



Published in final edited form as:

J Am Coll Surg. 2012 November ; 215(5): 722–730. doi:10.1016/j.jamcollsurg.2012.06.018.

Do Surgical Trials Meet the Scientific Standards for Clinical Trials?

Danielle M. Wenner, PhD^{1,2}, Baruch A. Brody, PhD^{3,2}, Anna F. Jarman, BA¹, Jacob M. Kolman, MA^{2,3}, Nelda P. Wray, MD MPH¹, and Carol M. Ashton, MD MPH¹

¹The Methodist Hospital Research Institute and Department of Surgery, The Methodist Hospital, 6550 Fannin Street, Houston, TX 77030

²Rice University Department of Philosophy, 6100 Main MS-14, Houston TX 77005-1827

³Center for Ethics, Medicine, and Public Policy, One Baylor Plaza, Houston, TX 77030

Unlike medications, the dissemination of surgical procedures into practice is not regulated. Before marketing, pharmaceutical products are required to be shown safe and efficacious in comparative clinical trials which utilize bias-reducing strategies designed to reduce the distortion of estimates of treatment effect by predispositions towards the investigational intervention or control. Unless an investigational device is involved, the corresponding process for surgical innovations is usually unregulated and therefore may not be based upon adequate evidence. Given these differences, we sought to evaluate the state of clinical research on invasive procedures.

We conducted a systematic review of publications from 1999–2008 that reported the results of studies evaluating the effects of invasive therapeutic procedures, focusing on trials which appeared to influence practice. Our objective was to determine what proportion of studies evaluating surgical procedures use a comparative clinical trial design and methods to control bias. This paper reports our results and raises concerns about the methodological, and therefore the ethical, quality of clinical research used to justify the implementation of surgical procedures into practice.

Methods

After reviewing the National Library of Medicine's list of medical subject headings and eliminating those for which surgery is not a viable treatment option, 36 non-rare medical conditions were identified for which a surgical treatment modality exists. We systematically searched these headings (subsequently consolidated to 33) for comparative clinical trials investigating the effectiveness of a surgical or minimally invasive procedure. Studies from a broad range of specialties were sought.

We conducted an electronic search of MEDLINE from January 1999 through December 2008 for English-language reports of comparative clinical trials of invasive therapeutic

© 2012 American College of Surgeons. Published by Elsevier Inc. All rights reserved.

Address correspondence to Danielle M. Wenner, Department of Bioethics, 9500 Euclid Ave., JJ60, Cleveland, OH 44195; telephone 216-445-1015; FAX 216-636-0712; wennerd2@ccf.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosure information: Nothing to disclose.

procedures using a modified version of the Cochrane Highly Sensitive Search Strategy (HSS).¹ Search terms were tested at each stage of an iterative revision of the HSS by reviewing a random sampling of the results of a test search for precision and sensitivity. See Table 1 for the final search algorithm used. Bibliographies of metaanalyses retrieved by the search were hand-reviewed for potential eligible studies.

All search results were reviewed for eligibility separately by one of four research assistants and one of two investigators (DMW or CMA) via a stepwise process. Studies were eliminated based on title, abstract, or a full review of the article text. Studies were excluded which evaluated perioperative adjuncts (medications, types of anesthesia, etc.) or minor technical elements of surgical procedures. Inclusion criteria were minimal and required at least two study arms (comparative trial), 30 participants, a description of allocation method within the report, and trial endpoints with discernible clinical significance to the participant. Disagreements regarding inclusion were resolved through discussion.

We tried to identify those trials which have had the greatest impact on clinical practice using an original unvalidated measure based on citation analysis of trial reports. Citation searches were conducted for each publication using the Web of Science database² to identify scholarly citations, and the news search function of the LexisNexis database³ to identify citations in mainstream news sources. Within each subject heading, the five reports with the most citations per year since publication were included in the review pool. For any heading with 5 eligible reports, all were included (Influence Criterion 1). For reports published within the last two years of the inclusion period, those which had 10 citations per year in the LexisNexis search were included (Influence Criterion 2), to correct for any bias in selection arising from temporal constraints of research dissemination and to retrieve trials most representative of current levels of rigor. Neither journal quality nor impact factor were explicitly taken into account by the influence criteria, although it is probable that such factors informed the results as studies published in higher-impact journals are likely to have been cited more frequently. Any report which met either influence criterion was included in our final pool for review.

