

も、薬物治療と外科治療は本質的に違う種類のものなので、上記の結果が外科領域に当てはまるか否かは不明である。

そこでわれわれは、この問題について、外科領域における初の検証をおこなうこととし、また一つの臨床科に限定することで網羅的な検討を行うにした。消化器外科領域を、系統的かつ網羅的に調査、検討することにより「消化器外科領域では、観察研究はランダム化比較試験に比べ、治療効果を大きく見積もりやすいか」に対する解答を求めることとした。

## B. 方法

### RCTのメタアナリシスを検索、トピックの選定

2004年の4月までに出版されている消化器外科領域におけるRCTのメタアナリシスを本研究において比較するトピックにすることとした。採用基準は：1)RCTのメタアナリシス；2)消化器外科に関するもの；3)「手術vs.何か」という形で手術効果を比較している；4)ヒトに関するもの、である。使用言語は制限しなかった。また、研究の主たる目的が治療効果の評価ではないもの（診断に関するものなど）は除外することにした。文献検索はPubMed(1966年から2004年4月)、EMBASE(1986年から2004年4月)、コクランライブラリ(2004年 Issue2)などのデータベースを使用した。以下のMesh用語とテキスト用語の組み合わせを用いた：

“Surgical Procedures, Operative”；  
“Digestive System Surgical Procedures”；  
“randomized”；“random”；  
“meta-analysis”；“review”。参考文献のマニュアル検索も追加した。

メタアナリシスのためのOSを検索

レビュー文献の中にRCTのメタアナリシスもOSのメタアナリシスも両方ともにある場合、その結果をそのまま比較に用いることができる。しかし、OSのメタアナリシスが行われていない場合、われわれがそれを行うこととした。つまり、入手したレビュー文献の中にOSのメタアナリシスが見つからなかった場合は、以下のOSを検索する過程の中でOSのメタアナリシスも検索した。

OSのメタアナリシスを行うために、まず、決定したトピックにおけるOSを検索した。それぞれのトピックについて、RCTのメタアナリシスで用いられた採用基準と同じ基準を用いることとした。ここでいうOSデザインとは、前向き非ランダム化試験、後ろ向きコホート試験、症例対照研究、コントロールのある症例集積、他の非特異的研究である。文献検索はPubMed(1966年から2004年4月)、EMBASE(1986年から2004年4月)、コクランライブラリ(2004年 Issue2)などのデータベースを使用した。PubMedには観察研究(Observational study)に該当する検索用語がないため、以下のようなテキスト用語を用いることにした：“observational”，“non-randomized”，“case series”，“case control study”，“cohort”，“retrospective”，“prospective”，などである。これに加え、入手したレビュー文献の参考文献からマニュアル検索を行うとともに、レビュー文献の著者に連絡をとることもおこなった。

### データ抽出とアウトカムの選定

OSにおける利用可能なアウトカムのバイナリデータは全て抽出することにした。英語か日本語以外の言語で書かれている場合は、翻訳家により英語に翻訳してからデータ抽出をおこなった。ここまでの作業を二人の独立した研究者(SS,TN)が独立して行い、意見

の相違は議論により解決した。

最終的には、あるトピックに関して同じアウトカムに対する少なくとも1つのRCTと1つのOSとのバイナリデータがあれば比較は可能である。もしレビュー文献において主要アウトカムが定義されていれば、それをメイン比較に用いる事ができる。しかし主要アウトカムが不明な場合には、臨床的に最も重要だと思われることをデータ抽出者の間で決めることとした。消化器外科の場合、通常、術後死亡率が他のアウトカムに比べて臨床的に重要だと思われることが多い。

### 統計解析

すべてのトピック毎に、OSからのデータを統合した。

一般的には、研究間の不均一性を検定した後、Peto オッズ比法または Mantel-Haenszel 法などの母数効果モデルが使われる[10,11]。研究間不均一性の検討はQ検定を用いるが、この検定は検出力が弱いので、有意水準を0.05ではなく0.01にする[12,13]。もし、研究間の不均一性が有ると判断した場合、DerSimonian-Laird 法などの変量効果モデルを使うことになる[14]。

しかし、本研究では、可能な限り同一の条件で、RCT結果とOS結果とを比較することが目的であるので、OSのメタアナリシスを行う方法として、RCTのメタアナリシスで用いられたのと同じ方法を採用することとした。また、アウトカムの統合に用いる指標として、統合オッズ比や統合リスク比などが使われるが、RCTのメタアナリシスで用いられたのと同じ指標を用いることとした。アウトカムの発生頻度が少なくなるにつれて、オッズ比とリスク比はほぼ同じ値になるので、統合値の比較の際はオッズ比とリスク比を同等に扱い比較した。信頼区間は95%として計算した。

もし、ひとつの群におけるアウトカムの発生が無い場合、これを2×2テーブルにおける「0セル」というが、この0セルは治療効果の比を計算する時に問題を生じる。この問題に対処する方法として、2×2テーブルのそれぞれのセルに0.5を加えることとした[15]。RCTの結果とOSの結果の違いを評価するために以下の解析をおこなうこととした。

1)それぞれのアウトカムにおいて、OSの統合オッズ比がRCTの統合オッズ比と2倍以上違うか。

2)同じトピックに関するRCTの統合オッズ比がOSの統合オッズ比よりも大きくなるのは偶然によるものか否かを下記の式で表すZスコアで評価した。

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

上記の式における $\ln(OR_{RCT})$ はRCT統合オッズ比の自然対数であり、 $\ln(OR_{OBS})$ はOS統合オッズ比の自然対数であり、varは分散を表す。Zスコアが1.96以上または-1.96以下の場合にはRCTとOSとの間に、偶然を超える違いが存在することを示唆する(有意水準0.05)[16]。

すべての統計解析はSTATA統計ソフトウェア8.1(Stata Corporation, College Station, TX, USA)を用いた。

### C. 結果

#### トピック、OSの特徴

まず、トピックを選定するためにRCTのメタアナリシスを検索し、1,184の文献が候補となった。図1に示す過程を経て、最終的に15編の消化器外科に関するRCTのメタアナリシスを同定した[7,17-30]。その15編のうち3編は2つのトピックを含んでいた。こうして、RCTとOSとを比較するための18

トピックが決定した(表1)。18トピックのうち10トピック(Topics 2,3,8-13,17,18)は、OSのメタアナリシスが存在しなかったため、われわれがこれを行うことにした。残りの8トピック(Topics 1,4-7,14-16)については、OSのメタアナリシスが存在したため、この結果を比較に使用することとした。OSのメタアナリシスが存在しない10トピックについて、メタアナリシスを行うために、まず文献検索を行ったところ、図2に示す過程を経て、10,960文献から111のOSを選出した。この111文献のうち17文献は英語または日本語以外の7つの言語で書かれていた。これらを翻訳家が英語に翻訳した後に、二人の研究者がデータを抽出した。本研究の中で、52のアウトカムについてRCTとOSとが比較可能であった。上記の過程を経て、18トピック52アウトカムを検討することになったのだが、これは101,170人の患者で構成する276文献(96のRCTと180のOS)のデータに基づくものである(表1)。180のOSは、36の前向き研究と144の後ろ向き研究により構成されている。

#### 研究間の結果の不均一性

われわれが本研究のために施行したOSのメタアナリシスに関しては(Topics 2,3,8-13,17,18)、研究間の不均一性を検証するためのQ検定に関するデータが利用できたが、残りのメタアナリシスについては記述がなかった。RCTでは、11トピック中1トピック(9%)に不均一性が認められた。OSでは、14トピック中7トピック(50%)に不均一性を認めた。

