Cancer Chemother Pharmacol (2009) 63:919–927
DOI 10.1007/s00280-008-0808-5

ORIGINAL ARTICLE

Postoperative adjuvant chemotherapy of gastric cancer: scrutiny into the clinical evidence based on quality assessment of medical literature of randomized controlled trials

Hei-Cheul Jeung · Sun Young Rha · Sang Joon Shin · Joong Bae Ahn · Jae Kyung Roh · Chan Hee Park · Sung Hoon Noh · Hyun Cheol Chung

Received: 10 May 2008 / Accepted: 15 July 2008 / Published online: 20 August 2008

Abstract The aim of this study was to scrutinize the evidence of adjuvant chemotherapy of gastric cancer by assessing the quality of the medical literature of randomized controlled trials (RCTs). A quality assessment (QA) scoring system was devised with the three parameters—control of bias, quality of report, and quality of design—which consisted 19 items. We searched for all the publications of the RCTs, from 1969 to 2007, with surgery-only arm, and their associated meta-analyses to score. Among the 26 RCTs, quality of three articles were graded as (2+), 10 articles as (1+), and 13 articles as (−). Recently published studies had overall better quality of report, but not necessarily better quality of design. Three studies demonstrating a positive survival benefit of adjuvant chemotherapy had a grade (1+). Hierarchical clustering revealed that the 26 articles were grouped into three major branches associated with study quality and a multi-institutional setting. We also obtained a statistically significant set of ten items ($P < 0.001$) that could differentiate articles of good (1–2+) and low quality (−) through supervised two-way hierarchical clustering. Finally, the level of recommendation for adjuvant chemotherapy in gastric cancer was to be a "B" according to the Scottish Intercollegiate Guidelines Network (SIGN) System. QA of medical literature should be an essential consideration for medical-related decision-making and the formation of evidence-based guidelines. Multidisciplinary discussion to develop and refine trial design is important for procuring better quality of RCTs of adjuvant chemotherapy of gastric cancer.

Keywords Adjuvant chemotherapy · Gastric adenocarcinoma · Information retrieval · Randomized controlled trial · Quality assessment

H.-C. Jeung · S. Y. Rha · J. K. Roh · C. H. Park · S. H. Noh · H. C. Chung (✉)
Cancer Metastasis Research Center, Yonsei Cancer Center, Yonsei University College of Medicine, 250 Seongsanno (134 Shinchon-Dong), Seodaemun-Gu, Seoul 120-752, South Korea
e-mail: unchung8@yuhs.ac

H.-C. Jeung · S. Y. Rha · S. J. Shin · J. B. Ahn · J. K. Roh · H. C. Chung
Department of Internal Medicine, Yonsei University College of Medicine, 250 Seongsanno (134 Shinchon-Dong), Seodaemun-Gu, Seoul 120-752, South Korea

S. H. Noh
Department of Surgery, Yonsei University College of Medicine, Seoul, South Korea

Introduction

Evidence-based medicine depends on reliable data from clinical trials in order to guide medical and public health decisions. Thorough examination of published clinical trials enables us to evaluate inherent bias, validity and applicability of the results, and the necessity for further investigation [1]. Scrutiny can also aid in the interpretation of conflicting results. However, it is not an easy task to determine the efficacy of a medical intervention from the limited range of published clinical trials. This challenge is especially true if studies addressing the same therapeutic problem produce conflicting results each other, such as these with postoperative adjuvant chemotherapy of gastric cancer. The quality of literature is also important with regard to the peer-review process and meta-analysis.

There are three widely accepted methods to assess the quality of clinical trials; individual markers, checklists, and numerical scales [2]. Of these, the numerical scale system

has a theoretical advantage of providing quantitative estimates of quality. It can be replicated easily and incorporated into the peer-review or systematic review process. However, there is a dearth of evidence to support the inclusion or exclusion of items and the numerical scores attached to each item, which vary according to researchers [2].

Therapeutic approaches in gastric cancer have long been focused on radical surgery, but many researchers believe that adjuvant chemotherapy can improve the prognosis of gastric cancer by eradicating micrometastasis. Thus, many randomized clinical trials (RCTs) have been designed to prove the positive effect of survival of adjuvant chemotherapy. However, despite enormous efforts, a 40-year history of randomized studies has not yet supported the concept of adjuvant chemotherapy as a standard therapy for gastric cancer. Rather, National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines recommend concurrent chemoradiotherapy for serosa- or lymph node-positive cancers, which were based on the landmark study of McDonald et al. [3, 4]. Many—especially Asian—researchers who have different surgical concept of gastric cancer, take a cautious attitude that the debate of adjuvant chemotherapy in gastric cancer should not be concluded so hastily due to the lacking of sufficient data so far enough to reach a definite role of adjuvant chemotherapy. Moreover, we believe that there has been a lack of scrutiny of both the quality of study design and the quality of report of publications. Many review articles and meta-analyses have failed to consider a QA of the publications they include.

The aim of this study was to scrutinize the clinical evidence of adjuvant chemotherapy of gastric cancer. To do so, we first searched for and selected publications on RCTs of adjuvant chemotherapy compared with surgery-only arm. We developed a scoring instrument for QA, and then scored the literature to classify them. Using this data, we then investigated which items are crucial for discriminating the quality of literature by statistical approach. Finally, we used the Scottish Intercollegiate Guidelines Network (SIGN) grading system to provide a recommendation level based on QA about adjuvant chemotherapy of gastric cancer [5].

