Comparison of Effects in Randomized Controlled Trials With Observational Studies in Digestive Surgery

Satoru Shikata, MD,*† Takeo Nakayama, MD, PhD,‡ Yoshinori Noguchi, MD, MPH,§ Yoshinori Taji, MD,† and Hisakazu Yamagishi, MD, PhD*

Objectives: To compare the results of randomized controlled trials versus observational studies in meta-analyses of digestive surgical topics.

Summary Background Data: While randomized controlled trials have been recognized as providing the highest standard of evidence, claims have been made that observational studies may overestimate treatment benefits. This debate has recently been renewed, particularly with regard to pharmacotherapies.

Methods: The PubMed (1966 to April 2004), EMBASE (1986 to April 2004) and Cochrane databases (Issue 2, 2004) were searched to identify meta-analyses of randomized controlled trials in digestive surgery. Fifty-two outcomes of 18 topics were identified from 276 original articles (96 randomized trials, 180 observational studies) and included in meta-analyses. All available binary data and study characteristics were extracted and combined separately for randomized and observational studies. In each selected digestive surgical topic, summary odds ratios or relative risks from randomized controlled trials were compared with observational studies using an equivalent calculation method.

Results: Significant between-study heterogeneity was seen more often among observational studies (5 of 12 topics) than among randomized trials (1 of 9 topics). In 4 of the 16 primary outcomes compared (10 of 52 total outcomes), summary estimates of treatment effects showed significant discrepancies between the two designs.

Conclusions: One fourth of observational studies gave different results than randomized trials, and between-study heterogeneity was more common in observational studies in the field of digestive surgery.

From the *Department of Digestive Surgery, Kyoto Prefectural University of Medicine, Kyoto, Japan; †Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine, Kyoto, Japan; ‡Department of Health Informatics, Kyoto University School of Public Health, Kyoto, Japan; and §Department of Medicine, Fujita Health University School of Medicine, Aichi, Japan.

The first randomized controlled trial in medicine was an investigation of streptomycin in 1948.[1] Since then, randomized controlled trials have been widely recognized as offering the gold standard for evaluating treatment efficacy and effectiveness and are classified as providing the highest grade of evidence in the hierarchy of research designs.[2]

Evaluations in the 1970s and 1980s suggested that observational studies may spuriously overestimate treatment benefits, yielding misleading conclusions.[3–6] In recent years, this debate has resurfaced. Some reports have suggested that for selected medical topics, both randomized and observational studies, may yield very similar results.[7,8] Conversely, opposing results have been reported from a large number of diverse medical topics.[9] Although these previous studies have contained some surgical topics, most have assessed topics involving pharmacotherapies. However, pharmacologic and surgical therapies differ in clinical nature, and results for pharmacologic investigations may therefore not apply to surgical fields.

This issue warrants investigation with a focus on the surgical area, and no previous studies appear to have undertaken an exhaustive assessment of a single clinical field. The present study investigated digestive surgery, allowing a systematic search and evaluation.

This systematic and exhaustive search of a large number of diverse articles on digestive surgery seeks to answer the following question: Do observational studies in digestive surgery tend to produce the same results as randomized controlled trials?

METHODS

Search for Meta-Analyses of Randomized Controlled Trials and Selection of Topics

Meta-analyses of randomized controlled trials in digestive surgery that had been published up to April 2004 were selected as topics in this study. Retrieved articles were judged suitable for use as topics only if all the following criteria were met: 1) meta-analysis of randomized controlled trials; 2) investigating digestive surgery; 3) assessing the treatment effects of at least one operative intervention versus any other intervention (operative or nonoperative); and 4) subjects were human. Searches were not limited to English language articles (any language). Studies were excluded if the main purpose was not evaluation of treatment effect, such as diagnosis. A literature search was

performed using the PubMed (1966 to April 2004), EMBASE (1986 to April 2004) and Cochrane Library (Issue 2, 2004) databases. A computer-assisted search was conducted using the following combination of Medical Subject Heading Terms and text words: "surgical procedures, operative," "digestive system surgical procedures," "randomized," "random," "meta-analysis," and "review." A manual search was also performed using references from the retrieved review articles.

Search for Observational Studies for Meta-Analysis

If meta-analyses of both randomized and observational studies had been performed on the same topic in each selected review article, the results could be used for comparison. However, if meta-analysis of observational studies had not been performed, we attempted to perform that by ourselves. Thus, when a meta-analysis of observational studies could not be identified in the selected review article, we needed to search for such meta-analyses while gathering observational studies under the following process.

For meta-analyses of observational studies, we first searched observational studies for all selected topics. In each topic, the same inclusion criteria used for meta-analysis of randomized controlled trials were used, with the exception of study design. Observational study designs were used if they could be categorized as prospective nonrandomized studies, retrospective cohort studies, case-control studies, case series with control groups, or other unspecified designs (provided a control group was used). A literature search was performed using the PubMed (1966 to April 2004), EMBASE (1986 to April 2004) and Cochrane Library (Issue 2, 2004) databases. PubMed contains no search term for observational studies, so a text-word strategy was used to search for "observational," "nonrandomized," "case series," "case control study," "cohort," "retrospective," and "prospective." In addition, a manual search was performed using references from the retrieved review articles. We also attempted to contact as many experts from the review articles as possible.

