



Published in final edited form as:

JAMA. 2005 February 23; 293(8): 970–978.

## Evaluation of New Treatments in Radiation Oncology:

### Are They Better Than Standard Treatments?

**Heloisa P. Soares, MD, Ambuj Kumar, MD, MPH, Stephanie Daniels, RN, MPH, Suzanne Swann, PhD, Alan Cantor, PhD, Iztok Hozo, PhD, Mike Clark, DPhil, Fadila Serdarevic, MD, MPH, Clement Gwede, RN, PhD, Andy Trotti, MD, and Benjamin Djulbegovic, MD, PhD**  
*H. Lee Moffitt Cancer Center & Research Institute at University of South Florida, Tampa (Drs Soares, Kumar, Cantor, Serdarevic, Gwede, Trotti, and Djulbegovic and Ms Daniels); Department of Statistics, Radiation Therapy Oncology Group, Philadelphia, Pa (Dr Swann); Department of Mathematics, University of Indiana Northwest, Gary (Dr Hozo); UK Cochrane Centre, Oxford, England (Dr Clark).*

### Abstract

**Context**—The superiority of innovative over standard treatments is not known. To describe accurately the outcomes of innovations that are tested in randomized controlled trials (RCTs) 3 factors have to be considered: publication rate, quality of trials, and the choice of the adequate comparator intervention.

**Objective**—To determine the success rate of innovative treatments by assessing preferences between experimental and standard treatments according to original investigators' conclusions, determining the proportion of RCTs that achieved primary outcomes' statistical significance, and performing meta-analysis to examine if the summary point estimate favored innovative vs standard treatments.

**Data Sources**—Randomized controlled trials conducted by the Radiation Therapy Oncology Group (RTOG).

**Study Selection**—All completed phase 3 trials conducted by the RTOG since its creation in 1968 until 2002. For multiple publications of the same study, we used the one with the most complete primary outcomes and with the longest follow-up information.

---

**Corresponding Author:** Benjamin Djulbegovic, MD, PhD, H. Lee Moffitt Cancer Center & Research Institute at the University of South Florida, Department of Interdisciplinary Oncology, 12902 Magnolia Dr, SRB 4, Room 24031, Tampa, FL 33612 (djulbebm@moffitt.usf.edu).

**Author Contributions:** Dr Djulbegovic had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Study concept and design:* Soares, Clark, Gwede, Djulbegovic.

*Acquisition of data:* Soares, Daniels, Swann, Gwede, Djulbegovic. *Analysis and interpretation of data:* Soares, Kumar, Cantor, Hozo, Clark, Serdarevic, Gwede, Trotti, Djulbegovic.

*Drafting of the manuscript:* Soares, Kumar, Daniels, Hozo, Clark, Gwede, Djulbegovic. *Critical revision of the manuscript for important intellectual content:* Kumar, Swann, Cantor, Clark, Serdarevic, Trotti.

*Statistical analysis:* Soares, Kumar, Swann, Cantor, Hozo, Djulbegovic.

*Obtained funding:* Djulbegovic.

*Administrative, technical, or material support:* Kumar, Daniels, Serdarevic, Gwede.

*Study supervision:* Clark, Djulbegovic.

**Financial Disclosures:** None reported.

**Funding/Support:** This research was supported by the Research Program on Research Integrity, an Office of Research Integrity/National Institute of Health collaboration, grant 1R01NS/NR44417-01, which exclusively provided financial support for this project.

**Role of the Sponsor:** The sponsor has not had any 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.

**Data Extraction**—We used the US National Cancer Institute definition of completed studies to determine the publication rate. We extracted data related to publication status, methodological quality, and treatment comparisons. One investigator extracted the data from all studies and 2 independent investigators extracted randomly about 50% of the data. Disagreements were resolved by consensus during a meeting.

**Data Synthesis**—Data on 12734 patients from 57 trials were evaluated. The publication rate was 95%. The quality of trials was high. We found no evidence of inappropriateness of the choice of comparator. Although the investigators judged that standard treatments were preferred in 71% of the comparisons, when data were meta-analyzed innovations were as likely as standard treatments to be successful (odds ratio for survival, 1.01; 99% confidence interval, 0.96-1.07;  $P=.5$ ). In contrast, treatment-related mortality was worse with innovations (odds ratio, 1.76; 99% confidence interval, 1.01-3.07;  $P=.008$ ). We found no predictable pattern of treatment successes in oncology: sometimes innovative treatments are better than the standard ones and vice versa; in most cases there were no substantive differences between experimental and conventional treatments.