#### すべてのアウトカムの比較

18トピック52アウトカムの結果を表2に示す。DerSimonian-Laird法を用いた変量

効果モデル、Petoオッズ比法またはMantel-Haenszel法を用いた母数効果モデルという3つのタイプの計算モデルを使用した。52アウトカムのうち21アウトカムについては、RCTのメタアナリシスがリスク比を評価指標に用いていたため、OSのメタアナリシスでもリスク比を指標に用いることとした。

52アウトカム中9アウトカムでは、OSの統合値が、RCTの統合値の2倍以上あった。10アウトカムに関してはこの逆の結果だった。Zスコアによる評価では52アウトカム中4アウトカムでRCTの統合値とOSの統合値で不一致が認められた(トピック12における"reoperation"、トピック7における"mortality"、トピック16における"major complication"、トピック16における"conversion to laparotomy") ( $1.96 < |Z|$ ,  $\alpha < 0.05$ )。これらの結果から、OSはRCTに比べて治療効果を大きくあるいは小さく見積もりやすいということはなく、両デザインでの結果は同様であることを示唆している。

#### 主要アウトカムの比較

すべてのトピックについて、主要アウトカムは、もとのレビュー論文で定義されていたか我々のコンセンサスで決めたが、そのほとんどはmortalityであった。しかし比較的安全な手技を扱うトピック、例えば、虫垂切除や肛門裂孔などでは、創感染率や裂孔残存率などといった合併症の一つを主要アウトカムにするのが適切だと判断した。18トピックの中でわずか10トピックの主要アウトカムだけ2つ以上の研究を統合していた。これらの点推定値と95%信頼区間を図3に示す。10主要アウトカムにおいて、OSの95%信頼区間とRCTのそれとが重なっていないものは無かった。10主要アウトカムのうちの3主要アウトカムではOSの統合値はRCTの統合

値の2倍以上あった(トピック7,8,17)。その逆は1主要トピックで生じた(トピック11)。Zスコアによる評価では10主要アウトカム中1アウトカムでRCTの統合値とOSの統合値で不一致が認められた(トピック7、 $Z=-2.26$ )。研究間の不均一性を主要アウトカムだけで評価した場合、RCTでは、6主要トピック中1トピック(16.7%)に不均一性を認めた。OSでは、8主要トピック中5トピック(62.5%)に不均一性を認めた。

#### D. 考察

18トピックに関する276文献のデータにより、消化器外科領域のRCT統合値とOS統合値を比較した。研究間の不均一性はRCTに比べてOSの方が頻度が高かった(9%, 50%)。消化器外科のほとんどの領域ではRCTの統合値とOSの統合値は同様であり、違いがあったのは52アウトカム中4アウトカムだけであった。

RCTの結果とOSの結果はほぼ同一である、というわれわれの結論は、Bensonら[7]やConcatoら[8]による研究結果と似ている。Bensonらは19の多様な治療法に関して136文献を検討した結果、OSがRCTに比べて過大評価することはないという結論を得た。Concatoらは5つのトピックに関して99文献を検討した結果、良くデザインされたOSはRCTに比べて過大評価する傾向があるとはいえない、と結論した。70年代、80年代の評価と近年行われたIoannidisらによる研究では、逆の結論となっている[3-6,9]。2001年、Ioannidisらは45の多種多様な薬物治療や外科治療に関するトピックに関して、408の文献を検証し、OSはRCTに比べて治療効果を過大評価しやすく(28/45トピック vs. 11/45トピック)、OSは研究結果の不均

一性も高い(41% vs. 23%)と結論した。これらすべての先行研究は、研究デザインの違いから生じる結果の解釈について問題提起をおこなったという点で多大なる貢献をした。しかしながら、多種の臨床分野についての治療法を検討したものであるため、結論は一般論という形にならざるをえない。本研究では、消化器外科という一つの臨床分野に限定することで、先行研究にはない2つの利点があると考えている。一つ目は、多種の臨床分野を扱うよりも網羅的な検索が可能になるということ、もう一つは、得られた結論は一般論よりも臨床的な適応性が高いということである。

消化器外科領域において治療効果の統合値は両デザイン間に差がない、というわれわれの結論にはいくつかの要因が関与していると思われる。まず、外科RCTの研究の質は低いという報告がある[31,32]。つまり、外科RCTの質は低いため、実質的な内容としてはOSと変わらない試験になっているのかもしれない。また、ほとんどのトピックにおいては、臨床的に重要な差を除外するために、十分なデータが足りなかった可能性もある。

OSはRCTに比べて研究間の不均一性が高かった理由として、通常のOSはリスクに関して広いスペクトラムの患者を含んでいることがある。これに対してRCTは厳密な採用基準を設けており、臨床現場における患者を代表しているとは必ずしも言えない。

本研究であつかったトピックでは、A対Bという形での比較をおこなった。一般的には、Aは新しい手技をBは従来の方法をさすのだが、いくつかのトピックに関しては、どちらが新しいものなのか判断が困難である。薬物治療試験においては、薬物の効果を評価するのだが、消化器外科領域においては、術後死亡率、合併症率といった手術のリスクを評価することが多い。したがって、本研究の命題

も「消化器外科において OS は RCT に比べて治療効果を過大評価しやすいか？」ではなく「消化器外科において OS は RCT に比べて治療リスクを過小評価しやすいか？」に変更せざるをえなかった。そのために、RCT の統合値と OS の統合値の不一致性を評価する必要があった。例えば、2つの研究デザイン間で統合値の違いが最も大きかったのは、トピック7 (Mortality) であるが、これは食道癌手術における経胸的と経裂孔的食道摘出術を比較したものである ( $Z=-2.26$ )。このトピックにおいて OS の統合値は RCT の統合値の12倍もあったが (1.43 vs. 0.12)、これは治療効果ではなく、リスクが大きく評価されたものである。

#### E. 結論

OS は RCT に比べて、低コストで、研究に要する時間も短く、幅広い患者層を組み入れられる、などの利点がある[33]。本研究により、消化器外科領域では両研究デザインの結果はほぼ同じになるが、しかし OS は RCT に比べて研究間の不均一性が高いことが示唆された。したがって、各 OS 結果は、ただちに臨床現場で活用できるわけではないが、多数の OS の結果を集積し統合することで適応性は高まり、RCT 結果を解釈するうえでも有用な情報になると考える。

#### F. 参考文献

1. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ*. 1948;2:769-82.
2. Preventive Services Task Force. Guide to clinical preventive services: report of the U.S. Preventive Services Task Force. 2nd ed. Baltimore, USA: Williams &

Wilkins, 1996.

3. Chalmers TC, Matta RJ, Smith H Jr, Kunzler A-M. Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med*. 1977;297:1091-6.
4. Sacks HS, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med*. 1982; 72: 233-40.
5. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I. Medical. *Stat Med*. 1989;8:441-54.
6. Miller JN, Colditz GA, Mosteller F. How study design affects outcomes in comparisons of therapy. II. Surgical. *Stat Med*. 1989;8:455-66.
7. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000; 342:1878-86.
8. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000; 342:1887-92.
9. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001; 286:821-30.
10. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*. 1985;27:335-71.
11. Mantel N, Haenszel WH. Statistical

aspects of the analysis of data from retrospective studies of diseases. *J Natl Cancer Inst.* 1959;22:719-48.

**12.** Fleiss JL. *Statistical Methods for Rates and Proportions.* New York, USA: Wiley, 1981.

**13.** Fleiss JL. Analysis of data from multiclinic trials. *Control Clin Trials.* 1986;7:267-75.

**14.** DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986;7:177-88.

**15.** Matthias E, George DS, Douglas GA. *Systematic reviews in health care: meta-analysis in context.* 2nd ed. London, England: BMJ books, 2001.