A standardized review was then conducted on each eligible report that met influence criteria to capture key descriptive elements of the trial and to collect information on attributes known to bolster internal validity – the confidence with which we can draw a conclusion about the difference between treatments based on study results – and those needed to assess the generalizability of study results, or external validity.⁴ Responses were limited to yes/no, true/false, and multiple choice elements in order to limit variability and ensure consistency across reviewers. Practice reviews were conducted by all six investigators on reports of ineligible trials to assess and correct any problems with inter-rater reliability. Subsequently, each individual trial was reviewed and evaluated by one investigator. Ambiguous or unclear cases were brought before the entire team for review and consensus resolution.

Results

The search yielded 37,944 unique articles across 33 medical subject headings. 2,890 trials (7.6%) met our inclusion criteria (Table 2). Of the 37,944 unique records, 166 were meta-analyses and 175 were unique reports identified via review of those meta-analyses. Proportions of eligible trials by medical subject heading varied from <0.5% (Carcinoma, Ductal, Breast) to 23.5% (Prostatic Hyperplasia). Application of influence criteria identified 290 trials for review (Table 2). Trials of cardiovascular procedures accounted for 18.3% (n=53) of the pool while trials of neurological procedures represented only 4.8% (n=14). See Figures 1 and 2 for sample flow charts of trial inclusion and exclusion.

After a full review of the reports, we evaluated the hypothesis that those studies identified by the two different influence criteria would not vary by fidelity to reporting and design standards. A Chi Square test employing the Bonferroni method for adjustment for multiple comparisons showed no significant difference between the two samples on 68 of 70 variables. Data are therefore presented in summary form for all studies surveyed. No other subgroup analyses were conducted.

Of trials surveyed, more than a third (104/290, 35.9%) did not identify the trial sponsor. Of 186 reporting sponsorship, 32.3% (60) were solely commercially funded, and another 15.6% (29) were had joint commercial and non-commercial funding. Included reports were roughly equally divided between device trials (142/290, 49%) and non-device trials (148/290, 51%). Of 95 device trials reporting sponsorship, 64 (67.4%) had some level of commercial sponsorship.

A significant number of reports failed to disclose basic information necessary to assess the internal validity of results. A formal statistical power calculation was reported in only two-thirds (63.4%; 184/290), and of these 13% (24) failed to include one or more elements of a complete power calculation: estimated treatment outcomes for each group, standard deviation for continuous variables, the alpha or significance level (Type 1 error), and the beta (Type 2 error) level or, conversely, power. Fewer than half of reports (42.4%, 78/184) made explicit allowance for attrition in the power calculation.

15.5% (45/290) failed to specify the primary endpoint used in analysis. Out of 245 specifying an endpoint, 14.3% (35) used a composite endpoint and 20.8% (51) used a subjective endpoint. Fewer than half of trials utilizing subjective endpoints (47.1%) reported the provenance (origin) of the subjective scale, and 2% (1), 3.9% (2), and 0% provided data to support the validity, reliability, and sensitivity of the scale, respectively.

Thirty of 290 trials (10.3%) used deterministic allocation schemes, while less than half of randomized trials reported the method of allocations sequence generation (41%, 105/256). Most studies did not report the use of blinding (60.7%; 176/290), and only 46 trials (40.4% of trials reporting blinding, 15.9% of total sample) reported blinding the primary outcome assessor. Only 20 out of 51 trials using subjective endpoints reported blinding of the primary outcome assessor (39.2%). Additional data regarding features affecting the internal validity of the trials surveyed can be found in Table 3.

Half of studies specifying the number of trial sites were single-center (133/266). The median number of sites in multi-center studies was 12 (range 2–217). 88% of multicenter trials (117/133) failed to identify any eligibility criteria for inclusion of a center. Of the few reports that disclosed the number of participating interventionists (65/290), roughly half (34/65) had multiple, with a median of 3.5 (range 2–66). Only 14/34 multi-interventionist trials specified eligibility criteria for interventionist participation (41.2%).

Of the 89 trials reporting some commercial sponsorship, 29.2% (26) reported the sponsor's role in the design of the trial, 34.8% (31) in data analysis, and 32.6% (29) in the preparation of the manuscript. Other trial features affecting the generalizability of trial results can be found in Table 4.

Discussion

This study assessed the proportion of published evaluations of invasive therapeutic procedures that used a comparative clinical trial design, and the methodological quality of the most influential of those trials. We note three main findings. First, a very low proportion of studies evaluating surgical procedures met the minimal design attributes we specified as

inclusion criteria – 7.6% overall. This may indicate the acceptance on face validity of practices which have not been adequately tested, or may be a reflection of the many barriers to conducting well-designed comparative trials of surgical procedures, such as the relatively high cost associated and correspondingly limited funding for nondevice surgical trials, or difficulties recruiting patients to randomized trials of operations.^{5, 6}

Second, despite focusing on comparative trials, our review uncovered deficient reporting of important aspects of clinical trials, such as funding source (36% failure), complete statistical power calculations (45% failure), and factors required to assess the generalizability of trial results, e.g., the number and inclusion criteria of interventionists (78% and 59% failure, respectively) and centers (8% and 88% failure, respectively).