**Conclusion**—The finding that the results in individual trials cannot be predicted in advance indicates that the system and rationale for RCTs is well preserved and that successful interventions can only be identified after an RCT is completed.

**R**ANDOMIZED CONTROLLED trials (RCTs) are considered one of the most reliable methods for assessing health care interventions.<sup>1-3</sup> However, “What is the probability that experimental/innovative treatments are superior to standard, established treatments?”<sup>4</sup> is, surprisingly, not known.

One of the reasons that this question has not been answered before is because all the factors that can affect the outcomes of RCTs have never been comprehensively accounted for. During the last decade, several factors were found to influence a distribution of the outcomes of RCTs. In general, it is considered that 3 main factors might influence and modify the trial's results: publication bias,<sup>5,6</sup> methodological quality of the trials' design,<sup>7-9</sup> and choice of comparator.<sup>10-12</sup> Therefore, to accurately describe the outcomes of RCTs and assess how often innovative treatments are better than the standard treatments, one needs to comprehensively take into account all 3 of these factors.

In this study, our objective was to evaluate treatment successes in oncology. We focused on the research efforts conducted by the Radiation Therapy Oncology Group (RTOG), one of the major cooperative groups (COGs) sponsored by the public funds from the US National Cancer Institute (NCI). By selecting NCI-sponsored trials, we were in a unique position to evaluate the success of research efforts from the perspective of a single funding agency (ie, NCI), which has a common platform for the development of preventive and therapeutic advances in oncology.<sup>13</sup>

## METHODS

Data were extracted from all RCTs conducted by the RTOG. These studies were performed after their research protocols had passed a rigorous peer-review process and been approved through the RTOG committee system and the NCI prior to their activation.<sup>14</sup> Thus, evaluation of the RTOG studies created a unique opportunity to assess the pattern of successes of new treatments in radiation oncology.

### Determination of Publication Rate

We identified and included in our study all completed phase 3 trials conducted by the RTOG since its creation in 1968 until 2002. Accurate determination of publication rate depends on accurate determination of the numerator and denominator of the studies considered for calculating the publication rate.<sup>15</sup> We used the NCI definition of completed studies to

determine the publication rate: a study is considered to be completed if it has been closed to accrual, all patients have completed therapy, and the study has met its primary objectives.<sup>16</sup> The studies that were initiated but closed early due to poor patient accrual or closed to accrual, but still required further follow-up to meet their primary outcomes, were excluded from the analysis. Some studies had multiple articles. If we identified more than 1 publication arising from a single research protocol, we used the publication containing the most complete data related to primary outcomes with the longest follow up information (FIGURE 1).

### Assessment of Quality

To facilitate the extraction and evaluation of the trials, we requested all research protocols related to all studies. We extracted data related to publication status, methodological quality, and treatment comparisons.<sup>17</sup> One investigator (H.P.S.) extracted the data from all studies and 2 independent investigators (A.K. and B.D.) extracted randomly about 50% of the data (25% each). Interobserver agreement was high ( $\kappa=0.92$ ).<sup>18</sup> In rare instances of disagreements, a consensus meeting was held to resolve any interobserver disagreement.

To assess the methodological quality of the trials, we extracted data on the methodological domains that have been acknowledged to be vital for minimizing bias and random error in the conduct and analysis of RCTs.<sup>19-21</sup> We recently reported our findings related to the methodological quality of the studies and research protocols performed by the RTOG.<sup>22</sup>

### Assessment of Distribution of the Outcomes Between Innovative and Standard Treatments

For each trial we identified the standard (conventional) intervention and the innovative (experimental) intervention(s). The definition of standard treatment was made on the basis of information provided by background sections of the articles and related protocols. When the conventional treated group was not explicitly defined (n=1), an expert in the field was consulted for clarification. The trials that compared 2 experimental treatment groups against each other were excluded. Similarly, our protocol specified exclusion of trials that had been designed to assess equivalence rather than superiority.<sup>23</sup> However, we found no equivalence studies in this sample of trials conducted by RTOG.