**16.** Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons of meta-analyses and large trials. *JAMA.* 1998; 279: 1089-93.

**17.** Leiboff AR, Soroff HS. The treatment of generalized peritonitis by closed postoperative peritoneal lavage. A critical review of the literature. *Arch Surg.* 1987;122:1005-10.

**18.** 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;16:338-45.

**19.** Urbach DR, Kennedy ED, Cohen MM. Colon and rectal anastomoses do not require routine drainage: a systematic review and meta-analysis. *Ann Surg* 1999;229:174-80.

**20.** Nelson RL. Meta-analysis of operative techniques for fissure-in-ano.

*Dis Colon Rectum.* 1999;42:1424-8.

**21.** Hulscher JB, Tijssen JG, Obertop H, van Lanschot JJ. Transthoracic versus transhiatal resection for carcinoma of the esophagus: a meta-analysis. *Ann Thorac Surg.* 2001;72:306-13.

**22.** Urschel JD, Blewett CJ, Bennett WF, Miller JD, Young JE. Handsewn or stapled esophagogastric anastomoses after esophagectomy for cancer:

meta-analysis of randomized controlled trials. *Dis Esophagus* 2001;14:212-7.

**23.** Urschel JD, Urschel DM, Miller JD, Bennett WF, Young JE. A meta-analysis of randomized controlled trials of route of reconstruction after esophagectomy for cancer. *Am J Surg.* 2001; 182: 470-5.

**24.** Urschel JD, Blewett CJ, Young JE, Miller JD, Bennett WF. Pyloric drainage (pyloroplasty) or no drainage in gastric reconstruction after esophagectomy: a meta-analysis of randomized controlled trials. *Dig Surg.* 2002;19:160-4.

**25.** Singer MA, Nelson RL. Primary repair of penetrating colon injuries: a systematic review. *Dis Colon Rectum.* 2002;45:1579-87.

**26.** Lustosa SA, Matos D, Atallah AN, Castro AA. Stapled versus handsewn methods for colorectal anastomosis surgery: a systematic review of randomized controlled trials. *Sao Paulo Med J.* 2002;120:132-6.

**27.** Sutherland LM, Burchard AK, Matsuda K, et al. A systematic review of stapled hemorrhoidectomy. *Arch Surg.* 2002;137:1395-406.

- 28.** McCulloch P, Nita ME, Kazi H, Gama-Rodrigues J. Extended versus limited lymph nodes dissection technique for adenocarcinoma of the stomach. *Cochrane Database Syst Rev.* 2003(4):CD001964.
- 29.** Merlin TL, Hiller JE, Maddern GJ, Jamieson GG, Brown AR, Kolbe A. Systematic review of the safety and effectiveness of methods used to establish pneumoperitoneum in laparoscopic surgery. *Br J Surg.* 2003; 90:668-79.
- 30.** Papi C, Catarci M, D'Ambrosio L, et al. Timing of cholecystectomy for acute calculous cholecystitis: a meta-analysis. *Am J Gastroenterol.* 2004;99:147-55.
- 31.** Solomon MJ, McLeod RS. Surgery and the randomized controlled trial: past, present and future. *MJA.* 1998; 169: 380-3.
- 32.** McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomized trials in surgery: problems and possible solutions. *BMJ.* 2002;324:1448-51.
- 33.** Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. *J Clin Epidemiol.* 1989;42:481-9.

表 1. ランダム化比較試験と観察研究とを比較したトピック

ID No.	トピック	ランダム化比較試験				観察研究	
		メタアナリシス	研究数 (患者数)	意向性 / 撤回意向性研究	合計 (患者数)	比較可能な アウトカム数	
1	Closed postoperative peritoneal lavage vs. no lavage for generalized peritonitis	Leiboff et al. <sup>17</sup> 1987	4 (173)	2 / 6	8 (1,034)	1	
2	Spleno-renal shunt vs. endoscopic sclerotherapy in prevention of variceal rebleeding	Spina et al. <sup>18</sup> 1992	4 (310)	0 / 2	2 (344)	1	
3	Routine drainage vs. no drainage after elective colorectal surgery	Urbach et al. <sup>19</sup> 1999	4 (414)	0 / 5	5 (1,767)	4	
4	Anal stretch vs. sphincterotomy for fissure-in-ano	Nelson et al. <sup>20</sup> 1999	6 (328)	0 / 4	4 (537)	2	
5	Open vs. closed lateral sphincterotomy for fissure-in-ano	Nelson et al. <sup>20</sup> 1999	2 (140)	0 / 4	4 (1,365)	2	
6	Laparoscopic vs. open appendectomy for acute appendicitis	Benson et al. <sup>7</sup> 2000	16 (1,703)	3 / 4	7 (1,502)	1	
7	Trans thoracic vs. transhiatal resection for carcinoma of the esophagus	Hulscher et al. <sup>21</sup> 2001	3 (138)	3 / 18	21 (2,466)	6	
8	Handsewn vs. stapled esophago-gastric anastomosis after esophagectomy	Urschel et al. <sup>22</sup> 2001	5 (467)	2 / 8	10 (3,196)	3	
9	Posterior vs. anterior route of reconstruction after esophagectomy	Urschel et al. <sup>22</sup> 2001	6 (342)	0 / 3	3 (329)	4	
10	Pyloroplasty vs. no drainage in gastric reconstruction after esophagectomy	Urschel et al. <sup>24</sup> 2002	3 (347)	0 / 2	2 (111)	1	
11	Primary repair vs. fecal diversion for penetrating colon injuries	Singer et al. <sup>25</sup> 2002	5 (467)	4 / 29	33 (5,745)	4	
12	Stapled vs. handsewn methods for colorectal anastomosis surgery	Lustosa et al. <sup>26</sup> 2002	9 (1,233)	2 / 13	15 (3,894)	6	
13	Stapled vs. conventional hemorrhoidectomy	Sutherland et al. <sup>27</sup> 2002	7 (591)	2 / 5	7 (910)	3	
14	Extended vs. limited lymph node dissection for adenocarcinoma of the stomach	McCulloch et al. <sup>28</sup> 2003	3 (1,729)	5 / 8	13 (4,058)	2	
15	Open (Hasson type) vs. closed (needle/trocar) access in laparoscopic surgery	Merlin et al. <sup>29</sup> 2003	4 (302)	4 / 6	10 (20,664)	3	
16	Direct trocar vs. closed (needle/trocar) access in laparoscopic surgery	Merlin et al. <sup>29</sup> 2003	3 (665)	0 / 2	2 (1,575)	3	
17	Early vs. delayed open cholecystectomy for acute cholecystitis	PapI et al. <sup>30</sup> 2004	9 (916)	2 / 16	18 (37,475)	3	
18	Early vs. delayed laparoscopic cholecystectomy for acute cholecystitis	PapI et al. <sup>30</sup> 2004	3 (228)	7 / 9	16 (3,705)	3	
		合計	96 (10,493)		180 (90,677)	52	



表 2. すべてのアウトカムの統合値

ID No.	トピック	アウトカム	ランダム化比較試験		観察研究		計算
			研究数	オッズ比 (95%信頼区間)	研究数	オッズ比 (95%信頼区間)	
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	Spleno-renal 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	Trans thoracic 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-year survival†	1	1.83 (0.70 - 4.78)	8	1.44 (1.12 - 1.86)	M-H
8	Handsewn vs. stapled esophago-gastric 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

表 2. すべてのアウトカムの統合値(続き)

