparoscopic appendectomy than open appendectomy (albeit with moderate heterogeneity), but the present study illustrates that a significant result turned insignificant. The findings of our study thus provide an example in which large intervention effects are not always conclusive, and they fluctuate over time. Fluctuation was not found in the wound infection outcome data, and the result was consistently in favor of laparoscopic appendectomy. Penninga et al. have shown strong evidence of favoring laparoscopic appendectomy for wound infection using a trial sequential analysis [18], and our result is consistent with this.

The nature of surgical trials

Evidential instability can be explained by the nature of surgical interventions, which are highly complex and difficult to evaluate [6]. Surgical interventions involve many factors, including the surgeons' skill and judgment, the skills of the treating team, the development of surgical devices, and pre- and post-surgical management. All of these factors change on a daily basis. Second, the effect of the learning curve influences outcomes [19]. For example, surgeons' performances improve to the point of acquiring expertise as they gain training and experience. The observed shift toward favoring laparoscopic appendectomy, which we observed in later trials, might be explained by the effects of these factors. Because of this phenomenon, surgical trials may differ from pharmaceutical trials, as in the latter, theoretically efficacy does not change over time. To minimize these factors, trial
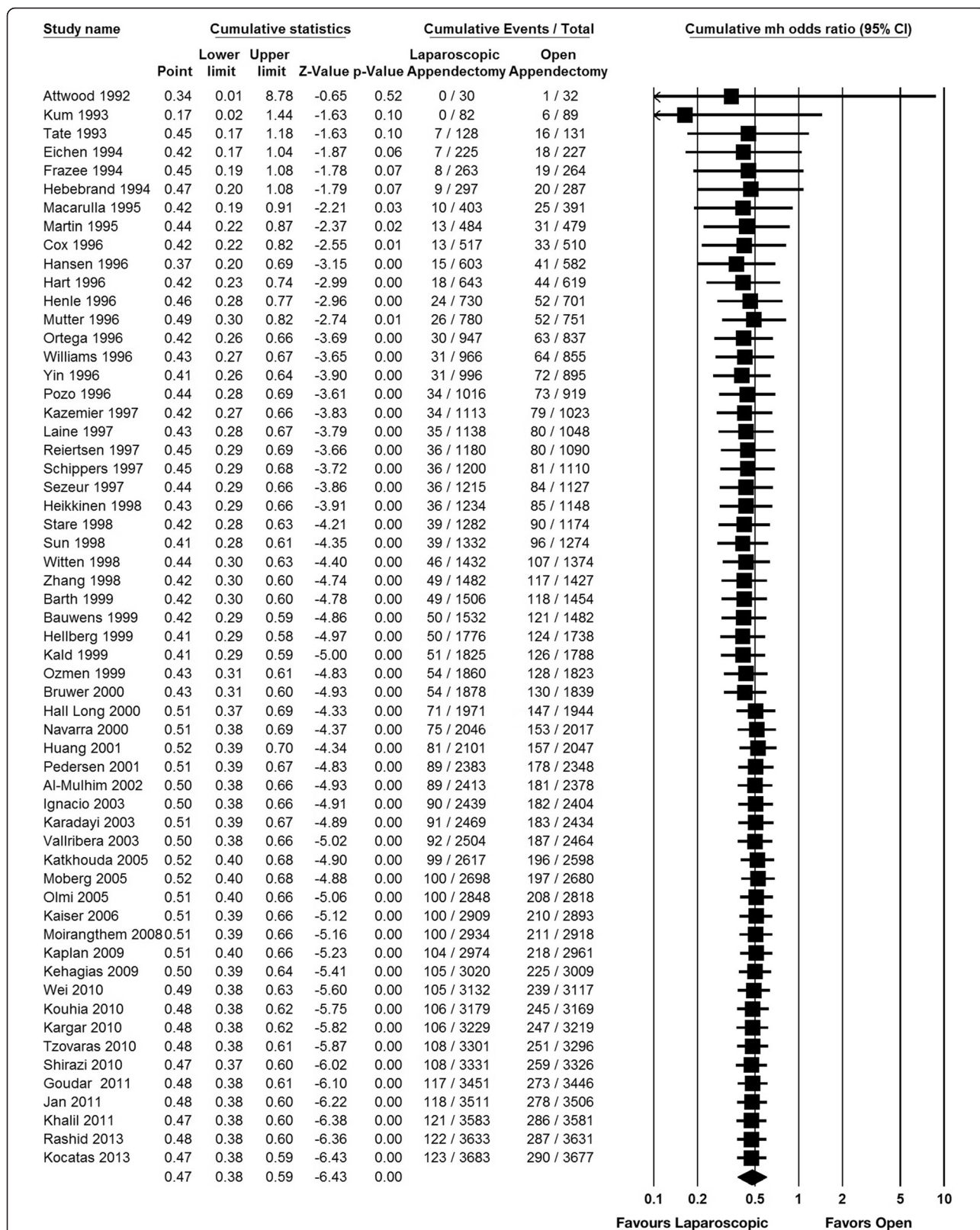
Study name	Cumulative statistics					Cumulative Events / Total		Cumulative mh odds ratio (95% CI)
	Point	Lower limit	Upper limit	Z-Value	p-Value	Laparoscopic Appendectomy	Open Appendectomy	