To determine if the experimental or standard treatment was superior, we recorded the primary outcomes specified in the trialists' null hypothesis for each study. The distribution of the outcomes was determined in 3 ways:

1. By determining what proportion of the trials were statistically significant according to the primary outcomes specified in a null-hypothesis.
2. By qualitative assessment (based on the trialists' claims) of which treatment was judged "better." This was necessary because a simple determination of the proportion of trials that achieved statistical significance on their primary outcomes does not capture fully the trade-off between benefits and harms of competing interventions. We assessed preferences between innovative and standard treatments based on the reports by the original trialists.<sup>10,24-26</sup>
3. Finally, quantitative meta-analytic techniques were used to pool data on the most important outcomes.<sup>17,21</sup> Since most events in cancer studies are reported as time-to-event data, and not as a proportion of successes or failures, we synthesized data on the most important events (deaths or relapses) to explore the distribution of the outcomes between experimental and standard treatments.<sup>27,28</sup> Summary effects were expressed as a hazard ratio or Peto odds ratio (OR)<sup>21,29</sup> with 99% confidence interval (CI), which was considered more appropriate to minimize the chance of obtaining false-positive results.<sup>30</sup> Although our main analyses focused on pooling data related to primary outcomes, we also pooled data on survival from all studies since survival

was one of the outcomes in almost all studies and it is certainly the most important outcome in cancer.

The unit of analysis was the comparison within each trial. For trials or reports that included more than one innovative treatment group, we pooled the data from all such groups if the results suggested that these groups were not better than the standard treatment. However, if, within a trial, one of the innovative treatments was classified as successful compared with standard treatment and other innovative treatments were not, this treatment only was compared against the standard group while the remaining groups were excluded from the analysis. (In this way, the analyses were biased against the standard treatments and thereby provided the best case scenario for the success of innovative treatments in oncology.) This situation was uncommon (only 13 trials were multi-group trials), and the results of our analyses did not change when the trials were reclassified according to the most important comparison in multi-group trials.

Ultimately, judgments about the value of any given intervention should include evaluation of both benefits and harms, even if a null hypothesis did not focus on simultaneous assessment of treatments' benefits and harms.<sup>31</sup> To provide the overall picture about the effects of treatment interventions evaluated in the RTOG trials we pooled data on treatment-related mortality and morbidity in the manner described above. Data about quality of life were not reported in any of the publications.

All the analyses were performed using SYSTAT<sup>32</sup> and REVMAN software.<sup>33</sup>

### Classification of Comparator

Outcomes of the trials may also be affected by the choice of an inappropriate comparator. Even if the study adheres to all contemporary standards of the good trial design,<sup>34</sup> it may result in biased findings if the control intervention selected in the initial stages of the trial design was inadequate or inferior.<sup>12</sup> In fact, solid evidence exists to point out that trials that incorporate placebo or no therapy as comparative controls may lead to biased results.<sup>10,12,35-38</sup> Studies that use inadequate comparators are thought to violate a fundamental scientific and ethical principle of equipoise, which requires that the patients should be enrolled in an RCT only if there is substantial uncertainty about which of the trial treatments might be most beneficial.<sup>10,12,35-37</sup> Therefore, we classified comparator as active vs placebo or no active treatment and analyzed the data accordingly in order to evaluate the hypothesis that if equipoise is truly preserved in this subgroup of (placebo or no therapy) trials, the rate of successful intervention would not differ from the rate of successful intervention in the main analysis.

This study was approved by the institutional review board of the University of South Florida.

## RESULTS

We identified 60 completed phase 3 trials. One trial was excluded from our analysis because both interventions were considered to be innovative. The publication rate was high: 56 of 59 trials (95%) had been published as full articles. Data were also extracted from one of the unpublished manuscripts. Data from the remaining 2 unpublished trials were not available. One of the studies is currently being analyzed. The other trial had some flaws in the data collection and RTOG decided that publication of the study would generate misleading conclusions. Figure 1 illustrates the details of identification, selection, and determination of publication rate of the RTOG studies. The average time to a complete publication was 10.7 years, ranging from 5 to 22 years.

TABLE 1 shows the characteristics of the studies included in our analysis. The majority of the trials studied interventions in cancer of head and neck, lung, central nervous system, and metastasis. Most of the trials studied definitive treatments, followed by palliative or adjuvant

treatment. A large number of the trials evaluated different types or doses of radiotherapy. Only 3 trials included a placebo or no therapy group as the standard treatment.