ID No.	トピック	アウトカム	ランダム化比較試験		観察研究		計算
			研究数	オッズ比 (95%信頼区間)	研究数	オッズ比 (95%信頼区間)	
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
		Intra-abdominal 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. handsewn 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
13	Stapled vs. conventional hemorrhoidectomy	Wound infection	6	1.43 (0.67 - 3.04)	5	1.28 (0.83 - 1.97)	Peto
		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 weeks)†	2	1.07 (0.36 - 3.17)	3	0.55 (0.16 - 1.83)	M-H
		Mortality	2	2.39 (1.50 - 3.82)	2	0.63 (0.43 - 0.93)	M-H
		5-year survival	2	0.92 (0.72 - 1.17)	2	1.17 (0.97 - 1.42)	M-H
14	Extended vs. limited lymph node dissection for adenocarcinoma of the stomach	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
15	Open (Hasson type) vs. closed (needle/trocar) access in laparoscopic surgery	Conversion to laparotomy	2	0.32 (0.05 - 1.96)	4	0.43 (0.16 - 1.21)	M-H



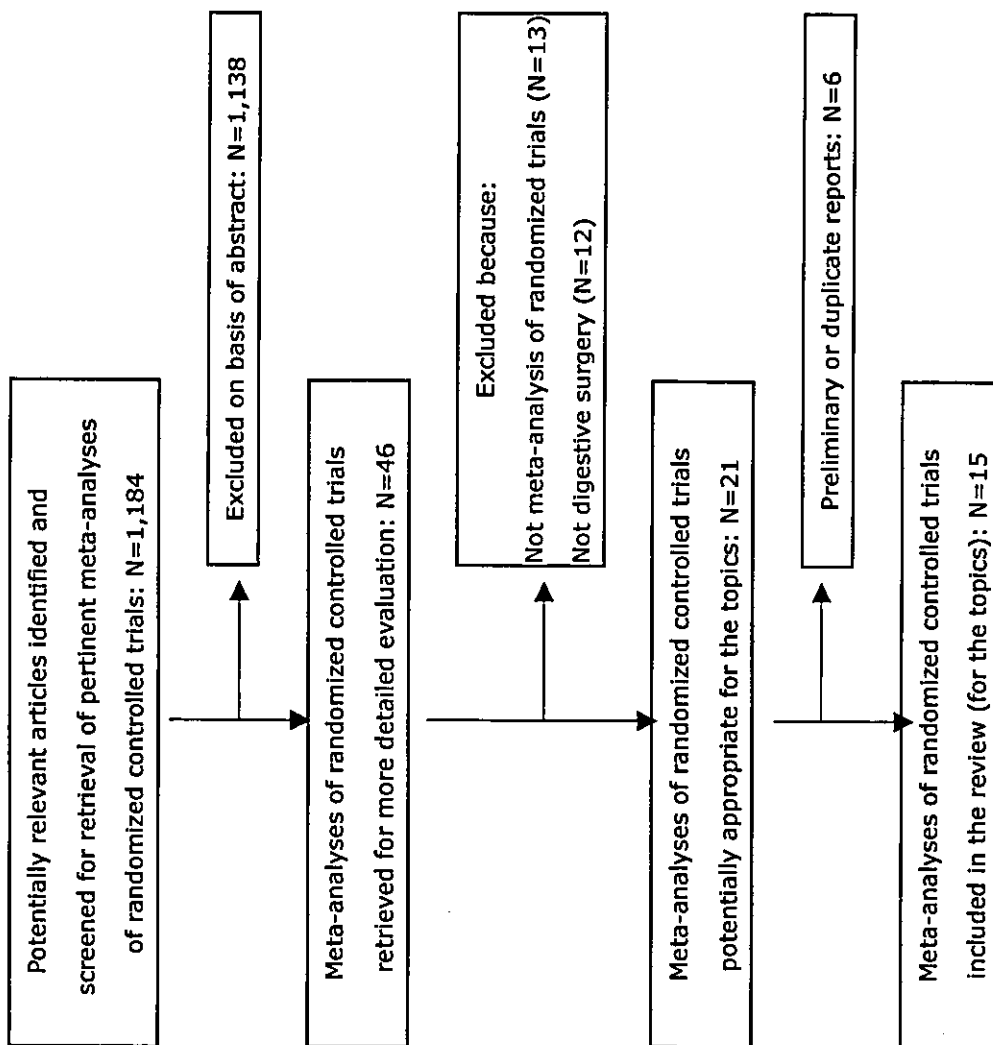


図 1. ランダム化比較試験のメタアナリシスの検索過程

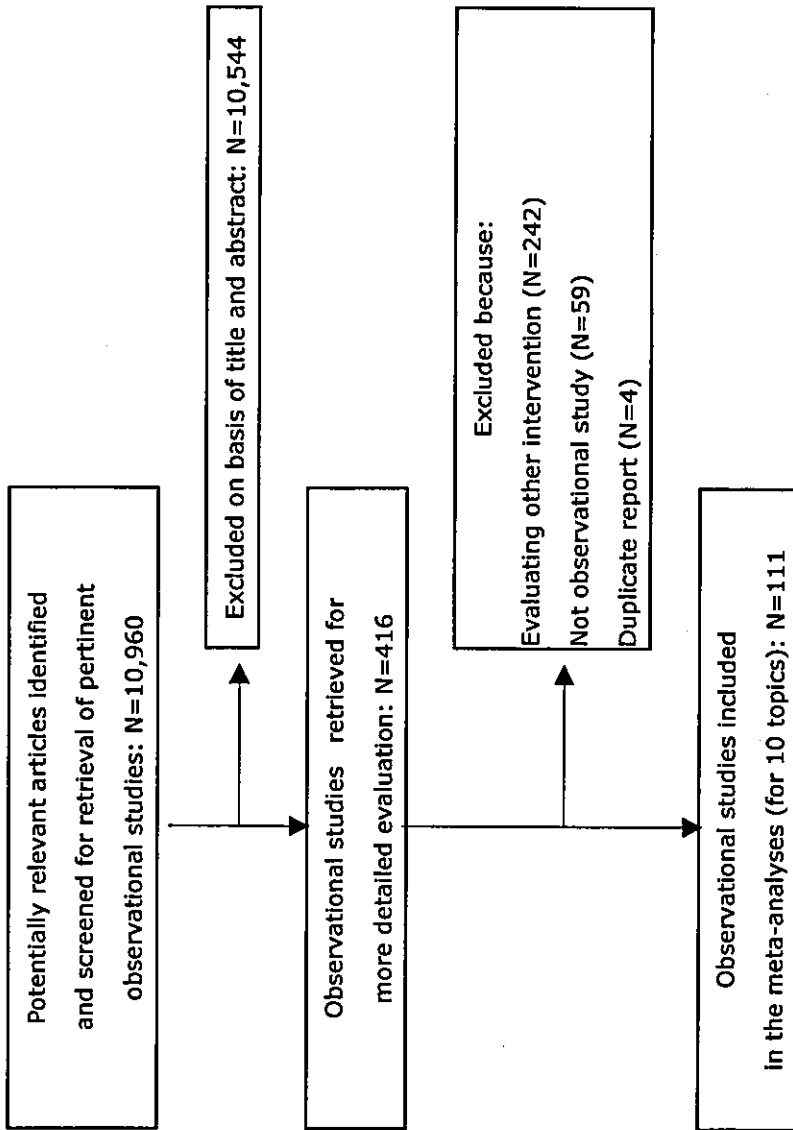


図 2. 観察研究の検索過程

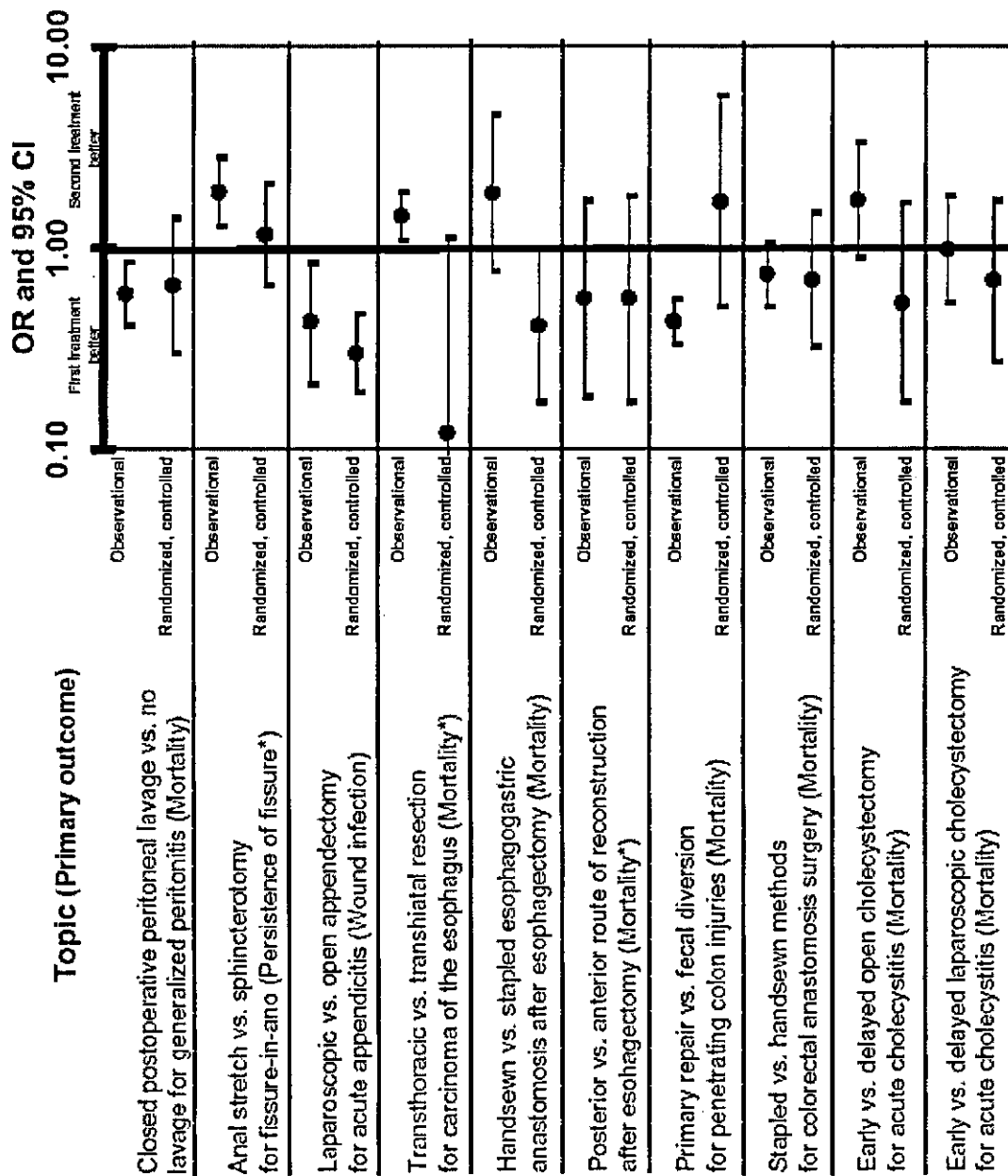


図 3. ランダム化比較試験と観察研究、主要アウトカムの比較

# 日本から発表されたランダム化比較試験の論文報告の質 : CONSORT 声明に基づく評価

研究協力者 植谷 可恵 (京都大学大学院医学研究科健康情報学分野)  
主任研究者 中山 健夫 (同 助教授)