Data Extraction and Selection of Outcomes

All available binary data were extracted from the outcomes of the gathered observational studies. Data extraction was performed after translation of the article into English if the article had not been written in English or Japanese. Up to this point, 2 authors (S.S., T.N.) undertook the literature searches and data extraction independently, and disagreements were resolved by consensus.

For final inclusion of a topic in the present evaluation, binary data for the same outcome had to be available from at least one randomized trial and at least one observational study. When primary outcomes had been defined in the review article, these were used for the main comparison. Whenever the primary outcome was unclear, the outcome that was considered a priori as the most clinically important was selected, using consensus among the data extractors. In digestive surgery, mortality was generally given priority in clinical importance over other outcomes.

Statistical Analysis

For all selected topics, data from observational studies were combined. Generally, the fixed-effects model weighted by Peto's odds ratio method or the Mantel-Haenszel method was used for data pooling, followed by a test of heterogeneity.[10,11] Heterogeneity between studies was assessed using Q statistics.[12] Given the low power of this test, a significance level of 0.10 was used, rather than 0.05.[13] If the hypothesis of heterogeneity was accepted, the random-effects model using the DerSimonian-Laird method was used.[14] However, this study sought to compare summary estimates of randomized controlled trials with observational studies under equivalent conditions to the maximum extent possible. Thus, when performing meta-analysis of observational studies, we used the same method that had been used in the meta-analysis of randomized controlled trials. In this study, the quantity I^2 was used for assessing heterogeneity between trials in meta-analyses, calculated as: $I^2 = [(Q - df)/Q] \times 100$, where Q is the χ^2 statistic and *df* is the degrees of freedom. A value greater than 50% may be considered indicative of substantial heterogeneity.[15]

Although pooled odds ratio or pooled relative risk could be used as the indicator of summary estimates of outcomes, the present study used the same indicator that had been used in the meta-analysis of randomized controlled trials. In this context, odds ratios and relative risks will inevitably be similar in magnitude, as the rates of outcome events are low. Relative risks were therefore considered as odds ratios in comparisons of summary estimates. Confidence intervals were always calculated at 95%. When one arm of an outcome contained no events, this was considered a "zero cell" in the 2×2 table. Zero cells create problems in computing ratio measures of treatment effect. This problem was dealt with using a common method of adding 0.5 to each cell of the 2×2 table for the trial.[16]

To evaluate concordance between the results of randomized and observational studies, the following analyses were performed: 1) assessment of the number of cases in which the summary estimates of the observational studies suggested an effect at least double that of the randomized trials; and 2) evaluation of whether differences in the summary odds estimates of randomized controlled trials and observational studies for the same topic were larger than what would be expected by chance alone. To accomplish this, Z scores were calculated as follows:

$$Z = [\ln(OR_{RCT}) - \ln(OR_{OBS})] / \{var[\ln(OR_{RCT})] + var[\ln(OR_{OBS})]\}^{1/2},$$

where $\ln(OR_{RCT})$ is the natural logarithm of the odds ratio or relative risk of randomized controlled trials, $\ln(OR_{OBS})$ is the natural logarithm of the odds ratio or relative risk of observational studies, and var is variance. A Z score above 1.96 or less than -1.96 suggests a nonrandom difference between randomized controlled trials and observational studies (0.05 level of statistical significance).[17]

All statistical analyses were performed using STATA statistical software version 8.1 (STATA Corporation, College Station, TX).

RESULTS

Characteristics of Topics, Observational Studies

A literature search was first performed to select meta-analyses of randomized controlled trials for the topics, identifying 1184 potentially relevant articles. The process finally identified and selected 15 meta-analyses of randomized controlled trials for digestive surgical topics in this research (Fig. 1).[7,18–31] Three of the 15 reviews contained two topics.[21,30,31] Thus, 18 topics were identified for comparison of summary estimates between randomized controlled trials and observational studies (Table 1).

Meta-analyses of observational studies could not be identified for 10 of the 18 topics (topics 2, 3, 8–13, 17, and 18), so additional meta-analyses were required. Meta-analyses of observational studies had been identified for the remaining 8 topics (topics 1, 4–7, and 14–16), and the results were used for comparisons in this study.

For meta-analyses of observational studies for the 10 topics without existing meta-analyses, a literature search was performed and 111 observational studies were selected from 10,960 articles using the process outlined in Figure 2. Of the 111 selected articles, 17 had not been written in English or Japanese, instead appearing in 7 different languages, and the 2 trial assessors therefore abstracted data from the articles after translation into English by independent translators. A total of 52 common outcomes for both randomized controlled and observational studies were available for comparison in this study.

Using the described processes, 52 outcomes of 18 topics were investigated in 276 original articles (96 randomized trials, 180 observational studies) with a total of 101,170 study patients (Table 1). The 180 observational studies comprised 36 prospective and 144 retrospective studies. Randomized and observational studies on the same topic generally administered treatment in the same way and outcome measures were similarly defined.