Attwood 1992	0.34	0.01	8.78	-0.65	0.52	0 / 30	1 / 32	
Kum 1993	0.17	0.02	1.44	-1.63	0.10	0 / 82	6 / 89	
Tate 1993	0.45	0.17	1.18	-1.63	0.10	7 / 128	16 / 131	
Eichen 1994	0.42	0.17	1.04	-1.87	0.06	7 / 225	18 / 227	
Frazee 1994	0.45	0.19	1.08	-1.78	0.07	8 / 263	19 / 264	
Hebebrand 1994	0.47	0.20	1.08	-1.79	0.07	9 / 297	20 / 287	
Macarulla 1995	0.42	0.19	0.91	-2.21	0.03	10 / 403	25 / 391	
Martin 1995	0.44	0.22	0.87	-2.37	0.02	13 / 484	31 / 479	
Cox 1996	0.42	0.22	0.82	-2.55	0.01	13 / 517	33 / 510	
Hansen 1996	0.37	0.20	0.69	-3.15	0.00	15 / 603	41 / 582	
Hart 1996	0.42	0.23	0.74	-2.99	0.00	18 / 643	44 / 619	
Henle 1996	0.46	0.28	0.77	-2.96	0.00	24 / 730	52 / 701	
Mutter 1996	0.49	0.30	0.82	-2.74	0.01	26 / 780	52 / 751	
Ortega 1996	0.42	0.26	0.66	-3.69	0.00	30 / 947	63 / 837	
Williams 1996	0.43	0.27	0.67	-3.65	0.00	31 / 966	64 / 855	
Yin 1996	0.41	0.26	0.64	-3.90	0.00	31 / 996	72 / 895	
Pozo 1996	0.44	0.28	0.69	-3.61	0.00	34 / 1016	73 / 919	
Kazemier 1997	0.42	0.27	0.66	-3.83	0.00	34 / 1113	79 / 1023	
Laine 1997	0.43	0.28	0.67	-3.79	0.00	35 / 1138	80 / 1048	
Reiertsen 1997	0.45	0.29	0.69	-3.66	0.00	36 / 1180	80 / 1090	
Schippers 1997	0.45	0.29	0.68	-3.72	0.00	36 / 1200	81 / 1110	
Sezeur 1997	0.44	0.29	0.66	-3.86	0.00	36 / 1215	84 / 1127	
Heikkinen 1998	0.43	0.29	0.66	-3.91	0.00	36 / 1234	85 / 1148	
Stare 1998	0.42	0.28	0.63	-4.21	0.00	39 / 1282	90 / 1174	
Sun 1998	0.41	0.28	0.61	-4.35	0.00	39 / 1332	96 / 1274	
Witten 1998	0.44	0.30	0.63	-4.40	0.00	46 / 1432	107 / 1374	
Zhang 1998	0.42	0.30	0.60	-4.74	0.00	49 / 1482	117 / 1427	
Barth 1999	0.42	0.30	0.60	-4.78	0.00	49 / 1506	118 / 1454	
Bauwens 1999	0.42	0.29	0.59	-4.86	0.00	50 / 1532	121 / 1482	
Hellberg 1999	0.41	0.29	0.58	-4.97	0.00	50 / 1776	124 / 1738	