Materials and methods

Information retrieval

In July 2007, we did a keyword search of the Cochrane and PubMed database for the following items: gastric neoplasms [MeSH term] OR gastric cancer [text word] AND adjuvant chemotherapy [text word], and randomized controlled trials [text word] conducted from 1969 to 2007. References in retrieved papers were also evaluated for other possible pertinent studies. When the same results were reported or updated in more than one publication, only the most informative publication was included. There was no language restriction or exclusion of papers from the Asian region. Finally, each retrieved article was classified into one of six categories based on Nygren et al. [6]: meta-analysis, RCT (phase III trial), phase I/II trial, retrospective study, review, and others (case report, editorial comments, animal experiment, technical report, and so forth). Inclusion for further analysis and QA was performed only for reports of RCTs and meta-analyses of RCTs. A search for unpublished results was not performed. Other eligibility criteria were as follows: (1) a control arm of surgery alone, (2) an experimental arm of surgery plus systemic chemotherapy, and (3) surgical resection of curative aim (studies of metastatic or palliative surgery were excluded). Studies exploring immunotherapy (±chemotherapy), intraperitoneal chemotherapy, neoadjuvant or perioperative chemotherapy, or radiotherapy (±chemotherapy) were excluded. Survival data had to be extractable in numerical graphical form as the number of deaths and total number of patients in each group.

A predesigned extraction form was used which included the following data: eligibility, year of publication, accrual years, study scale (single or multicenter), follow-up duration, treatment regimen, number of patients accrued, stratification factors, number of deaths in each arm, percentage of lymph node metastasis in each arm, and author's conclusion.

Development of QA instrument

Instrument for QA comprised three categories: (1) control of bias, (2) quality of report, and (3) quality of trial design compatible for adjuvant chemotherapy. Each category included 2, 10, and 7 items, respectively, which were generated and transformed into questions. Each answer was graded by five levels: well conducted (WC), appropriately conducted (AC), poorly conducted (PC), not reported (NR), and not assessed (NA) (Fig. 1). The overall quality score of an article was generated by summing the numbers of items scored WC or AC, and was graded as (2+) (almost all criteria are fulfilled; $\geq$16 items), (1+) (some of the criteria are fulfilled; 9–15 items), or (−) (if few or no criteria are fulfilled; $\leq$8 items), based on the report of Liddle et al. [7]. Three reviewers independently reviewed all publications for eligibility and quality assessment.

Statistical analysis

Inter-rater reliability between two results was analyzed using Kappa statistics. Unsupervised two-way (articles and items) hierarchical clustering was conducted based on the

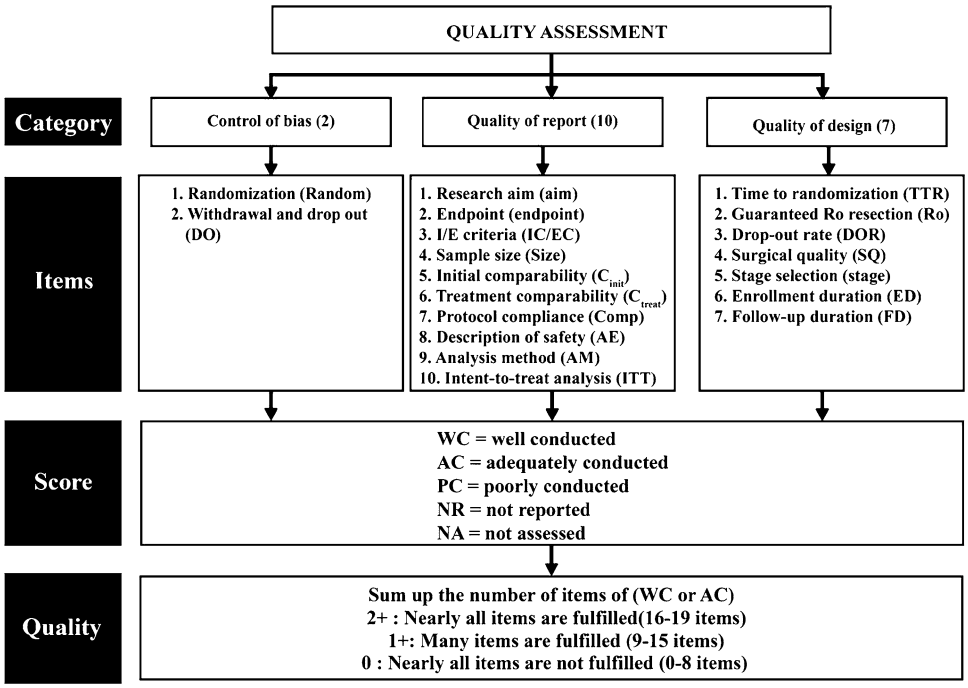

Fig. 1 Study algorithm of quality assessment of medical publicatrions. *Characters shown in the parenthesis* next to each item indicate the abbreviation corresponding to each item for further analysis. *I/E criteria* denotes inclusion/exclusion criteria

scores of all the items with GeneSpring 7.2 (Agilent, Palo Alto, CA, USA). Selection of parameters that determined the quality of an article was performed using multiple linear regression analysis. A statistically significant set of items was further selected using a two-sample t test at $P < 0.001$.