Third, we found that while some bias-reducing strategies were frequently used (such as randomization), others were often forgone. Blinding was reported in only four of ten trials overall, and in trials using subjective endpoints, where blinding is key to unbiased assessments, six of ten trials did not blind the primary outcome assessor.

There is international consensus that scientifically unsound trials are *ipso facto* ethically problematic, because they do not produce generalizable knowledge sufficient to justify the use of human participants.^{7, 8} In light of this, our findings indicate failures on the part of investigators, independent review boards, and peer reviewers. Such failures suggest that highly motivated trialists and reviewers may nevertheless be inadequately trained to ensure the use of the best scientific methods in clinical research. In a previous work, we have documented both that the standards governing the conduct of clinical trials are numerous, as well as the inherent difficulty in locating, assimilating, and applying such standards.⁹ These factors may present a real barrier to the ethical conduct of surgical trials, and as such may indicate the need for greater attention to the institutional structure(s) in which surgical research is developed, reviewed, and conducted.

The dissemination of surgical innovations in advance of evidence of effectiveness from randomized trials has long been a matter of concern.¹⁰ Our study provides the best estimate yet of how often such procedures are evaluated in comparative trials generally, as well as with randomized controlled trials implementing bias-reducing strategies. With some exceptions,^{11–13} previous studies counted only the number of studies or clinical trials in major surgical journals,^{14–17} or in one case identified only studies published in general medical journals.¹⁸ Most did not distinguish between trials evaluating surgical procedures and those evaluating surgical adjuncts or drugs related to surgical procedures.¹⁵ In contrast, we examined trials across journal types, and differentiated between comparative studies and case series. Our approach enabled us to identify a sizeable number of prospective comparative trials of invasive therapeutic procedures which could indicate the current state of the art of research into invasive medicine. Because we identified those studies most widely cited within the literature, our analysis focused on those trials most likely to be impacting current clinical practice.

The practical implications of these findings are at least threefold. First, these findings suggest that more emphasis on applying the CONSORT requirements to nonpharmaceutical treatment modalities⁴ could vastly improve the evidence available to surgical interventionists. If pressure was consistently exerted by journal editors and peer reviewers, fidelity to reporting criteria would likely increase.

Second, improvement in clinical research training in the surgical sub-specialties is necessary, as is the establishment of stronger research infrastructure for surgeons.¹⁹ Although some research institutions have begun to incorporate research fellowships into their surgical training, such practices have yet to become mainstream. Additional progress

can be made in the education and training of independent review board members, who must approve surgical research programs before they are implemented.

Finally, although there are unique hurdles to surgical clinical trials,²⁰⁻²⁴ our findings suggest that scientific rigor is not always prohibitively demanding. Almost 90% of reviewed studies did randomize participants, suggesting that the barriers to successful randomization may not be as large as is often presumed. Our results also suggest that blinding the participant in surgical research is achievable,¹⁵ at least when surgical procedures are tested against other invasive interventions. Methods of using blinding in surgical trials have been suggested in the literature,²⁵ and regardless of the nature of the comparator, there is often no substantial obstacle to blinding independent outcome assessors,²⁶ a practice which we saw in only 16% of our full sample. Something as simple as blinding outcome assessors can do a great deal to limit biases in estimates of treatment effects.

There are at least three limitations to our results. First, we evaluated the published reports of trials, which are significantly limited by the length restrictions of most medical journals. Therefore, it is unclear what amounts of the failures we identify reflect mere reporting deficiencies and what proportion indicate a lack of methodological rigor in our sample. However, reporting failures are also a concern, as deficits in methods reporting are correlated with inflated estimates of treatment effects,²⁷ and interventionists must be able to evaluate a procedure on the basis of the published reports.

Second, our analysis did not evaluate whether there was general improvement in the quality of surgical trials over the period of review. This is an important question which we were unable to address due both to a limited sample size and the manner in which our data was coded, which did not easily convert into a summary quality metric. The extension of the CONSORT statement to non-pharmacological treatment modalities, as well as the growing attention paid to research methodology in the surgical literature have hopefully had a positive impact on the quality of surgical trials over recent years.⁴

Finally, our study utilized an unverified measure of influence to reduce the size of our sample. This method assumes both that the publication of research results has a direct causal impact on clinical practice, as well as that the citation rate of trial reports is an appropriate surrogate for a trial's influence on clinical practice. Although there is some evidence in support of the former claim,^{28, 29} it is unclear how those data generalize to the dissemination of surgical research results. The latter assumption is untested and may have biased our results.