Between-Study Heterogeneity

Data on between-study heterogeneity using the I^2 statistic were available for all 10 meta-analyses of observational studies that we performed specifically for the present study (topics 2, 3, 8–13, 17, and 18). Conversely, data had not been described in 8 of the remaining meta-analyses that had been reported (topics 1, 4–7, and 14–16). In primary outcomes of 16 topics, significant heterogeneity was noted between randomized controlled trials in 1 of 9 topics (11.1%). Significant between-study heterogeneity was identified between observational studies in 5 of 12 topics (41.7%). There was no significant difference between the rates of heterogeneity ($P = 0.18$ by Fisher exact test).

Comparison of Primary Outcomes

In almost all topics, the primary outcome defined in the review or decided by author consensus was mortality. However, in topics dealing with safety of procedures, such as appendectomy and operation for fissure-in-ano, one of the complications, such as risk of wound infection or persistence
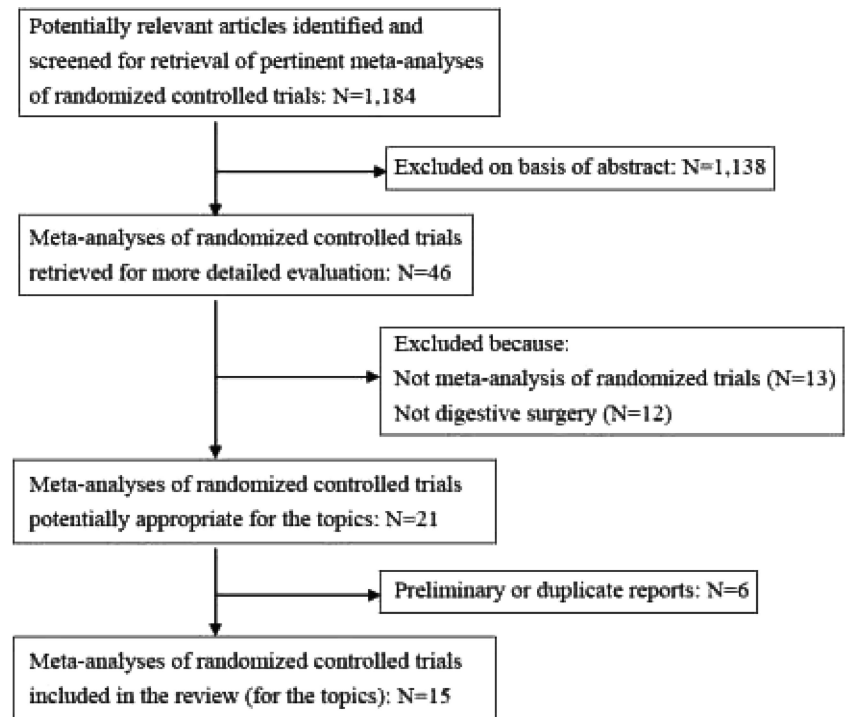

FIGURE 1. Summary profile of search for meta-analyses of randomized controlled trials.

TABLE 1. Topics of Meta-Analyses Considering Both Randomized Controlled Trials and Observational Studies

Identification No.	Topic	Meta-Analysis	Randomized Controlled Trial — No. of Studies (no. of patients)	Observational Study (no. of studies) — Prospective/ Retrospective	Observational Study (no. of studies) — Total (no. of patients)	No. of Comparable Outcomes
1	Closed postoperative peritoneal lavage vs. no lavage for generalized peritonitis	Leiboff et al[18] (1987)	4 (173)	2/6	8 (1034)	1
2	Splenorenal shunt vs. endoscopic sclerotherapy in prevention of variceal rebleeding	Spina et al[19] (1992)	4 (310)	0/2	2 (344)	1
3	Routine drainage vs. no drainage after elective colorectal surgery	Urbach et al[20] (1999)	4 (414)	0/5	5 (1767)	4
4	Anal stretch vs. sphincterotomy for fissure-in-ano	Nelson et al[21] (1999)	6 (328)	0/4	4 (537)	2
5	Open vs. closed lateral sphincterotomy for fissure-in-ano	Nelson et al[21] (1999)	2 (140)	0/4	4 (1365)	2
6	Laparoscopic vs. open appendectomy for acute appendicitis	Benson et al[7] (2000)	16 (1703)	3/4	7 (1502)	1
7	Transthoracic vs. transhiatal resection for carcinoma of the esophagus	Hulscher et al[22] (2001)	3 (138)	3/18	21 (2466)	6
8	Hand-sewn vs. stapled esophagogastric anastomosis after esophagectomy	Urschel et al[23] (2001)	5 (467)	2/8	10 (3196)	3
9	Posterior vs. anterior route of reconstruction after esophagectomy	Urschel et al[24] (2001)	6 (342)	0/3	3 (329)	4
10	Pyloroplasty vs. no drainage in gastric reconstruction after esophagectomy	Urschel et al[25] (2002)	3 (347)	0/2	2 (111)	1
11	Primary repair vs. fecal diversion for penetrating colon injuries	Singer et al[26] (2002)	5 (467)	4/29	33 (5745)	4
12	Stapled vs. hand-sewn methods for colorectal anastomosis surgery	Lustosa et al[27] (2002)	9 (1233)	2/13	15 (3894)	6
13	Stapled vs. conventional hemorrhoidectomy	Sutherland et al[28] (2002)	7 (591)	2/5	7 (910)	3
14	Extended vs. limited lymph node dissection for adenocarcinoma of the stomach	McCulloch et al[29] (2003)	3 (1729)	5/8	13 (4058)	2
15	Open (Hasson type) vs. closed (needle/trocar) access in laparoscopic surgery	Merlin et al[30] (2003)	4 (302)	4/6	10 (20,664)	3
16	Direct trocar vs. closed (needle/trocar) access in laparoscopic surgery	Merlin et al[30] (2003)	3 (665)	0/2	2 (1575)	3
17	Early vs. delayed open cholecystectomy for acute cholecystitis	Papi et al[31] (2004)	9 (916)	2/16	18 (37,475)	3
18	Early vs. delayed laparoscopic cholecystectomy for acute cholecystitis	Papi et al[31] (2004)	3 (228)	7/9	16 (3705)	3
Total			96 (10,493)		180 (90,677)	52