Results

Search and selection of publications

The selection procedure is diagrammed in Fig. 2. We found 658 citations and 465 were excluded after evaluation of the abstracts. Among the remaining 183 papers, there were 77 meta-analyses or RCTs, 37 of which were excluded because of pooling of chemotherapy ($n = 10$) and immunochemotherapy ($n = 19$) arms, and intraperitoneal chemotherapy ($n = 8$). In the remaining 33 articles with a surgery-control arm, seven papers were further excluded for the following reasons: repetition of the same data in different papers ($n = 4$), publication before 1980 ($n = 2$), and those written in statistical view ($n = 1$). Finally, 26 publications describing RCTs in which a surgery-only arm was included, and six meta-analyses were considered for inclusion in this study [8–39]. The RCTs included in this study are summarized in Table 1. Although the report of Grau is an updated form of Alcobendas from the same study group, we included both articles for analysis because the number of accrued patients was doubled, and we thought that a 10-year interval was enough time for a change in overall

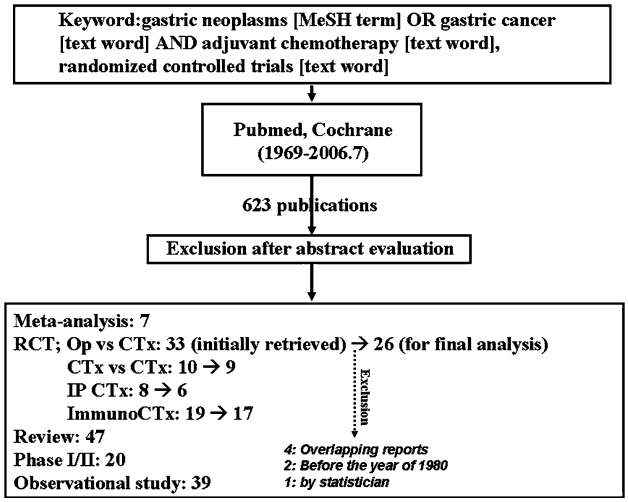

Fig. 2 Publication search and selection process. *Op* operation, *CTx* systemic chemotherapy, *IP CTx* intraperitoneal chemotherapy, *immunoCTx* immunochemotherapy

literature quality [13, 20]. Nine were multi-institutional trials and five were Asian studies. In total, 26 studies included 5,435 patients, out of whom 2,826 were treated and 2,609 were control patients. Six individual studies found that the adjuvant chemotherapy showed positive effect for survival that was statistically significant compared to the surgery-only group.

Quality assessment of RCT reports

Quality assessment of the literature according to the 19 items described in "Materials and methods" is depicted

Table 1 Randomized controlled trials included in the quality assessment of adjuvant chemotherapy in resectable gastric cancer with surgery-only arm

Study	Year	Accrual	Multicenter	Stage	No. patient treated/control	FUD (month)	Time-to-randomization	Drop out (%)	R0 resection	Surgery quality description
Huguier (France)	80	1969–1973	No	–	23/26	–	<2 months	-	-	–
Nakajima (CIHJ)	80	1970–1974	No	I–IV	42/40	–	Immediate	5	-	–
Blake (UK)	81	1973–1978	No	–	29/34	–	6–8 weeks	–	Residual	–
Schlag (German)	82	1976–1979	No	II–III	49/54	–	42–46 days	3	–	–
GITSG	82	1975–1980	Yes	–	71/71	–	<6 weeks	14	R0	Yes
VASOG	83	1974–1980	Yes	I–III	156/156	–	–	5	Residual	–
Alcobendas (Spain)	83	1977–unknown	No	IB–III	33/37	62	<4 weeks	–	–	–
Nakajima (CIHJ)	84	1974–1977	No	I–IV	164/79	–	<10 days	9	–	–
ECOG	85	1975–1980	Yes	–	91/89	64	<6 weeks	0	R0	–
BSCG	89	1976–1981	Yes	II–III	281/130	100	<12 weeks	10	Residual	–
ICCG	90	1981–1984	Yes	II–III	133/148	68	2 weeks	8	Incomplete	Yes
NCCTG	91	–	Yes	I–III	61/64	84	4–6 weeks	2	R0	–
Grau(Spain)	93	1977–1983	No	IB–III	68/66	134	<6 weeks	–	Residual	–
BSCG	94	1981–1986	Yes	II–IV	138/145	–	4 weeks	17	Residual	Partial
Taiwan	94	1986–1992	No	II–III	59/56	22	2 weeks	9	Incomplete	Partial
EORTC	95	1980–1989	Yes	II–III	155/159	78	4 weeks	–	Residual	Yes
SWOG	96	1978–1991	Yes	IB–III	93/100	–	8 weeks	0	Residual	Yes
Neri (Italy)	96	1989–1991	Yes	II–IV	48/55	–	4–6 weeks	–	R0	Yes
Greece	96	1988–1994	No	I–III	42/42	60	2–3 weeks	9	R0	Yes
CCHCGGC	99	1988–1994	Yes	III	76/72	37	<4 weeks	5	R0	–
Nakajima (CIHJ)	99	1988–1992	Yes	IB–IV	288/285	72	Immediate	3	R0	Partial
ITMO	02	1992–1997	Yes	III–IV	135/136	66	<60 days	4	R0	Yes
JCOG	03	1993–1994	Yes	IB–III	127/123	69	Immediate	<1	R0	–
AURC	04	1989–1997	Yes	–	101/104	101	<4 weeks	5	R0	Partial
FFCDG	05	1989–1997	Yes	II–IV	127/133	98	14 days	4	R0	Yes
EORTC/IGCC	06	1990–1998	Yes	IB–IV	194/203	78	<4 weeks	5	Residual	Yes

Abbreviation: *FUD* follow-up duration

in Fig. 3. There were three articles of grade (2+) and ten articles of grade (1+). The remaining 13 articles were of grade (−). We examined the agreement of our results with the previous critical review by Janunger et al. [40], including 18 publications of RCTs. When we extracted the 18 corresponding publications between the two studies and compared them, the inter-rater reliability designated by Kappa value was 0.333, which was statistically significant ($P = 0.023$). When we divided Janunger's scores into two groups (high-moderate evidence vs. low evidence) and analyzed again the degree of agreement, the Kappa value increased to 0.483 ($P = 0.017$). The concordance rate between the two scoring methodologies was 72%. All discordant cases were scored lower in our system compared to Janunger's, implying that our scoring system was stricter in the estimation of quality.