## 研究要旨

【背景】ランダム化比較試験 randomized controlled trial (RCT) は、予防や治療の効果を検証する際に科学的な妥当性を最も高く保証しうる研究デザインである。RCT の情報は単独の成果としてのみでなく二次研究への利用価値が認識されつつある。RCT の報告において、結果の解釈に必要な項目の記載がなければ、読者による妥当性の判断ができず情報の利用が困難となる。CONSORT (Consolidated Standards of Reporting Trials) 声明は RCT の報告の質の改善を目的として発表された、報告に記載すべき項目を示すチェックリストである。CONSORT 声明は論文執筆や査読に活用できる国際的な指針であり、現在では 150 以上の医学雑誌で投稿規定に採用されている。また、国外では CONSORT 声明を用いて報告の質の評価が多く行われているが、研究対象は欧米の論文が中心である。日本で実施された RCT の報告についても世界中で利用される情報として高い質を確保すべきであるが、その現状は把握されていない。

【目的】日本の RCT の報告に対して CONSORT 声明に基づいた評価を行い、報告の質に関する現状と課題を明らかにする。また、報告の質を規定する要因として、雑誌の CONSORT 声明の支持(投稿規程への採用)の有無と報告の質との関係を調査する。

【方法】記述的な文献調査研究である。対象は日本で実施された RCT の報告で、2004 年 1 月から 3 ヶ月間に掲載された全ての論文を MEDLINE より選出した。調査項目は、① CONSORT 声明に基づく報告の質の評価 32 項目、②本研究のために追加した 3 項目(倫理審査、インフォームド Consent、利害の衝突)であり、各項目の記載有無を調査した。

① CONSORT 声明に基づく報告の質の評価: 主要評価項目は全 32 項目中の総記載項目数、副次的評価項目は各項目の記載論文数とした。解析は、全評価対象論文の結果を記述統計により示した。また、論文を掲載した雑誌の CONSORT 声明の支持の有無によるクロス集計を行った。

②追加した 3 項目: 全評価対象論文中の記載論文数を示し、副次的評価項目とした。

【結果】評価対象論文として 100 報が得られた。

① CONSORT 声明に基づく報告の質の評価: 全 32 項目中の総記載項目数は中央値が 15 項目であった。CONSORT 声明の支持ありの論文数は 10 報、支持なしの論文数は 90 報であった。総記載項目数の中央値は支持ありの論文で 16.5 項目、支持なしの論文で 15 項目であった。

②追加した 3 項目: 全評価対象論文中の記載論文数は、倫理審査 83 報、インフォームド Consent 92 報、利害の衝突 19 報であった。

【考察】日本の RCT の報告には CONSORT 声明の項目を記載していないものが多く存在することが示され、報告の質の向上が必要である。世界中で利用されるような臨床試験の成果を日本から発表し、研究実施の説明責任を果たすためには、透明で明確な報告を行うことが望まれる。CONSORT 声明の支持による報告の質の違いは大きくなかったことから、CONSORT 声明を投稿規定において推奨するような雑誌側の取り組みだけでは充分とはいえない。現在の状況を改善するためには、研究成果の論文執筆を行う際に CONSORT 声明のような国際的な指針を理解し、報告に反映することが日本において必要である。その基盤として、臨床試験の方法論の教育や、専門職(臨床疫学、生物統計学、メディカルライティング等)との協力が実用的な対策になると考える。