of fissure, was considered as a more appropriate primary outcome.

In 16 of 18 topics, primary outcomes could be compared between observational studies and randomized controlled trials. These summary estimates and associated 95% confidence intervals are shown in Figure 3. One of 16 primary outcomes displayed a magnitude of effect in the combined observational studies that was outside the 95% confidence interval for the
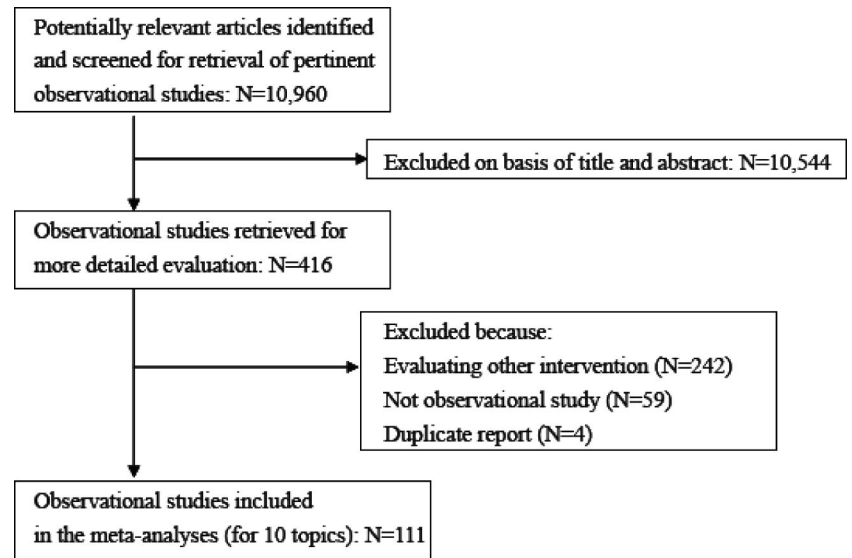
FIGURE 2. Summary profile of search for observational studies.

combined randomized controlled trials (topic 14). In 4 of 16 primary outcomes, summary estimates from observational studies were at least double those from randomized controlled trials (topics 7, 8, 15, and 17). The converse occurred in 3 topics (topics 11, 14, and 16) (exact $P = 0.45$ by Wilcoxon test). Evaluation by Z score revealed significant discrepancies between randomized trials and observational studies for 4 of 16 primary outcomes (topic 7, $Z = -4.28$; topic 8, $Z = -2.36$; topic 11, $Z = 2.19$; topic 14, $Z = 4.34$).

Comparison of All Outcomes

All summary estimates for 52 outcomes of 18 topics are shown in Table 2. Three types of calculation model were used: random effects calculation using the DerSimonian-Laird method; and fixed effects calculation using Peto's odds ratio method or the Mantel-Haenszel method. In 21 of 52 outcomes, relative risk was evaluated rather than odds ratio in meta-analyses of observational studies, as the original meta-analyses of randomized controlled trials had used relative risks for evaluations.

In 9 of 52 outcomes, summary estimates from observational studies were at least double those from randomized controlled trials. The converse occurred in 10 outcomes (exact $P = 0.943$ by Wilcoxon test). Evaluation by Z score revealed significant discrepancies between randomized trials and observational studies in 10 of 52 outcomes.

Overall, these data suggest that about one fourth of observational studies gave different results than randomized trials.

DISCUSSION

Using data from 276 articles in 18 topics, summary estimates were compared between randomized controlled trials and observational studies in digestive surgery. Significant between-study heterogeneity occurred more often between observational studies than between randomized controlled trials. One fourth of the summary estimates of treatment effects in randomized controlled trials and observational studies differed significantly from each other. From

this study, observational studies in digestive surgery tend to have similar results to those by randomized controlled trials. At least, they do not tend to overestimate or underestimate more than randomized controlled trials.