Quality assessment of articles reporting a positive survival benefit of chemotherapy

Six publications advocated the role of adjuvant chemotherapy as improving overall survival [9, 11, 20, 22, 25, 27]. However, three studies were excluded for further analysis. The reported survival benefit of the study of Grau was not consistent. The authors described an observed/expected ratio as significant ($P = 0.025$), but a calculation of overall hazard on the given figures has revealed a difference that is not statistically significant (95% CI, 0.23–1.03) [20, 40]. The Taiwan study was excluded because survival analysis was performed with the only half of patients [22]. Japanese study was also excluded because a later larger-scaled study of better quality yielded no survival benefit with adjuvant chemotherapy [9, 30]. The remaining three studies advocated survival benefit, particularly in serosa- or lymph

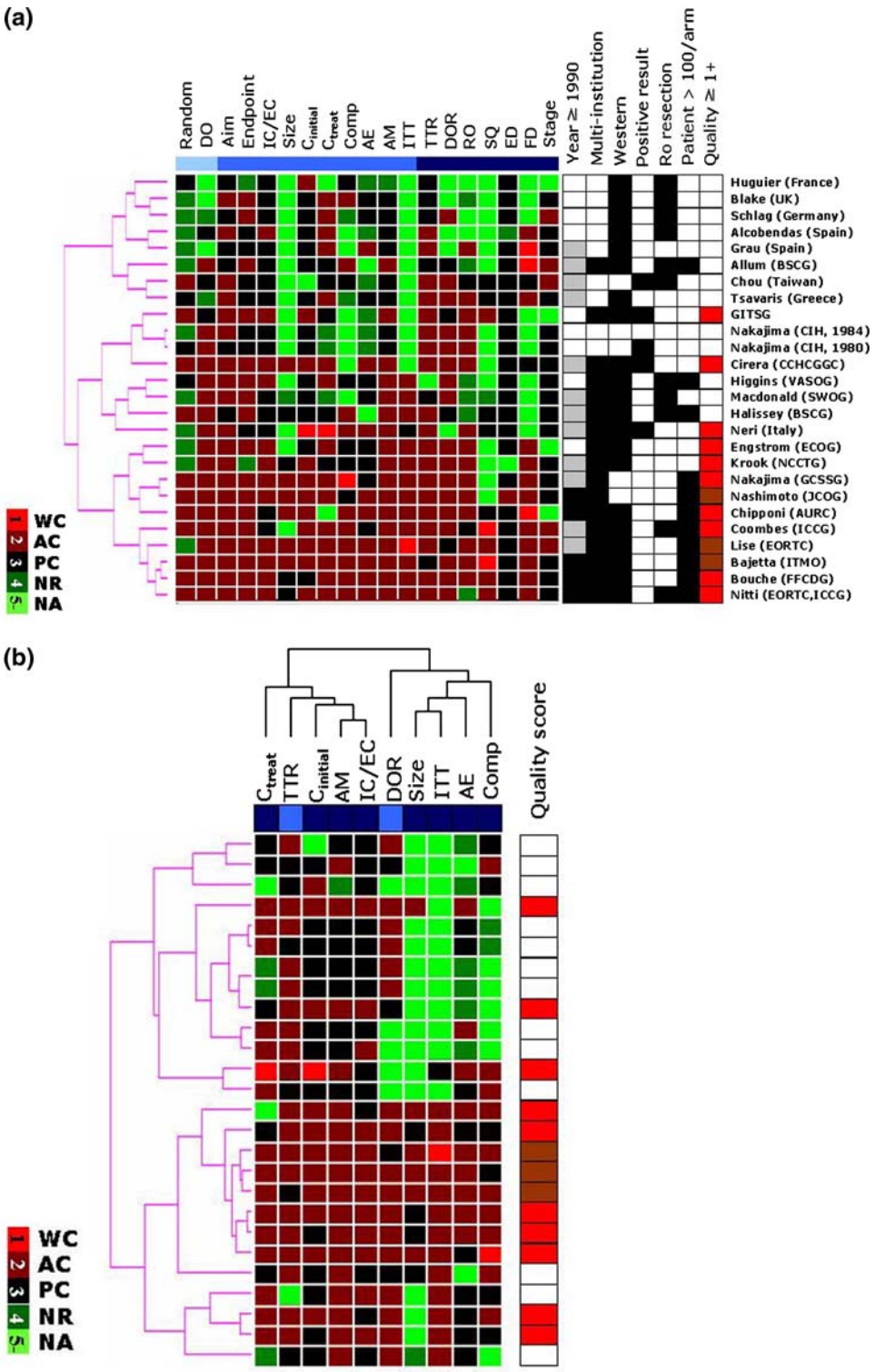

Fig. 3 Unsupervised two-way (articles and items) hierarchical clustering was conducted based on the scores of all the items. *Brown blocks* indicate a quality score of (2+). *Gray blocks* indicate publication period from 1985 to 1990. Note that 26 articles (*vertical axis*) were grouped into three major branches (*purple lines* on the left side) by the QA items (*horizontal axis*). Abbreviations for items are the same as in Fig. 1a. Supervised clustering of the selected 10 items was associated with quality score (**b**)

node-positive, and in radically resected cancer (R0 resection). All these studies had a quality score of (1+), but only one study reported sufficient statistical power to prove a survival benefit of adjuvant chemotherapy. The GITSG trial included 165 patients (of which 142 were analyzed) [11]. This was the first report in which surgery qualification was described. Patient distribution and stratification were described well, and the mean inclusion rate was 3.3 patients per year in an institution. Neri et al. [25] evaluated 103 patients. About 20% of control patients had died before the first patient in the treatment group died. Randomization, statistical power, and follow-up duration were not

described. Finally, Cirera et al. [27] a multi-institutional study in Spain, enrolled 148 patients. Statistical power and comparability factors were adequately described. But, follow-up duration was only 37 months, and the treatment compliance was not clearly described.