As described in the “Methods” section, we first assessed distribution of the outcomes according to the statistical significance and treatment preferences by the original trialists. Surprisingly, the standard treatments were preferred in 42 (71%) of the 59 comparisons (from 57 trials) by the researchers who conducted the studies. However, only 7 trials were statistically significant (6 favoring innovative treatments). In 52 (88%) of the comparisons, there was no statistically significant difference favoring either standard or innovative treatment. In 34 (65%) of the 52 trials, the lack of difference between innovative and standard treatments was a real one, ie, true-negative finding (evidence of no treatment effect) rather than no evidence of effect (false-negative results).<sup>8</sup> We focused on excluding potential benefits of innovative treatments and made this interpretation based on the expected differences in primary outcomes between treatment groups.<sup>39</sup> However, we should mention that in hindsight, many trials were expected to detect unrealistically large effects of new therapies, often as large as 50% to 100%. Thus, in reality, more trials should be considered as inconclusive (no evidence of effect) instead of true-negative trials (evidence of no effect).<sup>39</sup>

FIGURE 2 illustrates the effect of the treatments on survival based on all the trials that reported survival data (n=12734 patients). The Peto OR for the combined analysis was 1.01 (99% CI, 0.96-1.07;  $P=.5$ ) indicating no overall statistically significant difference between innovative and standard treatments in the RTOG cohort of the studies. We did not discern the change in a trend in success rate over time (data not shown). Similar results were seen for all others outcomes (FIGURE 3). When we further analyzed data for possible harms, we found that the treatment-related mortality was better in the standard groups (Figure 3). Patients allocated to innovative treatment groups had more deaths associated with treatment than the ones in the standard treatment groups. In absolute terms, the number of deaths related to treatment toxicity was, however, small representing only 1% of the population of all deaths. This is probably the reason overall survival was not affected by increase in treatment-related mortality. Unfortunately, we were able to pool data on treatment-related morbidities from 51% of the trials only, despite the fact that toxicities were reported in 54 (92%) studies. This was because it often was not clear if the toxic events occurred in a single patient or in many patients preventing proper use of meta-analytic techniques. However, based on extractable data we found that innovative treatments were associated with more adverse effects than standard treatments (OR, 1.27; 95% CI, 1.08-1.5).

TABLE 2 describes the methodological quality of the studies based on the combination of research protocols, published data, and verification by the RTOG Department of Statistics. In general, the methodological quality of the trials was high and did not influence the distribution of the finding in our study. FIGURE 4 represents sensitivity analysis of survival according to the methodological domains described in the articles and compared to the information extracted from the research protocols and data provided by the RTOG statistical office (ie, we compared the quality of reporting to what was actually done). Since the quality of these trials were high, sensitivity analysis did not show any significant difference in the impact of the methodological quality on outcome of the trials regardless of the methodology used to assess the trial quality.

When we evaluated the effect of the choice of comparator on the treatment effect, we found that in 3 trials that used placebo or no active treatment as standard treatment, the overall OR was 1.39 (99% CI, 1.08-1.79;  $P<.001$ ; Figure 3), favoring the standard intervention. This result excludes a possibility that innovative treatments may be favored because of the use of inferior comparators.<sup>35</sup>

## COMMENT

Clinical trials are conducted to reduce uncertainty, and they should only be done if health care professionals and the patients are uncertain which of the alternative interventions is better.<sup>11,40,41</sup> This requirement is sometimes referred to as “the uncertainty principle”<sup>42</sup> or “equipoise.”<sup>36,41,43,44</sup> We previously hypothesized that there is a predictable relationship between the uncertainty principle, that is, the moral principle, upon which trials are based and the ultimate outcomes of clinical trials.<sup>11</sup> That is, if all trials (published and unpublished) are analyzed, the quality of the trials is high, and if the investigators cannot predict what they are going to discover (ie, they are uncertain about the relative effects of the competing treatment alternatives), we hypothesize that in aggregate, innovative and standard treatments should be expected to be equally successful. We would expect that innovative treatments will sometimes be superior, sometimes standard will be better, and sometimes there will be no difference between experimental and standard treatments. Our results indicate that this hypothesis may be correct.

We found that none of the possible factors that might affect the distribution of the findings of RCTs could explain our results. The publication rate of RTOG trials was very high (95%) effectively excluding publication bias as a factor affecting the distribution of the findings in this set of trials.<sup>6</sup>

Similarly, the methodological quality of the trials was very high, probably highest of any cohort of the trials reported in the literature.<sup>22</sup> Thus, concern that poor methodological quality of the trials may have distorted results,<sup>34,45</sup> does not appear to apply to our results.