## 1. 緒言

### 1.1. 背景

ランダム化比較試験 randomized controlled trial (RCT) の成果が社会健康医学 (Public Health) の向上に活用されるためには、質の高い試験が実施されるだけでなく、質の高い報告がされなければならない。RCT の成果を把握するためには、その実施内容や結果の解釈を理解しなければならず、これは論文報告における完全な透明性によってのみ達成されるからである<sup>1</sup>。EBM (evidence-based medicine) の普及に伴い、RCT は単独の試験としてのみでなくシステマティックレビューなどの二次研究における利用価値が認識されることとなった。一方で、その情報を批判的に吟味する重要性に対する認識も高まりつつある。RCT は治療や予防の効果を検証する際に科学的な妥当性を最も高く保証しうる研究デザインであり、利用可能な場合には治療や予防を評価するための適切な方法とされている。しかし、試験の結果を真の値から系統的に乖離させるバイアスの混入は、RCT を実施することで全て排除されるわけではない。RCT の計画、実施、報告、情報の利用といった各段階においてバイアスは生じる<sup>2</sup>。従って、RCT の情報を利用する場合には提示された結果を理解するだけでなく、結果をみちびいた試験の実施に関わる情報を取得し、その妥当性を判断しなければならない<sup>1</sup>。

RCT の論文報告を研究対象とし、その成果を評価する試みは古くから行われてきた。国外では RCT を論文から評価するための様々なツールが開発され<sup>3</sup>、RCT の評価研究がなされている。しかしながら、試験の内容を読み解くための情報が論文上に記載されていない場合には、論文から RCT を評価することは困難である。この限界から、「報告の質」と「試験の質」は明確に区別<sup>4</sup>され「報告の質」に焦点をあてた研究が多くなされるようになった。しかし、日本の論文を対象とした研究は試験の実施内容の調査に限られている<sup>5</sup>。近年では、多くの RCT が日本でも行われている<sup>6</sup>ことから、その報告の質を評価し、情報としての利用可能性を高めるための取り組みが必要な時期にある。

CONSORT (Consolidated Standards of Reporting Trials) 声明<sup>7</sup>は、RCT の報告の質の改善を目的として発表された (1996 年 8 月初版, 2001 年 4 月改訂)、報告に記載すべき項目を示すチェックリストである。「報告の質の改善」とは、読者が試験結果を解釈するために必要な全ての情報を論文中から見出せることを保証することであ

る。CONSORT 声明は論文執筆や査読の際に活用できる指針であり、臨床試験実施者 (clinical trialist)、疫学者、統計学者、医学雑誌編集者らで構成される国際的なグループ (CONSORT グループ<sup>8</sup>) により作成され、現在も追補資料の発表<sup>10</sup>などの発展を続けている。CONSORT 声明には学会、医学雑誌編集者団体、企業等、各方面からの支持が集まっている<sup>9</sup>。特に、この声明に賛同する医学雑誌は増加し、現在では主要医学雑誌を含む 150 以上の海外の雑誌で投稿規定に採用されているが、日本の医学雑誌の採用は 3 誌のみにとどまっている<sup>9</sup>。CONSORT 声明に主張されるように、RCT を報告する際には試験結果の妥当性を判断するのに十分な情報を提示することで、その成果が利用可能な情報となるための準備が整えられる。

CONSORT 声明の発表以降、この指針との適合性に基づいて RCT の報告の質の評価が多くされるようになった<sup>11-14</sup>。これらの研究では CONSORT 声明に含まれる項目の全部もしくは一部を用いて、論文中の記載との項目の一致を評価し、その適合性をもって報告の質としている。これらの研究成果は国外で実施され、研究対象は欧米の雑誌や論文が中心である。また、地域ごとの比較を行った研究は PubMed で調査した限りは得られなかった。さらに、先行研究の多くは、特定の疾患領域や医学雑誌を限定して調査対象としている。雑誌や領域を広く対象とした研究では、二次情報雑誌 3 誌に掲載された内科領域の元論文についての評価がある<sup>11</sup>。評価の結果、CONSORT 声明を投稿規定に採用している雑誌の論文は採用していない雑誌の論文より CONSORT 声明の適合性が高く、報告の質が高いことが示された。また、主要医学雑誌 4 誌の RCT 論文の記載項目を評価し、CONSORT 声明の発表前後を比較して報告の質の向上を示した研究成果がある<sup>12</sup>。この研究では投稿規程への CONSORT 声明の採用の有無では報告の質の差が示されなかった。以上のように、国外では CONSORT 声明を用いて RCT の報告の質を向上させるための啓発が積極的に行われている。日本においても報告の質の改善にむけた取り組みを行い、実施される RCT の成果が有効に活用されることが必要である。

### 1.2. 本研究の目的と位置づけ

本研究の目的は、日本国内で実施され、日本人によって発表された RCT の論文報告に対して CONSORT 声明に基づいた適合性判定による評



価を行い、報告の質に関する現状を明らかにすることで。また、報告の質を規定する要因として投稿規程との関係を調べる。

研究全体の展望は、日本の論文報告の質に関する現状の把握を行い、問題点を明らかにした上で、今後の情報基盤の整備、具体的には論文報告の質向上にむけた課題についての提言を行う。そのためにまず、RCTの論文について日本と国外の報告の質を比較する研究を計画しており、その前段階として論文報告の質の評価方法と国内の実態調査の結果を示すことが本研究の位置づけである。疾患領域を限らず、特定の国で実施されたRCTの報告の質を評価する試みは本研究において初めてであり、日本特有の状況や問題が明らかになる可能性がある。

## 2 方法

### 2.1. 調査対象論文

調査対象は医学文献データベースMEDLINEに収録されている論文とし、日本国内で実施されたRCTの報告とした。選択基準は、1)臨床医学研究、2)原著論文(短報や総説論文は除く)3)RCT(介入群の割付けをランダムに行った旨の記述がある)、4)日本国内で実施された試験、5)記載言語が日本語か英語、6)出版日が2004年1月1日から3月31日(3ヵ月間)とした。出版期間は、多様な出版形態(毎月、隔月、季刊等)が含まれることと調査の量的な実施可能性から、100報程度が見込まれる3ヵ月間が妥当と判断した。また年月は最近の状況を調査する目的より決定した。除外基準は、1)既に報告されたRCTの再解析の論文、2)RCTの計画や予定の論文とした。検索は、日常的にPubMedのタグを活用した単純な検索結果がよく利用されることを想定し、タグによるアドレス(日本)、パブリケーションタイプ(RCT)、研究対象(ヒト)および出版日を特定した検索式「Japan[ad] AND Randomized Controlled Trial[ptyp] AND "humans"[MeSH Terms] AND ("2004/01/01"[PDAT]: "2004/03/31"[PDAT])」を使用した。他のデータベースや検索式の利用、ハンドサーチ等による論文の抽出は実施しなかった。調査期間中に新たにMEDLINEに追加された論文で、評価対象の基準をみたすものについては随時、評価対象論文に追加した。

### 2.2. 調査項目

調査項目は、CONSORT声明を基にした報告の質評価(32項目)のほか、独自に追加した調査項目(3項目)、および背景要因とした。

CONSORT声明は22項目からなるチェックリストであるが、先行研究<sup>13</sup>では複合的な項目を分割し32項目としたことに倣い、本研究でも同じ項目を分割した。分割した項目は9項目あり、「参加者」「アウトカム」「症例数」「順番の作成」「ブラインド化」「統計学的手法」「参加者の流れ」「解析された人数」の8項目は2項目に分割し16項目とした。「介入」については3項目に分割した。残りの13項目は分割を行わず、総項目数を32項目とした。