Quality assessment of retrieved articles in meta-analyses

From 1980 to 2007, there were six publications on the meta-analyses of RCTs [34–39]. The median number of articles included in a meta-analysis was 15.5 (range 13–21). Asian studies were excluded in one study, and none contacted authors or dealt with unpublished results. All but one study were restricted to English language papers. The proportion of good-quality (≥1+) articles included in the analyzed publications was only 22–42% (Table 2), implying that the meta-analyses of adjuvant chemotherapy contains a range of publications with a high risk of bias with respect to QA.

Determination of factors associated with quality of RCTs

To obtain comprehensive information about quality patterns in the publications, we used an unsupervised two-way (articles and items) hierarchical clustering algorithm to group all the articles which share to similarities in their quality pattern, by the 19 evaluation items (Fig. 3a). Hierarchical clustering showed that the articles were grouped into three major branches. We found that two Japanese studies of similar trial design or the aforementioned two Spanish studies of the same study group were clustered together [9, 12, 15, 19]. The three major clusters were independent of study characteristics such as geographical area (Western or Asian), inclusion of R1-2 resected patients, or the benefit of adjuvant chemotherapy (positive or negative for survival). Rather, the clusters seemed to be associated with the publication year (<1990 or ≥1990), study scale (number of enrolled patients per arm, <100 or >100), and study quality [(−) or (1–2+)].

Next, we applied a two-sample *t* test and multiple regression analysis to select a set of statistically significant QA factors ($P < 0.001$) that discriminated the articles as good [(1–2+)] or poor quality (−). Eight items showed differential scores according to the article quality: description of initial and treatment comparability (C_{init}, C_{treat}), analytical method (AM), inclusion/exclusion criteria (IC/EC), sample size (Size), intent-to-treat analysis (ITT), adverse events (AE), and treatment compliance (Comp). Two items of the quality of design—time-to-randomization (TTR) and drop-out rate (DOR)—were also added. Then, we performed the supervised two-way hierarchical clustering accommodating these ten factors. It showed clustering of two groups based on score quality [(1–2+) vs. (−)], except six articles (three each in each groups) which were misclustered (Fig. 3b). Overall, the concordance rate was 77%.

Grading of literature of adjuvant chemotherapy for evidence-based guidelines by the SIGN system

The published evidence for adjuvant chemotherapy in gastric cancer based on a QA algorithm was examined. We chose the revised SIGN grading system for this purpose, which has the advantage of considering both study design and methodological quality of individual studies. Grades of recommendation were based on the strength of supporting evidence, taking into account its overall level and the considered judgment of the guideline developers. We conferred

Table 2 Approaches for quality assessment in the six meta-analysis studies

Literature	Hermans	Earle	Mari	Panzini	Hu	Janunger
Searched database	1980–1991	1966–1999	1965–1999	1980–2000	1980–2001	1969–1996
Hand searching	Not started	Yes	Yes	Yes	Yes	Yes
Contacting authors	Not done	Not done	Not done	Not done	Not done	Not done
Non-published papers	Not done	Not done	Not done	Not done	Not done	Not done
Asian studies	Included	Excluded	Included	Included	Included	Included
Language restriction	Yes	Yes	Yes	Yes	No	Yes
Adjuvant treatment	Mixed	sCTx	sCTx	sCTx	Mixed	sCTx
Excluding						
Incomplete resection	No	Incomplete	No	Incomplete	No	No
No. articles in analysis	14	13	21	17	14	21
Proportion of (1+ to 2+)						
Articles included	29%	39%	34%	42%	22%	24%
No. quality score-	4	6	9	5	2	10
No. quality score 1+ to 2+	3	5	7	7	3	5

Abbreviation: *sCTx*, systemic chemotherapy

a score of (1−) when the level of evidence for adjuvant chemotherapy was based on meta-analyses, systematic reviews, or RCTs with a high risk of bias. Overall, we scored the level of recommendation for adjuvant chemotherapy in gastric cancer as (B) from a body of extrapolated evidence provided by (grade 1).

Discussion

RCT is accepted as the most robust way of study design with the least risk of bias [7]. Regardless of whether its result reaches statistical significance or not, the design, conduct, and report for publication should be of high quality. It is important to distinguish between the quality of a trial and the quality of its report [2]. Quality of a trial is defined by how well the trial was designed, conducted, and the analysis has minimized biases in its treatment comparisons. This definition focuses on the methodological quality. The quality of a report can be defined by how well it provides adequate information about the design, conduct, and analysis of a trial [2]. These two aspects can be mutually exclusive sometimes. A well-reported trial designed with several biases could receive a high-quality score and vice versa [2]. More than 30-year history of RCTs has not yet supported the concept of adjuvant chemotherapy as a standard approach in gastric cancer. We thought that it would be helpful to assess the quality of the literature for evidence-based clinical guidelines for adjuvant chemotherapy, which was the primary aim of this study.