Although we uncovered important results indicating the state of scientific rigor in surgical trials, further study is warranted. The field would benefit from research into the most effective means of educating surgical trialists and review board members in the use of bias-reducing strategies in surgical research. Additional research into the best methods for implementing methodological rigor in the testing of treatment modalities which appear hostile to the most common bias reducing strategies would also be valuable.

Our results show that the methodological quality of surgical trials, as well as their reporting, is in need of significant improvement. Clinical research ethics require that the risks of harm to participants be outweighed by the value of the knowledge to be generated. The value and clinical relevance of research-generated data is limited by both the internal and external validity of findings. When trials are conducted without the use of adequate bias-reducing techniques, participants are subjected to invasive procedures in order to produce findings of questionable scientific value and generalizability to the medical community at large. The ethical implications of these findings are compounded outside of the experimental setting,

when clinical interventionists treat patients with procedures which have not been proven efficacious or safe.

Acknowledgments

All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. The work was supported by the National Institutes of Health grant # R01 CA134995. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

References

1. The Cochrane Collaboration. [Accessed May 4, 2011] Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Available at: www.cochrane-handbook.org
2. Thomson Reuters. [Accessed June 16, 2011] Web of Science. Available at: http://thomsonreuters.com/content/science/pdf/Web_of_Science_factsheet.pdf
3. [Accessed May 4, 2011] LexisNexis. Available at: <http://www.lexisnexis.com/>.
4. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008; 148(4):295–309. [PubMed: 18283207]
5. Plaisier PW, Berger MY, van der Hul RL, et al. Unexpected Difficulties in Randomizing Patients in a Surgical Trial: A Prospective Study Comparing Extracorporeal Shock Wave Lithotripsy with Open Cholecystectomy. *World J Surg*. 1994; 18:769–773. [PubMed: 7975698]
6. Wilt TJ, Brawer MK, Barry MJ, et al. The Prostate cancer Intervention Versus Observation Trial: VA/NCI/AHRQ Cooperative Studies Program #407 (PIVOT): Design and baseline results of a randomized controlled trial comparing radical prostatectomy to watchful waiting for men with clinically localized prostate cancer. *Contemp Clin Trials*. 2009; 30(1):81–87. [PubMed: 18783735]
7. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Biomedical Research Involving Human Subjects. Geneva, Switzerland: Available at: http://www.cioms.ch/publications/layout_guide2002.pdf.
8. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA*. 2000; 283(20):2701–2711. [PubMed: 10819955]
9. Ashton CM, Wray NP, Jarman AF, et al. A Taxonomy of multinational ethical and methodological standards for clinical trials of therapeutic interventions. *J Med Ethics*. 2011; 37:368–373. [PubMed: 21429960]
10. McCulloch P, Altman DG, Campbell WB, et al. Surgical Innovation and Evaluation 3: No Surgical Innovation without Evaluation: The IDEAL Recommendations. *Lancet*. 2009 Sep 26.374:1105–1112. 2009. [PubMed: 19782876]
11. Jacquier I, Boutron I, Moher D, et al. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann Surg*. 2006 Nov; 244(5):677–683. [PubMed: 17060758]
12. Balasubramanian SP, Wiener M, Alshameeri Z, et al. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg*. 2006; 244(5):663–667. [PubMed: 17060756]
13. Solomon MJ, Laxamana A, Devore L, McLeod R. Randomized controlled trials in surgery. *Surgery*. 1994; 115(6):707–712. [PubMed: 8197563]
14. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996; 347:984. [PubMed: 8606606]
15. Wente MN, Seiler CM, Uhl W, Buchler MW. Perspectives of Evidence-Based Surgery. *Digest Surg*. 2003; 20:263–269.
16. Chang DC, Matsen SL, Simpkins CE. Why Should Surgeons Care about Clinical Research Methodology? *J Am Coll Surgeons*. 2006 Dec; 203(6):827–830. 2006.
17. Hall JC, FRACS, Mills B, et al. Methodologic standards in surgical trials. *Surgery*. 1996; 119:466–472. [PubMed: 8644014]