独自に追加した3項目は「倫理委員会の承認」「インフォームドコンセント」「利害の衝突」である。CONSORT声明はRCTの内的および外的妥当性に関わる項目を中心とする<sup>7</sup>ことから、これらの項目は含まないとされるが、いずれもRCTの実施状況に関わる情報であり、その重要性はCONSORT声明の解説<sup>8</sup>においても記されている。倫理委員会の承認とインフォームドコンセントは臨床試験を行う際に必ず実施されるべき項目であるため、論文への記載状況を調査した。「利害の衝突」は資金提供者など研究に関連した利害関係の記載の有無を判定する項目である。企業が出資した研究では企業の利益になる結果のみが報告される傾向にある<sup>15</sup>という出版バイアスの問題から、近年では本項目の記載が重要視されていることから調査した。

対象論文の背景要因として、雑誌(雑誌名、発行国、CONSORT声明の支持)、論文(報告頁数、構造化抄録の有無と項目数)、RCT(比較群数、介入の種類、参加者数、RCTのデザイン)について調査した。CONSORT声明の支持については、公式ウェブサイト<sup>9</sup>の支持雑誌一覧を調査した。支持雑誌とは、雑誌の投稿規程にCONSORT声明を採用したことを公式に表明している雑誌であり、「支持あり」の論文とは支持雑誌に掲載された論文を意味する。

主要評価項目はCONSORT声明(32項目)の総記載項目数とした。副次的評価項目はCONSORT声明の各見出し(方法、結果、考察)の記載項目数、CONSORT声明の各項目および追加3項目の記載論文数とした。

### 2.3. 評価基準

CONSORT声明に基づく報告の質の評価(32項目)については、CONSORT声明および解説書<sup>1,7,8</sup>をもとに本研究における独自の評価基準を項目毎に定め、文章化した評価用マニュアルを作成した。基準設定の際には複数担当者による協議を重ね、評価基準の統一をはかった。評価は基本的に各項目について記載の有無を単純に判

定するものとした。CONSORT 声明中で 1 項目に対し複数の推奨をしている場合は、どのような記載があれば「あり」とするかを基準として定め明記した。各項目に対する記載の有無を評価基準としたため、記載内容が試験と照らして適切かどうかという判定や、前後の記述との整合性についての判定は行わなかった。また、記述量の多寡に対する一様な判定は不可能と判断し、評価基準には含めなかった。

#### 2.4. 評価手順

評価については、一次評価、QC (Quality Control), QA (Quality Assurance) の手順を全論文に対し行う。段階ごとに独立した担当者に関わることとし、担当者は全員、臨床試験の方法論および CONSORT 声明に関する知識をもつ者とした。開始前には、各担当者は評価基準について十分に理解した。評価に際してはまず、評価担当者が一次評価を行った。次に QC 担当者が一次評価結果の記録をもとに、評価の妥当性を全論文について確認する。最後に QA 担当者が、QC 担当者による確認結果が適切に最終評価に反映されたことを確認する。

#### 2.5. 統計解析

全評価対象論文について解析を行った。背景要因の各項目について論文数および構成割合 (%) を求め、論文の報告頁数と構造化抄録の項目数は中央値、最小値、最大値を示した。RCT の参加者数は最小値と最大値を求め、50 名ごとにカテゴリー化し論文数および構成割合 (%) を求めた。CONSORT 声明の総記載項目数と見出しごとの記載項目数は平均値、標準偏差、中央値、最小値、最大値を示した。

CONSORT 声明の各項目は、記載論文数および構成割合 (%) を示した。全評価対象論文の解析結果に併せて、CONSORT 声明の支持の有無によるクロス集計を解析計画通り行った。追加 3 項

目は、全評価対象論文中の記載論文数を示した。

### 3. 結果

#### 3.1. 調査状況と背景要因

対象論文選定の流れについて図 1 に示す。PubMed 検索の結果、133 報が得られたが、33 報は適格基準を満たさず除外となり、対象論文 100 報の調査を行った。調査は現在進行中で、本報告は担当者 1 名による評価の解析結果をもとに行った (2005 年 1 月 10 日)。なお、対象論文の本文記載言語は全て英語であった。対象論文の背景を表 1 に示す。CONSORT 声明の支持ありの論文は 10 報、支持なしの論文は 90 報であった。日本国内で実施された 100 試験の RCT の報告が得られた。

#### 3.2 評価結果

CONSORT 声明の全項目 (32 項目) および各見出しと各項目の評価について、全評価対象論文 (100 報) と支持の有無によるクロス集計の結果を表 2 に示す。CONSORT 声明の適合性に基づく報告の質の評価の結果、主要評価項目である全 32 項目中の総記載項目数の中央値は、全評価対象論文で 15 項目、支持ありの論文では 16.5 項目、支持なしの論文では 15 項目であった。追加した 3 項目について、全評価対象論文中の記載論文数は、倫理委員会 83 報、インフォームドコンセント 92 報、利害の衝突 19 報であった。

### 4. 考察

#### 4.1. 主要評価項目

日本で実施された RCT の論文には CONSORT 声明の適合性が低い報告が多く存在することが示された。全 32 項目中の総記載項目数は平均値 15.6 項目、中央値 15 項目であり全項目の 50% に満たない。主要医学雑誌 4 誌の評価<sup>12</sup> では、1994 年の RCT 論文で平均 55% (CONSORT 声明を 40 項目に分割した中の 22 項目) 以上の項目の記載があった。1994 年は初版 CONSORT 声明の発表以前であり、1998 年の論文を対象とした評価では 4 雑誌全てが質の向上を示した<sup>12</sup>。また、小児代替医療領域の RCT 論文を評価した研究でも、10 年ごとの比較の結果、時間の経過とともに CONSORT 声明の適合性が向上したことが示された<sup>13</sup>。CONSORT 声明が広範に普及し報告に反映されるには時間を要する<sup>13</sup>

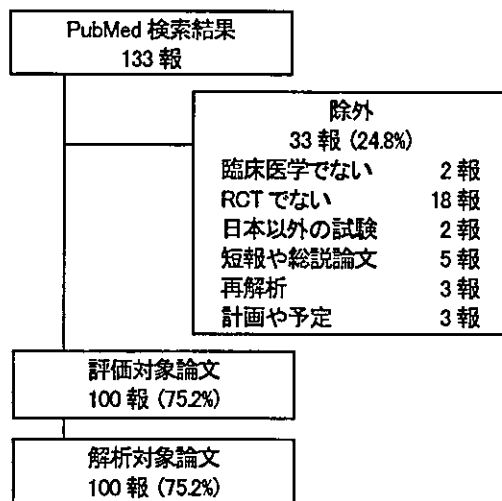


図 1 対象論文選定の流れ

が、1996年の初版 CONSORT 声明発表から8年が経過したこと、その間に CONSORT グループを中心とする積極的な取り組みが続けられている<sup>10</sup>こともあり、RCTの報告の質は年月とともに向上する傾向にある。しかし、本研究の結果からは2004年初頭における日本のRCT論文の報告の質は高いとはいえず、改善の余地が大いにありと考える。

#### 4.2. 雑誌の投稿規程と報告の質との関係

クロス集計の結果、CONSORT 声明の支持ありと支持なしとの間に報告の質の大きな違いはみられず、投稿規程による指示にも関わらず記載すべき項目がない論文が高い割合で存在することが示された。投稿規程による違いが大きい理由として2点を考える。まず日本において CONSORT 声明の意義と内容が十分に理解されておらず、この指針が投稿規程にあっても論文執筆の際に十分に活用されていない可能性がある。次に投稿規定以外の要因によって報告の質が規定されている可能性である。後者については、他の要因(疾患領域、雑誌のインパクトファクター等)を検討する必要がある。ただし、支持ありにおける総記載項目数の最小値(12項目)が支持なし(6項目)より明らかに大きいことや平均値では3.4項目の差があることから、投稿規程による報告の質の改善傾向はみられる。平均値にみられた差は主に、支持ありで少数の論文が示した高い記載項目数(30項目1報、22項目3報)が反映されたと考える。支持ありは対象論文数が10報と少なかったため、少数の高値が平均値に大きく影響したことがわかる。このような高い記載項目数の論文には、執筆に際し CONSORT 声明を遵守する姿勢が明確にあらわれている。また、支持ありでも記載項目数の低い論文が多くみられたことから、CONSORT 声明が雑誌の投稿規程に採用されていても、その推奨内容を論文の記載に反映させるような強制力が大きく働いていない場合があることがわかった。