Kald 1999	0.41	0.29	0.59	-5.00	0.00	51 / 1825	126 / 1788	
Ozmen 1999	0.43	0.31	0.61	-4.83	0.00	54 / 1860	128 / 1823	
Bruwer 2000	0.43	0.31	0.60	-4.93	0.00	54 / 1878	130 / 1839	
Hall Long 2000	0.51	0.37	0.69	-4.33	0.00	71 / 1971	147 / 1944	
Navarra 2000	0.51	0.38	0.69	-4.37	0.00	75 / 2046	153 / 2017	
Huang 2001	0.52	0.39	0.70	-4.34	0.00	81 / 2101	157 / 2047	
Pedersen 2001	0.51	0.39	0.67	-4.83	0.00	89 / 2383	178 / 2348	
Al-Mulhim 2002	0.50	0.38	0.66	-4.93	0.00	89 / 2413	181 / 2378	
Ignacio 2003	0.50	0.38	0.66	-4.91	0.00	90 / 2439	182 / 2404	
Karadayi 2003	0.51	0.39	0.67	-4.89	0.00	91 / 2469	183 / 2434	
Vallribera 2003	0.50	0.38	0.66	-5.02	0.00	92 / 2504	187 / 2464	
Katkhouda 2005	0.52	0.40	0.68	-4.90	0.00	99 / 2617	196 / 2598	
Moberg 2005	0.52	0.40	0.68	-4.88	0.00	100 / 2698	197 / 2680	
Olmi 2005	0.51	0.40	0.66	-5.06	0.00	100 / 2848	208 / 2818	
Kaiser 2006	0.51	0.39	0.66	-5.12	0.00	100 / 2909	210 / 2893	
Moirangthem 2008	0.51	0.39	0.66	-5.16	0.00	100 / 2934	211 / 2918	
Kaplan 2009	0.51	0.40	0.66	-5.23	0.00	104 / 2974	218 / 2961	
Kehagias 2009	0.50	0.39	0.64	-5.41	0.00	105 / 3020	225 / 3009	
Wei 2010	0.49	0.38	0.63	-5.60	0.00	105 / 3132	239 / 3117	
Kouhia 2010	0.48	0.38	0.62	-5.75	0.00	106 / 3179	245 / 3169	
Kargar 2010	0.48	0.38	0.62	-5.82	0.00	106 / 3229	247 / 3219	
Tzovaras 2010	0.48	0.38	0.61	-5.87	0.00	108 / 3301	251 / 3296	
Shirazi 2010	0.47	0.37	0.60	-6.02	0.00	108 / 3331	259 / 3326	
Goudar 2011	0.48	0.38	0.61	-6.10	0.00	117 / 3451	273 / 3446	
Jan 2011	0.48	0.38	0.60	-6.22	0.00	118 / 3511	278 / 3506	
Khalil 2011	0.47	0.38	0.60	-6.38	0.00	121 / 3583	286 / 3581	
Rashid 2013	0.48	0.38	0.60	-6.36	0.00	122 / 3633	287 / 3631	
Kocatas 2013	0.47	0.38	0.59	-6.43	0.00	123 / 3683	290 / 3677	
	0.47	0.38	0.59	-6.43	0.00			