Finally, we found no evidence that the comparator intervention that had been chosen for testing in these RCTs was inferior, which would have violated the uncertainty principle or equipoise.<sup>11,35</sup> As shown in Figure 3, in all 3 RTOG trials in which placebo or no active treatment was used as a control, these interventions proved to be superior to the experimental treatments against which they were tested. This result strongly suggests that the RTOG investigators did not violate equipoise when they designed these trials. Previously, we showed that equipoise is violated in the industry-sponsored trials.<sup>10</sup> In fact, we found that the main mechanism of bias in commercially sponsored trials is via selection of inferior comparator.<sup>12</sup> Our results indicate that equipoise may not be violated in the publicly sponsored trials. Despite the fact that the NCI trials are elite, it is not, however, clear at this time if the results based on the research of one group (ie, RTOG) can be generalized to the other publicly sponsored cooperative groups. Currently, we are in the process of collecting data from other COGs to answer this question. Further studies to confirm the generalizability of our results are required and would be useful.

We believe that the results presented herein are in concordance with the equipoise hypothesis.<sup>11</sup> Our results seem to indicate that there is an equal chance for innovative and standard treatments to result in successful interventions or in the outcomes that may not differ between 2 types of treatment. That is, there is no evidence that the outcomes of RCTs, in aggregate, consistently and predictably favor innovative over control interventions, in spite of the trialists' hopes that the new interventions would be better. It is important to note that this does not apply to individual treatments, which can be more or less successful in an unpredictable way. Whether a particular intervention would be more or less successful, could only be known after the relevant RCT had been conducted. Ethically, this is also a welcome finding, because if treatment success can consistently be predicted, the patients would be naturally expected to request such a treatment, making randomization impossible. Therefore, with the caveats expressed above, we conclude that a system of RCTs, at least the ones conducted by RTOG, is well preserved.

However, one should note that our results are consistent with this interpretation based on the quantitative pooling of treatment effects (Figure 3). When data were interpreted according to the trialist's judgments of the treatment successes, we found that researchers favored standard treatment in 71% of comparisons. This occurred, even though the research protocols that we examined clearly showed that investigators had been hoping that the innovative treatments would be better. We are not sure why we saw this distribution of the investigators' conclusion about treatments' superiority. The usual interpretation is that researchers are biased in their conclusion and they are often at variance with objective data.<sup>46</sup> However, RTOG investigators did not favor new treatments they were hoping would be better, but to the contrary, they concluded that the standard treatments were, in aggregate, superior. This effectively rules out interpretation bias in favor of experimental treatments.<sup>47</sup> We should note that many trials were in fact inconclusive, possibly affecting the trialists' interpretation of the value of treatments tested. We also found that treatment-related mortality and treatment-related morbidity were, on average, higher in the innovative arms. Although none of the articles reported quality of life data, it appears that the increase in toxicity associated with those in innovative treatment groups may have swayed the investigators to conclude that introduction of new treatments in practice is not worth the price of toxicity. This finding makes sense since no physician would add toxic or unnecessary treatments or would administer a new treatment with which he or she is not familiar without consistent supporting data. The future RTOG studies should routinely include quality of life as the one of the main outcomes, particularly because few treatments proved to be curative.

Ultimately, our study addresses an epistemological question related to advances in cancer. Preferences of the RTOG researchers for standard treatments can be understood in terms of Kuhn's notion that the standards for adoption of new hypotheses and theories are generally high before a particular scientific group can accept their results.<sup>48,49</sup> Thus, the reluctance of the RTOG investigators to accept new treatments and replace those that are better understood is an expected result from the perspective of theory of knowledge. In the context of our results, this notion is further enforced by the fact that individually and collectively standard treatments are more likely to be used optimally than new treatments that are associated with increased treatment-related mortality. The lack of standards for reporting of treatment-related harms further prevents the researchers from developing criteria under which new treatments may become more acceptable. Thus, the reluctance to adopt new equally efficacious treatments is likely reinforced by increase in treatment-related mortality and morbidity and perception that innovative treatments are usually more costly than the standard treatments. Recent development of the CONSORT statement on the standard for reporting of harms may increase understanding about the issues related to treatment harms.<sup>50</sup>

We should understand that even though our interobserver agreement on the assessment of the investigators' preferences was excellent, we did not have access to the researchers' subtle thinking about the overall value of the interventions that they tested. For example, the issue of cost, or the logistics of installing expensive equipment to deliver the innovative treatment more widely, which usually was not explicitly raised in the discussion section of the articles that we examined, may have played a role in concluding that new treatments are not better than the existing ones and therefore may not be worth introducing in practice.