Our findings support the conclusions of earlier evaluations in the 1970s and 1980s.[3–6] In 2001, Ioannidis et al investigated 45 diverse pharmacologic and surgical topics in 408 articles and concluded that observational studies tend to indicate larger treatment effects (28 of 45 topics vs. 11 of 45 topics) and between-study heterogeneity is more frequent among observational studies than among randomized controlled trials (41% vs. 23%).[9] On the other hand, previous studies by Benson and Hartz[7] and Concato et al[8] reached the opposite conclusion. Benson and Hartz[7] investigated 19 diverse pharmacologic and surgical treatments in 136 articles and found little evidence of larger or differing estimates of treatment effects in observational studies compared with randomized controlled trials. Concato et al[8] evaluated 5 clinical topics and 99 articles, concluding that well-designed observational studies do not systematically overestimate the magnitude of treatment effects when compared with randomized controlled trials on the same topic.

All these previous studies have made substantial contributions toward identifying the problems caused by differing study designs. However, conclusions have inevitably been in the form of general statements, as the studies addressed diverse topics in various clinical fields. The present study was limited to a single clinical field, digestive surgery and thus offers two advantages over previous studies: a more exhaustive search is possible in studies of diverse clinical fields; and higher applicability to clinical practice than a general statement.

In 25% of digestive surgical topics, summary estimates of treatment effects in observational studies yielded different results than randomized trials, but both designs reached similar results in the remaining topics. This may be attributable to various factors. First, quality of surgical randomized controlled trials is low according to some review articles and
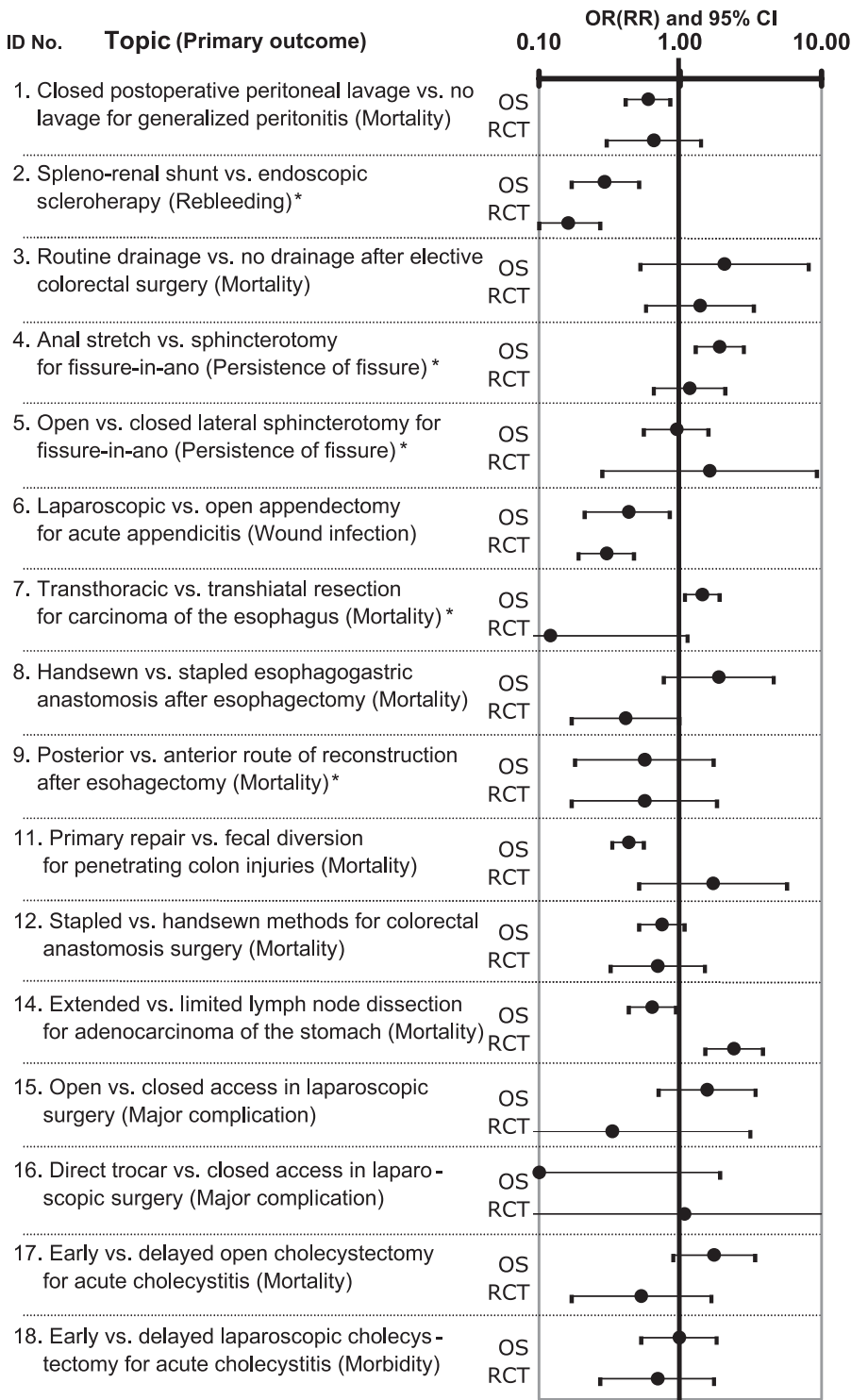
FIGURE 3. Comparison of primary outcomes between observational studies and randomized controlled trials. This figure is based on data from 13 review articles[7,18–24,26,27,29–31] and 10 meta-analyses of observational studies by the authors. OR, odds ratio; RR, relative risk; CI, confidence interval. *Outcome reporting relative risk rather than odds ratio.

may be so low that the essential contents of randomized trials do not differ from those of observational studies.[32,33] Second, for most topics, sample sizes may be too small to detect clinically important differences between the results of two types of study. Actually, 12 of 18 topics used fewer than 500 randomized patients. Combined with the use of a rare end-point, mortality, we could expect to see very large confidence intervals in the randomized evidence. The wide confidence intervals mean that demonstrating any significant discrepancy between the two designs will be very difficult.