0.1 0.2 0.5 1 2 5 10

Favours Laparoscopic   Favours Open

Fig. 7 Cumulative odds ratio in wound infection comparing laparoscopic appendectomy and open appendectomy. Studies with zero event in both laparoscopic and open appendectomy are not included

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 11 of 12

designs that consider the effects of the learning curve or perioperative management should be used [20, 21].

The shift favoring conservative treatment in the early to middle period

Although we observed a shift favoring laparoscopic appendectomy in later trials in our analysis, an apparent shift in the opposite direction was observed in the early to middle period after good results were obtained for laparoscopic appendectomy in very early trials. Early trials tend to overestimate treatment effects for a variety of reasons, such as the under-reporting of disappointing results or the selection of favorable subgroups [22–24]. However, as new interventions are disseminated and the study participant inclusion criteria are broadened, positive results become less extreme. Relevant examples can be found elsewhere [25, 26]. We assume that the results favoring open appendectomy in the early-middle period are another example of this phenomenon.

Limitations

This study has several limitations. First, we observed the change from statistically significant to insignificant findings using the 95 % CI of odds ratios. There are other methods to analyze the results of meta-analyses chronologically, such as the trial sequential analysis (TSA) which takes into account random errors due to repetitive meta-analyses [27, 28]. We performed a TSA for the intra-abdominal abscess outcome, and it did not show significance throughout; i.e., the required information size was not reached and the Z-curve did not cross the trial sequential monitoring boundaries. Therefore, we cannot dismiss the possibility of random errors which brought the statistical significance in the analysis. This strengthens the importance of not relying only on the 95 % CI of the effect size and of performing a TSA to conclude the comparison.

Second, this analysis can be prone to publication bias. It is likely that trials with contradicting results would be published more often than those confirming the existing evidence. Small study effects could also have existed in our analysis. A visual inspection of the funnel plot showed slight asymmetry in small studies (Fig. 6).

Third, we did not conduct an analysis of the operative time or the length of hospital stay due to considerable heterogeneity, and a small study effect could have a substantial impact on the heterogeneity.

Fourth, studies with a high risk of bias could skew the results. We conducted subgroup analyses for trials with low and high risks of bias, and the results showed that the intra-abdominal abscess outcome after laparoscopic appendectomy was considerably more frequent among the studies with a low risk of bias compared to those with a high risk of bias, although a cumulative meta-analysis showed a similar trend toward favoring laparoscopic appendectomy [see the Additional file 1].

Fifth, the rarity of intra-abdominal abscesses may have complicated the analysis. The small number of events can increase the uncertainty. Since there are many trials with zero-events in intra-abdominal abscess, we substituted the correction factor of 0.5 to 0.01 as a sensitivity analysis to test for robustness [29]. The cumulative meta-analysis showed that statistical significance was first observed in the trial published in 2001 and it disappeared as more results accumulated, with the overall OR of 1.24 (95 % CI 0.84–1.81). The results were similar between both correction factors.

Sixth, the number of studies published each year was unbalanced. We included more trials during the time period from 1996 to 2001, which might have biased the results. Nevertheless, considering that the shift from surgeons favoring open appendectomy to those favoring laparoscopic appendectomy occurred after 2002, new findings might have been more evident if more trials had been published after 2001.

Finally, although we report herein an example demonstrating evidential instability in the surgical field, we cannot generalize this observation to all surgical interventions or other fields. Although the numbers of RCTs addressing other surgical topics are generally low [30], more evidence should be accumulated for other topics to understand the stability of evidence by means of not only cumulative meta-analyses, but also TSAs.

Conclusion

Our cumulative meta-analysis of RCTs comparing laparoscopic and open appendectomy demonstrated that evidence can fluctuate over time in surgery with complex variables. Observing only the 95 % confidence interval of the effect size from meta-analyses may not provide conclusive results. More stringent analyses should be used to assess the results of meta-analyses.

Additional file

Additional file 1: Pooled odds ratio in intra-abdominal abscess for trials comparing laparoscopic appendcetomy and open appendectomy among studies with low risk of bias. Cummulative odds ratio in intra-abdominal abscess comparing laparoscopic appendectomy and open appendectomy among studies with low risk of bias. Pooled odds ratio in intra-abdominal abscess for trials comparing laparoscopic appendcetomy and open appendectomy among studies with high risk of bias. Cummulative odds ratio in intra-abdominal abscess comparing laparoscopic appendectomy and open appendectomy among studies with high risk of bias. (ZIP 1902 kb)

Ukai *et al. BMC Gastroenterology* (2016) 16:37

Page 12 of 12

Author details

[1]Department of Community Medicine, Mie University School of Medicine, 2-174 Edobashi, Tsu, Mie 514-8507, Japan. [2]Department of Family Medicine, Mie University School of Medicine & Graduate School of Medicine, 2-174 Edobashi, Tsu, Mie 514-8507, Japan. [3]Department of Family Medicine, Mie Prefectural Ichishi Hospital, 616 Minamiieki, Hakunsan-cho, Tsu, Mie 515-3133, Japan. [4]Department of Surgery, Flinders Medical Centre, Bedford Park, SA 5042, Australia. [5]General Internal Medicine, Japanese Red Cross Nagoya Daini Hospital, 2-9 Myoken-cho, Showa-ku, Nagoya 466-8650Aichi, Japan. [6]Department of Health Informatics, Kyoto University School of Public Health, Konoe-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan.