18. Walter CJ, Dumville JC, Hewitt CE, et al. The Quality of Trials in Operative Surgery. *Ann Surg.* 2007; 246(6):1104–1108. [PubMed: 18043117]
19. Jarman AF, Wray NP, Wenner DM, Ashton CM. Trials and Tribulations: The Professional Development of Surgical Trialists. *Am J Surg.* 2012 In press.
20. McLeod R. Issues in surgical randomized controlled trials. *World J Surg.* 1999; 23(12):1210–1214. [PubMed: 10552108]
21. McCulloch P, Taylor I, Sasako M, et al. Randomised trials in surgery: problems and possible solutions. *Brit Med J.* 2002; 324(7351):1448–1451. [PubMed: 12065273]
22. Lilford R, Braunholtz D, Harris J, Gill T. Trials in surgery. *Brit J Surg.* 2004; 91(1):6–16. [PubMed: 14716788]
23. Ergina PL, Cook JA, Blazeby JM, et al. Surgical Innovation and Evaluation 2: Challenges in Evaluating Surgical Innovation. *Lancet.* 2009 Sep 26.374:1097–1104. 2009. [PubMed: 19782875]
24. Abraham NS. Will the Dilemma of Evidence-Based Surgery Ever Be Resolved? *ANZ J Surg.* 2006; 76(9):855–860. [PubMed: 16922912]
25. Boutron I, Guittet L, Estellat C, et al. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med.* 2007; 4(2):e61. [PubMed: 17311468]
26. Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: does blinding of outcome assessors matter? *J Bone Joint Surg Am.* 2007; 89(3):550–558. [PubMed: 17332104]
27. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995; 273(5):408–412. [PubMed: 7823387]
28. Ketley D, Woods KL. Impact of clinical trials on clinical practice: example of thrombolysis for acute myocardial infarction. *Lancet.* 1993; 342:891–894. [PubMed: 8105166]
29. Lamas GA, Pfeffer MA, Hamm P, et al. Do the results of randomized clinical trials of cardiovascular drugs influence medical practice? *N Engl J Med.* 1992; 327(4):241–247. [PubMed: 1535419]