This study examined not only primary outcomes, but also the secondary outcomes. Generally, results about

TABLE 2. Summary Estimates for All Outcomes

Identification No.	Topic	Outcome	Randomized Controlled Trial		Observational Study		Calculation Model
			No. of Studies	Summary Estimate OR (95% CI)	No. of Studies	Summary Estimate OR (95% CI)	
1	Closed postoperative peritoneal lavage vs. no lavage for generalized peritonitis	Mortality	4	0.65 (0.30–1.40)	8	0.59 (0.41–0.85)	M-H
2	Splenorenal shunt vs. endoscopic sclerotherapy in the prevention of variceal rebleeding	Rebleeding*	4	0.16 (0.10–0.27)	2	0.29 (0.17–0.51)	M-H
3	Routine drainage vs. no drainage after elective colorectal surgery	Mortality.	4	1.38 (0.57–3.31)	2	2.05 (0.52–8.11)	M-H
		Anastomotic leak	4	1.47 (0.71–3.06)	5	1.99 (1.12–3.53)	M-H
		Pulmonary complication	4	0.81 (0.41–1.59)	2	0.94 (0.24–3.73)	M-H
		Wound infection	4	1.70 (0.87–3.30)	2	0.94 (0.24–3.73)	M-H
4	Anal stretch vs. sphincterotomy for fissure-in-ano	Persistence of fissure*	6	1.16 (0.65–2.08)	4	1.89 (1.28–2.81)	M-H
		Flatus incontinence*	4	6.63 (2.06–21.3)	4	1.34 (0.79–2.27)	M-H
5	Open vs. closed lateral sphincterotomy for fissure-in-ano	Persistence of fissure*	2	1.61 (0.28–9.28)	4	0.94 (0.55–1.58)	M-H
		Flatus incontinence*	2	0.79 (0.29–2.13)	4	1.16 (0.94–1.51)	M-H
6	Laparoscopic vs. open appendectomy for acute appendicitis	Wound infection	16	0.30 (0.19–0.47)	7	0.43 (0.21–0.84)	M-H
7	Transthoracic vs. transhiatal resection for carcinoma of the esophagus	Mortality*	3	0.12 (0.04–1.12)	20	1.43 (1.08–1.89)	M-H
		Cardiac complication*	2	0.77 (0.30–1.99)	5	1.19 (0.70–2.01)	M-H
		Pulmonary complication*	2	0.85 (0.53–1.38)	10	1.20 (0.99–1.46)	M-H
		Anastomotic leak*	3	1.20 (0.34–4.25)	14	0.49 (0.38–0.64)	M-H
		Vocal cord paralysis*	2	0.98 (0.14–6.59)	9	0.51 (0.33–0.78)	M-H
		3-yr survival*	1	1.83 (0.70–4.78)	8	1.44 (1.12–1.86)	M-H
8	Hand-sewn vs. stapled esophagogastric anastomosis after esophagectomy	Mortality	4	0.41 (0.17–0.98)	3	1.87 (0.76–4.57)	D-L
		Anastomotic leak*	5	0.79 (0.44–1.42)	10	1.77 (1.22–2.56)	D-L
		Anastomotic stricture*	4	0.60 (0.27–1.33)	7	0.79 (0.41–1.50)	D-L
9	Posterior vs. anterior route of reconstruction after esophagectomy	Mortality*	3	0.56 (0.17–1.82)	3	0.56 (0.18–1.72)	D-L
		Anastomotic leak*	4	1.01 (0.35–2.94)	3	0.28 (0.10–0.79)	D-L
		Pulmonary complication*	3	0.67 (0.34–1.33)	3	0.81 (0.50–1.34)	D-L
		Cardiac complication*	3	0.43 (0.17–1.12)	2	0.87 (0.44–1.74)	D-L
10	Pyloroplasty vs. no drainage in gastric reconstruction after esophagectomy	Pulmonary complication*	2	0.69 (0.42–1.14)	2	4.07 (0.91–18.3)	D-L
11	Primary repair vs. fecal diversion for penetrating colon injuries	Mortality	5	1.70 (0.51–5.70)	25	0.43 (0.33–0.55)	Peto
		Morbidity	5	0.28 (0.18–0.42)	20	0.73 (0.60–0.90)	Peto
		Intraabdominal infection	5	0.59 (0.38–0.94)	20	0.60 (0.49–0.74)	Peto
		Wound infection	5	0.55 (0.34–0.89)	18	0.78 (0.62–0.98)	Peto
12	Stapled vs. hand-sewn methods for colorectal anastomosis surgery	Mortality	7	0.69 (0.32–1.49)	12	0.74 (0.51–1.07)	Peto
		Anastomotic leak	9	0.99 (0.71–1.40)	11	1.16 (0.82–1.64)	Peto
		Anastomotic stricture	7	3.59 (2.02–6.35)	5	3.78 (1.40–10.2)	Peto
		Hemorrhage	4	1.78 (0.84–3.81)	2	0.59 (0.08–4.19)	Peto
		Reoperation	3	1.94 (0.95–3.98)	3	0.18 (0.12–0.26)	Peto
		Wound infection	6	1.43 (0.67–3.04)	5	1.28 (0.83–1.97)	Peto