References

1. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. BMJ. 2009;339:b2700. doi:10.1136/bmj.b2700.
2. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992;327(4):248–54.
3. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992;268: 240–8.
4. Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Lancet. 1986;1:397–402.
5. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet. 1988;2:349–360.
6. Ergina PL, Cook JA, Blazeby JM, Boutron I, Clavien P-A, Reeves BC, Seiler CM, Altman DG, Aronson JK, Barkun JS, et al. Challenges in evaluating surgical innovation. Lancet. 2009;374(9695):1097–104.
7. Shikata S, Nakayama T, Noguchi Y, Taji Y, Yamagishi H. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. Ann Surg. 2006;244(5):668–76.
8. Li X, Zhang J, Sang L, Zhang W, Chu Z, Li X, Liu Y. Laparoscopic versus conventional appendectomy— a meta-analysis of randomized controlled trials. BMC Gastroenterol. 2010;10:129.
9. Ohtani H, Tamamori Y, Arimoto Y, Nishiguchi Y, Maeda K, Hirakawa K. Meta-analysis of the results of randomized controlled trials that compared laparoscopic and open surgery for acute appendicitis. J Gastrointest Surg. 2012;16:1929–39.
10. Sauerland S, Jaschinski T, Neugebauer EA. Laparoscopic versus open surgery for suspected appendicitis. Cochrane Database Syst Rev. 2010;10:CD001546. doi:10.1002/14651858.CD001546.pub3.
11. Glanville JM, Lefebvre C, Miles JNV, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. J Med Libr Assoc. 2006;94:130–6.
12. Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc. 2006;94(1):41–7.
13. Wong SS, Wilczynski NL, Haynes RB. Optimal CINAHL search strategies for identifying therapy studies and review articles. J Nurs Scholarsh. 2006;38(2): 194–9.
14. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88.
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60.
16. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med. 2004;23:1351–75.
17. Habre C, Tramèr MR, Pöpping DM, Elia N. Ability of a meta-analysis to prevent redundant research: systematic review of studies on pain from propofol injection. BMJ. 2014;348:g5219. doi: http://dx.doi.org/10.1136/bmj.g5219.
18. Penninga L, Gluud C, Wetterslev J. Meta-analysis of randomised trials on laparoscopic versus open surgery for acute appendicitis: has firm evidence been reached? J Gastrointest Surg. 2014;18(7):1383–4.
19. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Assessment of the learning curve in health technologies. A systematic review. Health Technol Assess. 2001;5(12):1–79.
20. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J, Aronson JK, Barkun JS, Blazeby JM, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet. 2009;374:1105–12.
21. Harrysson IJ, Cook J, Sirimanna P, Feldman LS, Darzi A, Aggarwal R. Systematic review of learning curves for minimally invasive abdominal surgery: a review of the methodology of data collection, depiction of outcomes, and statistical analysis. Ann Surg. 2014;260(1):37–45.
22. Chan A-W, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, Krumholz HM, Ghersi D, van der Worp HB. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257–66.
23. Pereira TV, Ioannidis JPA. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. J Clin Epidemiol. 2011;64:1060–9.
24. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. JAMA. 2012;308(16):1676–84.
25. Klein JB, Jacobs RH, Reinecke MA. Cognitive-behavioral therapy for adolescent depression: a meta-analytic investigation of changes in effect-size estimates. J Am Acad Child Adolesc Psychiatry. 2007;46(11):1403–13.
26. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PLoS One. 2014;9:e102670.
27. Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E. The knowledge system underpinning healthcare is not fit for purpose and must change. BMJ. 2015;350:h2463.
28. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol. 2008;61(1):64–75.
29. Keus F, Wetterslev J, Gluud C, Gooszen HG, van Laarhoven CJ. Robustness assessments are needed to reduce bias in meta-analyses that include zero-event randomized trials. Am J Gastroenterol. 2009;104(3):546–51. doi:10.1038/ajg.2008.22.
30. Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, Balliol C, Altman DG, Barkun JS, Blazeby JM, Boutron IC, et al. Evaluation and stages of surgical innovations. Lancet. 2009;374(9695):1089–96.