There are many factors to be considered in the design of RCTs for adjuvant therapy (Fig. 4). The control of bias for randomization, double-blinding, and description of withdrawals and drop-outs is essential [1]. Although double-blinding is a prerequisite for RCTs, it is difficult to perform in the adjuvant setting, especially in cancer treatment. We found that only two early studies included a placebo in the control arm [13, 17]. As a result, we excluded double-blinding from the QA instrument. All other items selected for this QA instrument were based on the "widely accepted criteria" for clinical trials. The majority of scales included at least one item about patient assignment, follow-up, and analytical method, which are generic to all trials. This essential set of factors should be common across all the trials, and this allows useful comparison. However, some of them were based on convictions whereas others were based on empirical evidence. For example, we excluded the item "informed consent" because its systematic influence on the methodological quality cannot be estimated [2]. This distinction is more debatable when it comes to the quality of trial design. For example, many studies have allowed stage IV (M0) to be enrolled in adjuvant trial as far as it was radically resected. But high proportion of stage IV disease ultimately leads to low survival duration, and it sets off the—if any—survival benefit of the stage II–III disease. We are also not convinced of the treatment benefit in too early-staged (stage I) patients. Thus, we deemed studies that enrolled only stage II–III patients as "adequately conducted". Another example is the drop-out rate. We judged a study to be "adequately conducted" if the overall drop-out rate was less than 10% with equal proportions between treatment and control arms. The quality and extent of surgical dissection are yet another debate. Asian researchers

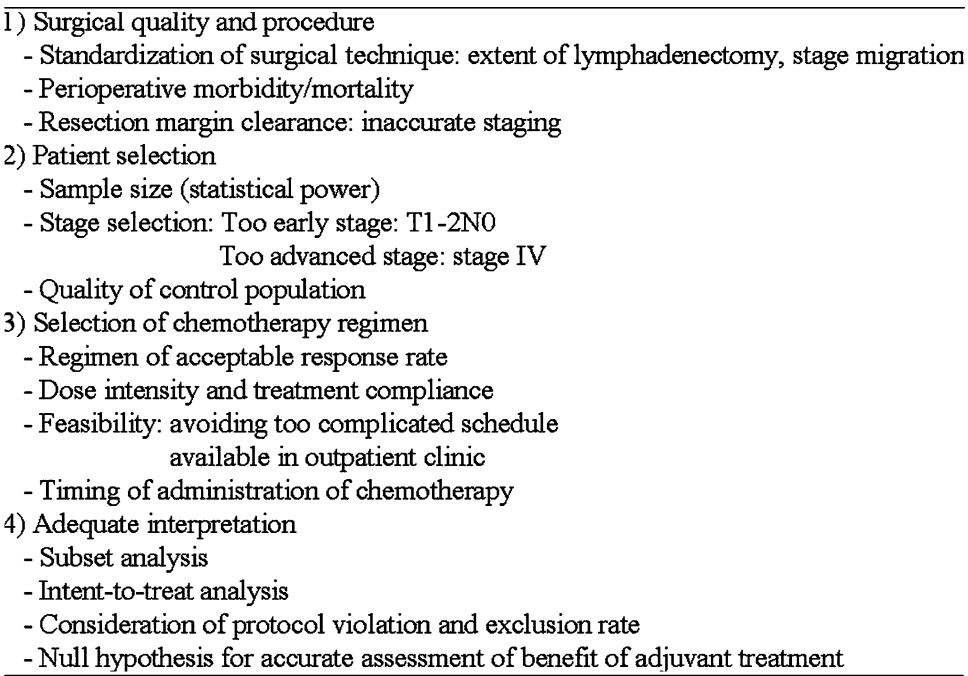

Fig. 4 Factors for consideration in the design of adjuvant clinical trials of gastric cancer

prefer the more extensive lymphadenectomy and more rigorous pathologic assessment of the lymphadenectomized specimen leads to secondary stage migration. Therefore, to establish a consensus on these items in the QA, we had to not only review large volume of literature, but also conduct rigorous multidisciplinary communications between medical or surgical oncologists and biostatisticians. We also addressed how compliance and safety analyses were incorporated in quality scores. These factors can influence the interpretation of the results and are likely to be more important for large collaborative groups for meta-analyses.

As far as we know, this is the first report focusing on the QA of RCTs of adjuvant chemotherapy. QA is a relatively new field but it is important. Of course, our results indicate a first step for improvement in how QA scales are developed. Invented instruments so far differ from one another in almost every aspect: why and how the items were selected for inclusion, the number of items included, their reliability, approximate time to completion, and scoring range [2]. Moreover, little attention has been given to the construct that the scales are assessing. Each publication contains highly heterogenous distribution of scores for each item, and we intended to use all these scores for analysis, not discriminating good or bad scores as most studies have done. Therefore, we introduced a novel statistical technique to incorporate all the data. We here adopted multiple linear regression analysis and (un)supervised hierachical clustering for this purpose, which was originated from analysis of more numerous or heterogeneous data such as microarray. This comprehensive view could help to minimize assessment bias and to grasp the true spectrum of quality. We believe that it could be useful for better analysis of QA by further refinement.

Our analysis demonstrated that quality is improving with recent reports, which might be the result of larger-scale collaborative studies. However, we also found that this increase in quality score often results from improvements in the report quality only, not necessarily accompanying a simultaneous improvement in trial design quality. Many items being used to assess trial designs are very heterogeneous in distribution. If we further add other provisional items for assessing adjuvant trial such as dose intensity, drug efficacy for gastric cancer, quality of life, and so forth, the scores grew poorer still (data not shown). Therefore, future efforts in assessing quality may be best spent in developing trial design and consensus generation.