(Continued)

TABLE 2. *(Continued)*

Identification No.	Topic	Outcome	Randomized Controlled Trial		Observational Study		Calculation Model
			No. of Studies	Summary Estimate OR (95% CI)	No. of Studies	Summary Estimate OR (95% CI)	
13	Stapled vs. conventional hemorrhoidectomy	Thrombosis of external piles*	2	0.56 (0.19–1.61)	2	0.71 (0.14–3.58)	M-H
		Urinary retention*	3	0.59 (0.28–1.24)	3	0.41 (0.23–0.72)	M-H
		Anal stenosis (2–6 wk)*	2	1.07 (0.36–3.17)	3	0.55 (0.16–1.83)	M-H
14	Extended vs. limited lymph node dissection for adenocarcinoma of the stomach	Mortality	2	2.39 (1.50–3.82)	2	0.63 (0.43–0.93)	M-H
		5-yr survival	2	0.92 (0.72–1.17)	2	1.17 (0.97–1.42)	M-H
15	Open (Hasson type) vs. closed (needle/trocar) access in laparoscopic surgery	Major complication	1	0.33 (0.04–3.13)	6	1.54 (0.70–3.40)	M-H
		Minor complication	2	0.82 (0.44–1.54)	5	0.52 (0.26–1.05)	M-H
		Conversion to laparotomy	2	0.32 (0.05–1.96)	4	0.43 (0.16–1.21)	M-H
16	Direct trocar vs. closed (needle/trocar) access in laparoscopic surgery	Major complication	1	1.07 (0.07–16.9)	1	0.09 (0.00–1.90)	M-H
		Minor complication	3	0.19 (0.09–0.40)	2	0.07 (0.04–0.14)	M-H
		Conversion to laparotomy	1	1.17 (0.16–8.58)	1	0.09 (0.00–1.90)	M-H
17	Early vs. delayed open cholecystectomy for acute cholecystitis	Mortality	9	0.53 (0.17–1.66)	13	1.73 (0.89–3.37)	D-L
		Morbidity	9	0.95 (0.66–1.38)	12	0.95 (0.59–1.54)	D-L
		Common bile duct injuries	9	0.66 (0.20–2.17)	3	1.35 (0.24–7.61)	D-L
18	Early vs. delayed laparoscopic cholecystectomy for acute cholecystitis	Morbidity	3	0.69 (0.27–1.73)	11	0.98 (0.53–1.80)	D-L
		Common bile duct injuries	3	0.70 (0.07–6.19)	11	1.27 (0.56–2.87)	D-L
		Conversion to laparotomy	3	0.62 (0.32–1.19)	16	0.39 (0.14–1.07)	D-L

OR, odds ratio; CI, confidence interval; M-H, Mantel-Haenszel method; D-L, DerSimonian-Laird method; Peto, Peto's odds ratio method.

*Outcome reporting relative risk rather than odds ratio.

concordance of different studies may vary depending on whether primary or secondary outcomes are examined. Discrepancies may be less apparent for secondary outcomes than for primary outcomes because secondary events are likely to be too uncommon to show any significant difference between arms except in extremely large trials (mega-trials).[17]

One possible explanation for the greater frequency of between-study heterogeneity in observational studies than in randomized trials is that each observational study usually includes a wide spectrum of subjects from the population at risk. In contrast, randomized trials use specific inclusion criteria and may not be representative of populations seen in clinical practice.

All topics examined in this study were comparisons in the form of A versus B. Generally, A represented a new procedure while B represented an accepted method, but deciding which was newer was difficult in some topics. Most trials in medicine estimate the benefits of pharmacologic effects, whereas 50 of 52 outcomes in this study estimate risks of operations, such as mortality and morbidity. Discrepancies in summary estimates were estimated accordingly between randomized trials and observational studies. For example, the greatest statistical discrepancy between the two types of study design was topic 14 (mortality), comparing extended and limited lymph node dissections for adenocarcinoma of the stomach ($Z = 4.34$). In this topic, although the summary estimate from observational studies was one fourth that from randomized trials (0.63 vs. 2.39), this represented an underestimation of risks, not of benefits.