#### 4.3. CONSORT 声明の各項目の記載状況

全評価対象論文における各項目の記載について、「方法(18項目)」「結果(9項目)」はRCTの内容を理解する上で重要な項目を多く含むが、記載のない報告が多く大幅な改善が必要である。「方法」で特に記載の不備が目立つ項目は「主要・副次的アウトカム評価尺度の定義」「目標症例数の設定方法」で、23報ずつの記載であった。「結果」で特に低い項目は「介入のエフェクトサイズとその精度(95%信頼区間)」である。CONSORT 声明は、

試験の結果はエフェクトサイズとして群間の対比(リスク比や平均値の差等)と、精度として95%信頼区間を示すことを推奨している。しかし、両方の記載があるのは100論文中8報のみで、大多数ではアウトカムに対し群間または各群の前後で比較し検定を行い、統計的な有意性についてのみ言及していた。また、参加者の数に対する解析上の取り扱いに関する項目「参加者の流れ(29報)」「プロトコル逸脱の人数と理由(44報)」「各解析の各群の参加者数(54報)」「ITT解析かどうか(58報)」の記載がない論文も多くみられた。以上の項目はいずれも統計解析と関係が深いことから、これらの記載が少ない背景として日本では試験実施や結果の報告に必要な生物統計学の基盤の脆さがあるのではないかと考える。

全32項目のうち、上記以外で特に記載論文数が少なかった項目はまず、「試験の中間解析と中止基準の説明(7報)」「サブグループ解析や調整解析の手法(12報)」「サブグループ解析や調整解析(3報)」であったが、これらは実際の試験で実施がない場合に記載がされにくい項目である。しかし、CONSORT 声明では「(計画あるいは実施を)しなかった」旨の明記を推奨している。同様に、「割付けの制限の詳細(22報)」については、ランダム割付けの際に層別化などの制限がある場合は、層別に用いた要因や割付け比の記載を指示しているが、制限がない場合も「単純ランダム化」や「1対1の割付け比」等の明記を推奨しており、CONSORT 声明の遵守の意識が高まることで、これらの項目の記載状況は改善すると考える。さらに、記載が全くなかった項目「ブラインド化の成功をいかに評価したか(0報)」は、主要医学雑誌を対象とした調査<sup>16</sup>でも一般医学論文の7%しかブラインド化の成功の評価がなかったとされ、今後の努力を要する項目である。

#### 4.4. 追加3項目の記載状況

追加した3項目について、倫理委員会とインフォームドコンセントは多くの論文にあったが、本来は全ての論文に記載されるべきである。利害の衝突として研究の資金源等を記載している論文は100報中19報であり、国外の雑誌を対象とした研究では56%に記載があった<sup>14</sup>ことと比較すると日本ではまだ記載への意識が低いことがわかる。

#### 4.5. 本研究の限界

CONSORT 声明に基づいた報告の質評価研究においては、CONSORT 声明の項目を評価用の項目に改変して用いている。CONSORT 声明はRCTの報告を執筆や査読を行う際の指針であり、

質評価のためのツールとすることを意図したものではなく<sup>8</sup>、このことから生じる限界がある。まず、「報告の質」を記載の有無のみで測ることに潜在的な限界がある。利用可能な情報としての質を論じるには、記載内容の妥当性や適切さなどに対する複合的な評価を行うことが必要であり、記載の有無の判定は、その内容の質を保証するものではない。また、総記載項目数を評価指標とすることは、各項目の重要性が同じでないことから適切とはいえない。この点については、本研究では日本の現状を把握する目的に対しては十分な指標となりうると判断して用いることとした。さらに、項目ごとに定めた評価基準は本研究における独自の判定基準であり、他の研究と一致した基準で評価しているとは限らず、先行研究と比較した結果の解釈には注意を要する。

CONSORT 声明を用いるもう一つの限界として、研究目的や RCT のデザインによっては評価対象として適さない項目が含まれている可能性がある。例えば、薬物動態の検討のように、効果を検証する目的以外の RCT の報告に対する CONSORT 声明の適用可能性は確立されていない<sup>16</sup>。また、サブグループ解析や調整解析に関する項目も、本来それらの解析を要しない RCT もあると考えられる。しかし、記載の要、不要の判断が困難な場合もあり、本研究ではその区別を行わず記載の有無を評価した。CONSORT 声明の全項目が適用できないような RCT が含まれる場合は、報告の質は適切に評価されていない可能性がある。この問題への対応には、RCT の研究目的や仮説、介入などの種類と CONSORT 声明の適合性に関する検討を進めることが必要で、場合によっては CONSORT 声明とは別の指針が必要かもしれない。

#### 4.6. 結論と提言

日本で実施された RCT の論文報告の質は高いとはいえず、利用可能な情報となるには質の向上が必要である。現時点では担当者 1 名による評価が終了した段階ではあるが、予定されている QC 及び QA を経たのちには、より信頼性の高い結果を得ることと方法の妥当性の確認を見込んでいる。本研究の「報告の質」とは、「結果を解釈するために必要な項目の記載が論文紙上にある」という限られた定義である。しかしながら、疾患領域や雑誌を限定することなく、一つの地域(国)から発表された RCT の報告の質を CONSORT 声明に基づいて評価する試みは初めてであり、本研究の意義は RCT の「報告の質」という観点から日本の現

状を把握する方法と結果を示したことにある。

現在の状況を改善するためには、CONSORT 声明を投稿規定に採用するような雑誌側の取り組みが必要であるが、本研究の結果より投稿規程による差は大きくなかったことから、これだけで充分とはいえない。研究成果の論文執筆を行う際に CONSORT 声明のような国際的な指針を理解し、報告に反映することが日本において必要である。その普及には、臨床疫学や生物統計学の知識を基盤とする臨床試験の方法論の教育が不可欠であり、専門家の養成や活動の場が必要である。試験実施に際しては、RCT の報告に必要な項目の多くは計画中に含まれていることが前提条件であり、計画段階から CONSORT 声明が求めている内容を理解して臨むべきである。実務としては、臨床試験の実施体制として生物統計学分野の協力が不可欠であるのと同様に、メディカルライティング分野の教育や専門職との共同作業も報告の質向上にむけた実用的な対策となるであろう。世界中で利用されるような臨床試験の成果を日本から発表し、研究実施の説明責任を果たすためにも、高い質の試験を計画、実施した上で、透明で明確な報告を行うことが望まれる。

#### 5. 参考文献

- 1) Moher D, Schulz KF, Altman D; CONSORT Group (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285(15):1987-91.
- 2) Jadad AR. *Randomised Controlled Trials: A User's Guide*. Canada: BMJ Books; 1998
- 3) Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16(1):62-73.
- 4) Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol*. 2002;55(8):783-6.
- 5) 金子善博, 津谷喜一郎, 中山健夫, 宇山久美子, 宮野昌明, 山崎茂明, 他. 2000 年に日本で報告された RCT の内容. *臨床薬理*. 2002;33(2):223S-4S.
- 6) 津谷喜一郎, 金子善博, 中山健夫, 宇山久美子, 宮野昌明, 山崎茂明, 他. 日本では