We also applied this assessment system to other forms of adjuvant RCTs practicing other designs such as intraperitoneal chemotherapy, chemotherapy versus chemotherapy, and chemoimmunotherapy (Fig. 2). However, no studies had a score of (2+), and none had the proportion of score (1+) more than 35% of given trial design (data not shown). Therefore, we have little evidence as yet to justify adjuvant treatment of gastric cancer based on other forms of RCTs than systemic chemotherapy compared with surgery-only arm.

QA is also important in the meta-analysis. Results of the meta-analysis may become less meaningful if quality of the publications is not ensured. Several meta-analysis reports have advocated adjuvant chemotherapy for its benefits in gastric cancer, but our results show that those studies included considerable amount of low-quality papers with high levels of bias because of a lack of scrutiny. Therefore, we suggest that any future meta-analysis should consider the quality of literature.

Finally, we chose the SIGN system to evaluate whether quality assessment could determine the strength of evidence for adjuvant chemotherapy in gastric cancer. It is a recently developed system that excludes discrepancies of other previous systems [41]. It is also based on study design and the methodological quality of individual studies. Moreover, all studies related to specific questions were summarized in an evidence table [41]. To elucidate the development process and weaknesses of the SIGN system is outside the scope of this study, but previous results, before applying this system in practice, recommend them to be among important parameters. Further research will be required to establish the extent to which this new system meets the objectives set for it.

To conclude, through QA of publications of RCTs for adjuvant chemotherapy, we found that a great proportion had inadequate trial design quality. Moreover, the improvement observed in quality scores of recent publications is not necessarily attributable to a higher quality of design. Trial design should be the focus of future studies evaluating the exact benefit of the adjuvant chemotherapy. Though QA is not sufficient to explain the value of clinical evidence, it should be a critical parameters leading to the appropriate medical decision-making by providing evidence-based guidelines in this era of overflowing medical resources.

Acknowledgments We are grateful to Eun Young Joo and Won Joo Roh of Cancer Metastasis Research Center for the sincere helps of hers for data documentation and secretarial assistance. This work was supported by the Korea Science and Engineering Foundation (KOSEF) with a grant funded by the Korean government (MOST) (R11-2000-082-03002-0), and in part by a grant from the Korea Health 21 R&D Project, Ministry of Health and Welfare, Republic of Korea (A040151).

Conflict of interest statement No conflicts of interest exist.

References

1. Jadad AR, Moore RA, Carroll D et al (1996) Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 17:1–12
2. Moher D, Jadad AR, Nichol G et al (1995) Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklist. Control Clin Trials 16:62–73