The authors revealed that one fourth of observational studies gave different results to randomized trials and between-study heterogeneity was more common in observational studies in the field of digestive surgery. Furthermore, even if clinical applicability is improved by combining a large number of observational studies, estimations of treatment effect sometimes differ from those obtained from randomized controlled trials. The present study confirmed such tendencies in the well-defined area of digestive surgery. However, observational studies offer several advantages over randomized controlled trials, including lower cost, greater timeliness, and a broader range of patients.[34] These benefits remain worthy of attention in real clinical settings, particularly where random allocation is not easily accepted by either clinicians or patients. In the field of digestive surgery, large

observational studies may actually be more reliable than small underpowered randomized controlled trials. To clarify how to interpret the findings of observational studies and randomized controlled trials, further analyses in other fields are eagerly awaited.

REFERENCES

1. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ.* 1948;2:769–782.
2. Preventive Services Task Force. *Guide to Clinical Preventive Services*: *Report of the U.S. Preventive Services Task Force*, 2nd ed. Baltimore: Williams & Wilkins, 1996.
3. Chalmers TC, Matta RJ, Smith H Jr, et al. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med.* 1977;297:1091–1096.
4. Sacks HS, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233–240.
5. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I. *Med Stat Med.* 1989;8:441–454.
6. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II. *Surgical Stat Med.* 1989;8:455–466.
7. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342:1878–1886.
8. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887–1892.
9. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA.* 2001;286:821–830.
10. Yusuf S, Peto R, Lewis J, et al. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis.* 1985;27:335–371.
11. Mantel N, Haenszel WH. Statistical aspects of the analysis of data from retrospective studies of diseases. *J Natl Cancer Inst.* 1959;22:719–748.
12. Fleiss JL. *Statistical Methods for Rates and Proportions*. New York: Wiley, 1981.
13. Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials.* 1986;7:267–275.
14. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7:177–188.
15. Julian PTH, Simon GT, Jonathan JD, et al. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327:557–560.
16. Matthias E, George DS, Douglas GA. *Systematic Reviews in Health Care*: *Meta-Analysis in Context*, 2nd ed. London: BMJ Books, 2001.
17. Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons of meta-analyses and large trials. *JAMA.* 1998;279:1089–1093.
18. Leiboff AR, Soroff HS. The treatment of generalized peritonitis by closed postoperative peritoneal lavage: a critical review of the literature. *Arch Surg.* 1987;122:1005–1010.
19. Spina GP, Henderson JM, Rikkers LF, et al. Distal spleno-renal shunt versus endoscopic sclerotherapy in the prevention of variceal rebleeding: a meta-analysis of 4 randomized clinical trials. *J Hepatol.* 1992;16:338–345.
20. Urbach DR, Kennedy ED, Cohen MM. Colon and rectal anastomoses do not require routine drainage: a systematic review and meta-analysis. *Ann Surg.* 1999;229:174–180.
21. Nelson RL. Meta-analysis of operative techniques for fissure-in-ano. *Dis Colon Rectum.* 1999;42:1424–1428.
22. Hulscher JB, Tijssen JG, Obertop H, et al. Transthoracic versus transhiatal resection for carcinoma of the esophagus: a meta-analysis. *Ann Thorac Surg.* 2001;72:306–313.
23. Urschel JD, Blewett CJ, Bennett WF, et al. Handsewn or stapled esophagogastric anastomoses after esophagectomy for cancer: meta-analysis of randomized controlled trials. *Dis Esophagus.* 2001;14:212–217.
24. Urschel JD, Urschel DM, Miller JD, et al. A meta-analysis of randomized controlled trials of route of reconstruction after esophagectomy for cancer. *Am J Surg.* 2001;182:470–475.
25. Urschel JD, Blewett CJ, Young JE, et al. Pyloric drainage (pyloroplasty) or no drainage in gastric reconstruction after esophagectomy: a meta-analysis of randomized controlled trials. *Dig Surg.* 2002;19:160–164.
26. Singer MA, Nelson RL. Primary repair of penetrating colon injuries: a systematic review. *Dis Colon Rectum.* 2002;45:1579–1587.
27. Lustosa SA, Matos D, Atallah AN, et al. Stapled versus handsewn methods for colorectal anastomosis surgery: a systematic review of randomized controlled trials. *Sao Paulo Med J.* 2002;120:132–136.
28. Sutherland LM, Burchard AK, Matsuda K, et al. A systematic review of stapled hemorrhoidectomy. *Arch Surg.* 2002;137:1395–1406.
29. McCulloch P, Nita ME, Kazi H, et al. Extended versus limited lymph nodes dissection technique for adenocarcinoma of the stomach. *Cochrane Database Syst Rev.* 2003;(4):CD001964.
30. Merlin TL, Hiller JE, Maddern GJ, et al. Systematic review of the safety and effectiveness of methods used to establish pneumoperitoneum in laparoscopic surgery. *Br J Surg.* 2003;90:668–679.
31. Papi C, Catarci M, D'Ambrosio L, et al. Timing of cholecystectomy for acute calculous cholecystitis: a meta-analysis. *Am J Gastroenterol.* 2004;99:147–155.
32. Solomon MJ, McLeod RS. Surgery and the randomized controlled trial: past, present and future. *Med J Aust.* 1998;169:380–383.
33. McCulloch P, Taylor I, Sasako M, et al. Randomized trials in surgery: problems and possible solutions. *BMJ.* 2002;324:1448–1451.
34. Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. *J Clin Epidemiol.* 1989;42:481–489.