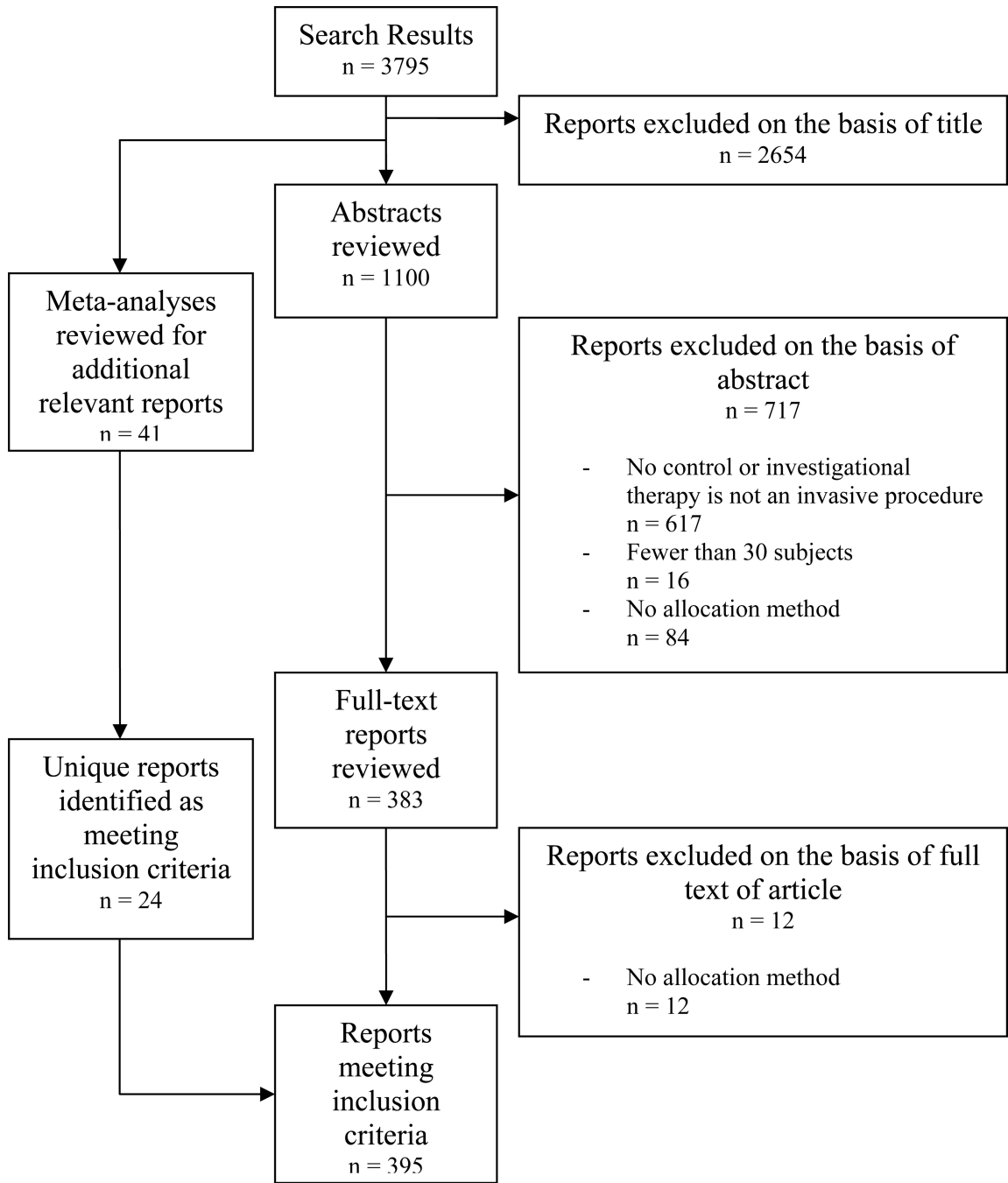


Figure 1. Review strategy and yield for trials of invasive therapeutic procedures for myocardial infarction

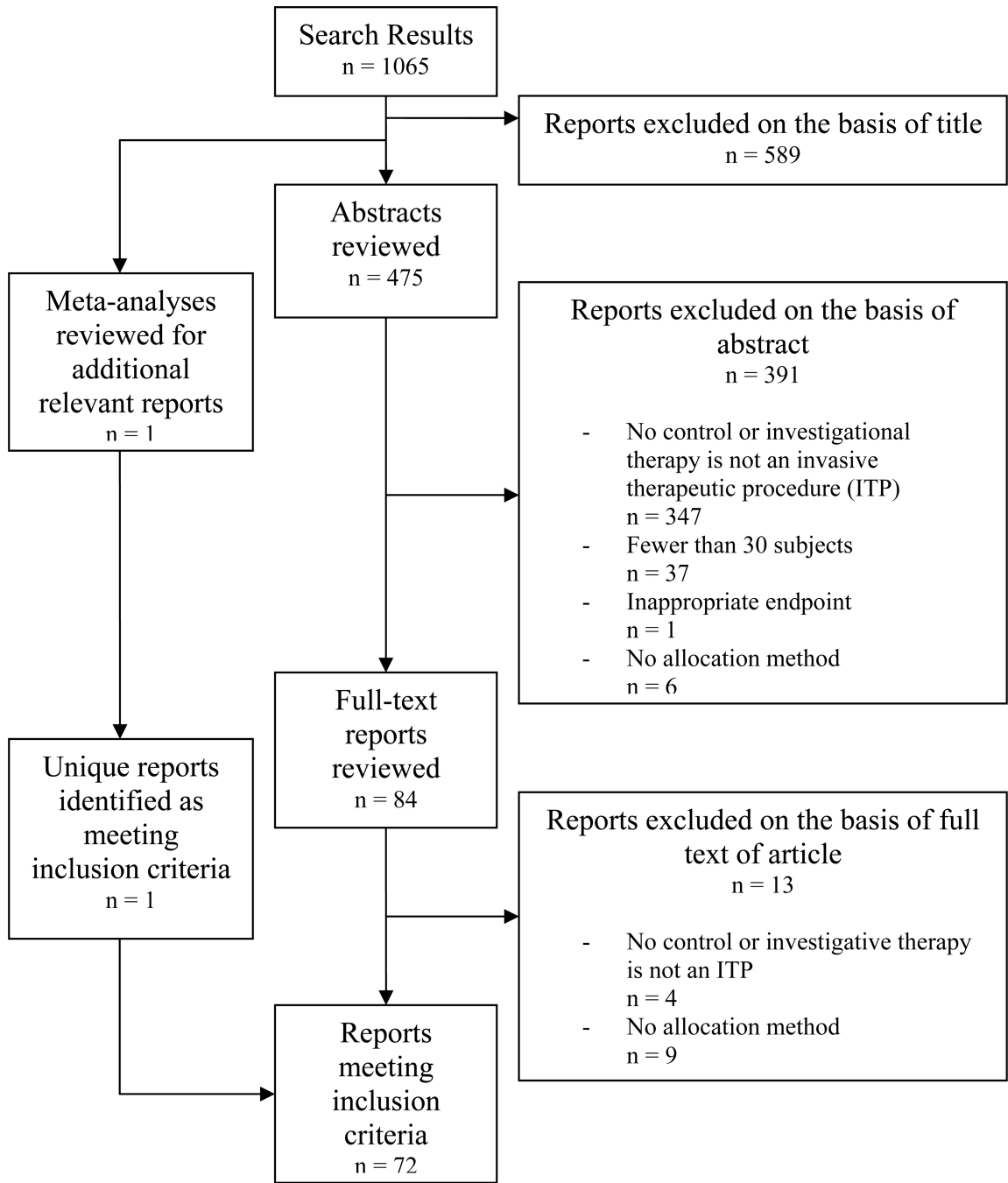


Figure 2. Review strategy and yield for trials of invasive therapeutic procedures overnutrition