3. NCCN clinical practice guidelines in oncology: Gastric cancer. V.1.2006. Available at URL: http://www.nccn.org/professionals/physician_gls/PDF/gastric.pdf
4. Macdonald JS, Smalley SR, Benedetti J et al (2001) Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. N Engl J Med 345:725–730
5. Harbour R, Miller J (2001) A new system for grading recommendations in evidence based guideline. Br Med J 323:334–336
6. Nygren P, Glimelius B (2001) The Swedish Council on Technology Assessment in Health Care (SBU) report on Cancer Chemotherapy—Project objectives, the working process, key definitions and general aspects on cancer trial methodology and interpretation. Acta Oncol 40:155–165
7. Liddle J, Williamson M, Irwig I (1996) Methods for evaluating research and guideline evidence (MERGE). New South Wales Department of Health, Sydney
8. Huguier M, Destroyes JP, Baschet C et al (1980) Gastric carcinoma treated by chemotherapy after resection: a controlled study. Am J Surg 139:197–199
9. Nakajima T, Fukami A, Takagi K et al (1980) Adjuvant chemotherapy with mitomycin C, and with a multi-drug combination of mitomycin C, 5-fluorouracil and cytosine arabinoside after curative resection of gastric cancer. Jpn J Clin Oncol 10:187–194
10. Blake JRS, Hardcastle JD, Wilson RG (1981) Gastric cancer: a controlled trial of adjuvant chemotherapy following gastrectomy. Clin Oncol 7:13–21
11. GITSG (1982) Controlled trial of adjuvant chemotherapy following curative resection for gastric cancer. The Gastrointestinal Tumor Study Group. Cancer 49:1116–1122
12. Schlag P, Schreml W (1982) Adjuvant 5-fluorouracil and BCNU chemotherapy in gastric cancer: 3-year results. Recent Results Cancer Res 80:277–283
13. Alcobendas F, Milla A, Estape J (1983) Mitomycin C as an adjuvant in resected gastric cancer. Ann Surg 198:13–17
14. Higgins GA, Amadeo JH, Smith DE (1983) Efficacy of prolonged intermittent therapy with combined 5-FU and methyl-CCNU following resection for gastric carcinoma. Cancer 52:1105–1112
15. Nakajima T, Takahashi T, Takagi K (1984) Comparison of 5-fluorouracil with ftorafur in adjuvant chemotherapies with combined inductive and maintenance therapies for gastric cancer. J Clin Oncol 2:1366–1371
16. Engstrom PF, Lavin PT, Douglass HO (1985) Postoperative adjuvant 5-fluorouracil plus methyl-CCNU therapy for gastric cancer patients: Eastern Cooperative Oncology Group Study. Cancer 55:1868–1873
17. Allum WH, Hallissey MT (1989) Adjuvant chemotherapy in operable gastric cancer. Lancet 18:571–574
18. Coombes RC, Schein PS, Chilvers CE (1990) A randomized trial comparing adjuvant fluorouracil, doxorubicin, and mitomycin with no treatment in operable gastric cancer. International Collaborative Cancer Group. J Clin Oncol 8:1362–1369
19. Krook JE, O'Conell MJ, Wieand HS (1991) A prospective, randomized evaluation of intensive-course 5-fluorouracil plus doxorubicin as surgical adjuvant chemotherapy for resected gastric cancer. Cancer 67:2454–2458
20. Grau JJ, Estape J, Alcobendas F (1993) Positive results of adjuvant mitomycin-C in resected gastric cancer: a randomized trial on 134 patients. Eur J Cancer 29A:340–342
21. Hallissey MT, Dunn JA, Ward LC et al (1994) The second British Stomach Cancer Group trial of adjuvant radiotherapy or chemotherapy in resectable gastric cancer: five-year follow-up. Lancet 343:1309–1312
22. Chou FF, Sheen-Chen SM, Liu PP et al (1994) Adjuvant chemotherapy for resectable gastric cancer: a preliminary report. J Surg Oncol 57:239–242
23. Lise M, Nitti D, Marchet A et al (1995) Final results of a phase III clinical trial of adjuvant chemotherapy with the modified fluorouracil, doxorubicin, and mitomycin regimen in respectable gastric cancer. J Clin Oncol 13:2757–2763
24. Macdonald JS, Fleming TR, Peterson RF et al (1995) Adjuvant chemotherapy with 5-FU, adriamycin, and mitomycin-C (FAM) versus surgery alone for patients with locally advanced gastric adenocarcinoma: a southwest oncology group study. Ann Surg Oncol 2:488–494
25. Neri B, de Neonardis V, Romano S et al (1996) Adjuvant chemotherapy after gastric resection in node-positive cancer patients: a multicentre randomized study. Br J Cancer 73:549–552
26. Tsavaris N, Tentas K, Kosmidis P et al (1996) A randomized trial comparing adjuvant fluorouracil, epirubicin, and mitomycin with no treatment in operable gastric cancer. Chemotherapy 42:220–226
27. Cirera L, Balil A, Batiste-Alentorn E et al (1999) Randomized clinical trial of adjuvant mitomycin plus tegafur in patients with resected stage III gastric cancer. J Clin Oncol 17:3810–3815
28. Nakajima T, Nashimoto A, Kitamura M et al (1999) Adjuvant mitomycin and fluorouracil followed by oral uracil plus tegafur in serosa-negative gastric cancer: a randomized trial. Gastric Cancer Surgical Study Group. Lancet 354:273–276
29. Bajetta E, Buzzoni R, Mariani L et al (2002) Adjuvant chemotherapy in gastric cancer: 5-year results of a randomized study by the Italian Trials in Medical Oncology (ITMO) Group. Ann Oncol 13:299–307
30. Nashimoto A, Nakajima T, Furukawa H et al (2003) Randomized trial of adjuvant chemotherapy with mitomycin, fluorouracil, and cytosine arabinoside followed by oral fluorouracil in serosa-negative gastric cancer: Japan Clinical Oncology Group 9206-1. J Clin Oncol 21:2282–2287
31. Chipponi J, Huguier M, Pezet D et al (2004) Randomized trial of adjuvant chemotherapy after curative resection for gastric cancer. Am J Surg 187:440–445
32. Bouche O, Ychou M, Burtin P et al (2005) Adjuvant chemotherapy with 5-fluorouracil and cisplatin compared with surgery alone for gastric cancer: 7-year results of FFCD randomized phase III trial (8801). Ann Oncol 16:1488–1497
33. Nitti D, Wils J, Guimaraes Dos Santos J et al (2006) Randomized phase III trials of adjuvant FAMTX or FEMTX compared with surgery alone in resected gastric cancer. A combined analysis of the EORTC GI Group and the ICCG. Ann Oncol 17:262–269
34. Hermans J, Bonenkamp JJ, Boon MC et al (1993) Adjuvant therapy after curative resection for gastric cancer: meta-analysis of randomized trials. J Clin Oncol 11:1441–1447
35. Earle CC, Maroun JA (1999) Adjuvant chemotherapy after curative resection after gastric cancer in non-Asian patients: revisiting a meta-analysis of randomized trials. Eur J Cancer 35:1059–1064
36. Mari E, Floriani I, Tinazzi A et al (2000) Efficacy of adjuvant chemotherapy after curative resection for gastric cancer: a meta-analysis of published randomized trials. Ann Oncol 11:837–843
37. Hu JK, Chen ZX, Zhou ZG et al (2002) Intravenous chemotherapy for resected gastric cancer: meta-analysis of randomized controlled trials. World J Gastroenterol 8:1023–1028
38. Panzini I, Gianni L, Fattori PP et al (2002) Adjuvant chemotherapy in gastric cancer: a meta-analysis of randomized trials and a comparison with previous meta-analysis. Tumori 88:21–26
39. Janunger KG, Hafstrom L, Glimelius B (2002) Chemotherapy in gastric cancer: a review and updated meta-analysis. Eur J Surg 168:597–608
40. Janunger KG, Hafstrom L, Nygren P et al (2001) A systemic overview of chemotherapy effects in gastric cancer. Acta Oncol 40:309–326
41. Evidence report/technology assessment Number 47. Systems to rate the strength of scientific evidence. AHRQ publication number 02-E016. Available at URL: http://www.ahrq.gov/clinic/epcsums/strenfact.pdf