Table 1

Invasive Therapeutic Procedure Clinical Trial Search Strategy

<p>Pubmed Search</p> <p>{MeSH Search Term}</p> <p>(randomized controlled trial [pt] OR controlled clinical trial [pt] OR comparative study [pt] OR evaluation studies [pt]) NOT (animal [mh] NOT human [mh]) AND (surgical procedures [mh] OR surgi* [tw] OR surger* [tw] OR operativ* [tw] OR endosc* [tw] OR laser* [tw] OR transplantation [mh] OR transpl* [tw])</p> <p>Limits</p> <p>Publication Date from January 1, 1999 to December 31, 2008</p> <p>Only items with Abstracts</p> <p>English</p>
--

Table 2

Medical Subject Headings (MeSH), number of publications 1999–2008 identified by the search strategy, and n (%) reports meeting inclusion criteria as a comparative clinical trial evaluating an invasive therapeutic procedure.

Medical Subject Heading	Total Published Reports Retrieved*	Eligible Reports, n (%)	N Eligible Reports Meeting Influence Criteria & Included in Review Pool
Aortic Aneurysm	1,466	113 (7.7)	9
Arrhythmia	2,100	181 (8.6)	7
Arthritis	1,701	171 (10.1)	15
Barrett Esophagus	300	18 (6.0)	7
Brain Neoplasm	1,423	15 (1.1)	5
Carcinoma, Bronchogenic & Non-Small-Cell-Lung	798	24 (3.0)	7
Carcinoma, Ductal, Breast	666	3 (0.5)	3
Carotid Artery Disease & Carotid Stenosis	1,016	76 (7.5)	13
Cerebrovascular Accident	2,584	302 (11.7)	8
Cholecystitis	225	32 (14.2)	5
Cholelithiasis	241	39 (16.2)	5
Colorectal Neoplasm	3,078	116 (3.8)	18
Coronary Disease	3,304	325 (9.8)	32
Dystonic Disorder	32	3 (9.4)	3
Embolism	875	72 (8.2)	7
Endometriosis	305	23 (7.5)	7
Esophageal Neoplasm	786	40 (5.1)	8
Esophagitis	501	33 (6.6)	7
Heart Failure, Congestive	2,571	125 (4.9)	6
Intracranial Arteriovenous Malformation	104	1 (1.0)	1
Laryngeal Neoplasm	396	19 (4.8)	6
Myocardial Infarction	3,819	395 (10.3)	13
Nephrolithiasis	188	23 (12.2)	12
Otitis Media	278	44 (15.8)	7
Ovarian Neoplasm	819	14 (1.7)	6
Overnutrition	1,066	72 (6.8)	6
Parkinson Disease & Parkinsonian Disorders	351	23 (6.6)	5
Peripheral Vascular Disease	823	75 (9.1)	10
Prostatic Hyperplasia	531	125 (23.5)	6
Prostatic Neoplasm	1,861	23 (1.2)	7
Pulmonary Disease, Chronic Obstructive	306	14 (4.6)	6
Thrombosis	2,556	186 (7.3)	10
Urinary Incontinence	874	134 (15.3)	23

Medical Subject Heading	Total Published Reports Retrieved[*]	Eligible Reports, n (%)	N Eligible Reports Meeting Influence Criteria & Included in Review Pool
TOTALS	37,944	2,890 (7.6)	290

* Total including eligible trial reports discovered by review of citation lists of meta-analyses

Table 3

Trial attributes influencing internal validity

Attribute	N(%) Reporting
Purpose of trial	
Demonstrate superiority ^a	235/290 (81.0)
Demonstrate equivalence or non-inferiority	43/290 (14.8)
Assess safety	11/290 (3.8)
Feasibility	1/290 (0.3)
Formal sample size and power calculation	
Not reported	106/290 (36.6)
Reported	184/290 (63.4)
Calculation included all elements	160/184 (87.0)
Calculation allowed for dropouts and attrition	78/184 (42.4)
Primary endpoint	
Not reported	45/290 (15.5)
Reported	245/290 (84.5)
Composite primary endpoint	35/245 (14.3)
Single objective primary endpoint ^b	159/245 (64.9)
Subjective primary endpoint	51/245 (20.8)
Reported provenance of scale	24/51 (47.1)
Reported validity	1/51 (2.0)
Reported reliability	2/51 (3.9)
Reported sensitivity	0/51 (0.0)
Reported blinding of primary outcome assessor	20/51 (39.2)
Type of control group	
Concurrent	282/290 (97.2)
Historical	8/290 (2.8)
Nature of comparator	
Placebo or sham procedure	11/290 (3.8)
Alternative operative intervention	220/290 (75.9)
Non-operative intervention	59/290 (20.3)
Allocation method	
Deterministic ^c	30/290 (10.3)
Randomization	256/290 (88.3)
Minimization ^d	4/290 (1.4)
Allocation concealment (256 randomized trials only)	
Method described for generation of random allocation sequence	105/256 (41.0)
Identification of person or entity generating the sequence	82/256 (32.0)
Provided assurance that sequence was concealed until allocation	132/256 (51.6)

Attribute	N(%) Reporting
Blinding	
Trial reported some element of blinding	114/290 (39.3)
Participants reported blinded	41/114 (36.0)
Interventionist reported blinded	14/114 (12.4)
Primary outcome assessors reported blinded	46/114 (40.4)
Maintenance of blind reported assessed	5/114 (4.4)
Trial execution	
Participants receiving treatment as allocated	
Treatment received by allocation not reported	27/290 (9.3)
<85% of intervention participants received intervention	13/263 (4.9)
<85% of control participants received control condition	8/263 (3.0)
Full crossover data not reported (256 randomized trials only)	29/256 (11.3)
>10% crossovers, intervention to control	17/227 (7.5)
>10% crossovers, control to intervention	9/227 (4.0)
Follow-up	
Full data regarding completeness of follow-up not reported	41/290 (14.1)
Reported follow-up in entire sample only	30/249 (12.0)
<90% intervention participants completing follow-up	59/219 (26.9)
<90% control participants completing follow-up	58/219 (26.5)

^aOne trial was powered to show non-inferiority but the authors interpreted findings of “no significant difference” as supporting the claim that the surgical procedure under investigation was superior on the grounds that surgery, in contrast to medical therapy, corrected the anatomic defect thought to be responsible for the pathophysiology of the disease. We classified this trial as a superiority trial due to the ultimate claims made by the authors.

^bObjective endpoints were defined as those which were externally verifiable, i.e. re-admission or morbidity.

^cOne trial was described as being randomized, but used hospital registration numbers for allocation (even to one group and odd to the other). This trial was recorded as deterministic.

^dOne trial began with a strategy of minimization, but switched to randomization midway. This trial was recorded as using minimization.

Table 4

Attributes needed for assessing external validity or generalizability

Attribute	N(%) Dossiers
N intervention centers	
N centers not reported	24/290 (8.3)
Single center	133/266 (50.0)
Multi-site	133/266 (50.0)
Median number of centers (range)	12 (2–217)
2–5 centers	40/133 (30.1)
6–10	23/133 (17.3)
>10	70/133 (52.6)
Eligibility criteria for inclusion of intervention centers (multi-center trials only)	
Not reported	117/133 (88.0)
Inclusion criteria described	16/133 (12.0)
N surgeons/interventionists	
N interventionists not reported	225/290 (77.6)
N interventionists reported	65/290 (22.4)
Single interventionist	31/65 (47.7)
Multiple interventionists	34/65 (52.3)
Median number interventionists (range)	3.5 (2–66)
2–5 interventionists	21/34 (61.8)
6–10	7/34 (20.6)
>10	6/34 (17.6)
Eligibility criteria for inclusion of interventionists (multi-interventionist trials only)	
Not reported	20/34 (58.8)
Inclusion criteria described	14/34 (41.2)
Pre-trial standardization of operative technique (multi-interventionist trials only)	
Not reported	26/34 (76.5)
Described critical elements of technique	8/34 (23.5)
Assessment of interventionist fidelity to standardized operative technique (multi-interventionist trials only)	
Not reported	31/34 (91.2)
Assessment method described	3/34 (8.8)
Sponsor role in conduct of trial (Out of trials reporting commercial sponsorship only)	
Trial design	26/89 (29.2)
Data analysis	31/89 (34.8)
Preparation of manuscript	29/89 (32.6)
Size of trial	
Median number of trial participants (range)	121.5 (30–3,120)
30 participants	4/290 (1.4)
31–60	56/290 (19.3)

Attribute	N(%) Dossiers
61-100	59/290 (20.3)
101-200	63/290 (21.7)
>200	108/290 (37.2)