Because formal statistical methods have not been developed to test a null hypothesis based on integration of all of the variables outlined above, we believe that treatment successes in oncology should be evaluated according to the trials' null hypothesis aiming to test differences between the effects of treatments on a specified primary outcome. Since in oncology survival is almost always primary or an important secondary outcome, treatment successes should also be evaluated against this outcome. When examined this way, the data presented herein show

that there is an equal chance for innovative and standard treatments to result in successful outcomes.

The finding that the result of an individual trial cannot be predicted, even if the majority of trials do not reveal important differences between innovative and standard treatments, is worth emphasizing: We can only know whether one treatment is better than another after an RCT was conducted. Improvements in clinical outcomes can only come from continuing empirical testing of new treatments in RCTs. There is no substitute to this painstaking process of continuing testing of new treatments in the clinical trial setting. Because of this, we believe that the current trend toward deemphasizing the role of RCTs in favor of other study designs is of concern. The public should be more aware of the amount of effort it takes to generate a new successful treatment.

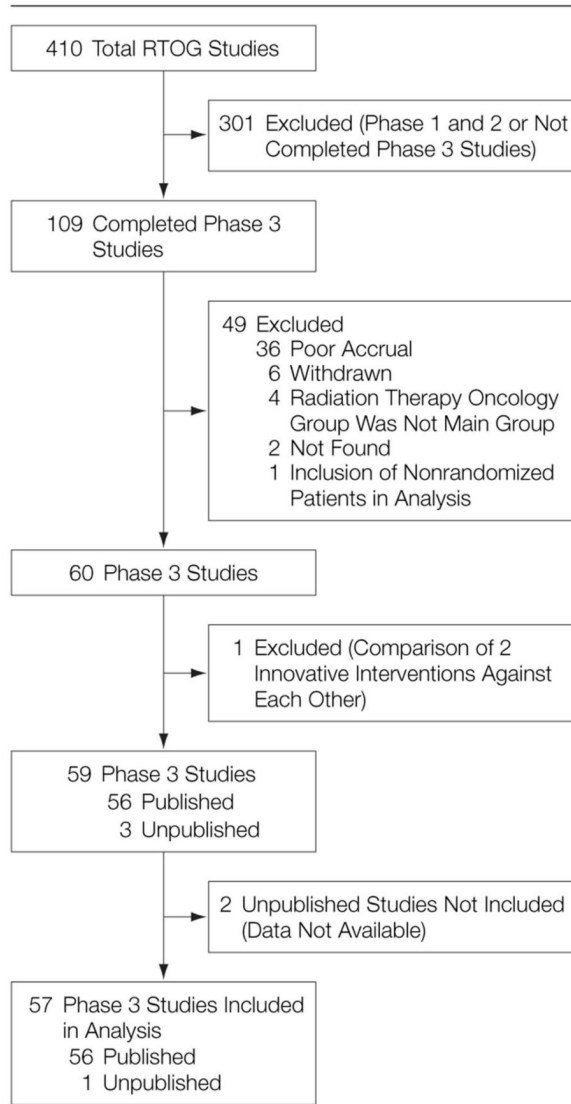
## REFERENCES

1. Jadad, A. *Randomized Controlled Trials*. BMJ Books; London, England: 1998.
2. Collins R, McMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001;357:373–380. [PubMed: 11211013]
3. Looking back on the millennium in medicine. *N Engl J Med* 2000;342:42–49. [PubMed: 10620649]
4. Chalmers I. What is the prior probability of a proposed new treatment being superior to established treatments? *BMJ* 1997;314:74–75. [PubMed: 9001503]
5. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional boards. *JAMA* 1992;267:374–378. [PubMed: 1727960]
6. Dickersin K. How important is publication bias? a synthesis of available data. *AIDS Educ Prev* 1997;9 (1 suppl):15–21. [PubMed: 9083596]
7. Schulz KF, Chalmers I, Hayes R, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–412. [PubMed: 7823387]
8. Altman D, Bland M. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485. [PubMed: 7647644]
9. Altman D. The scandal of poor medical research. *BMJ* 1994;308:283–284. [PubMed: 8124111]
10. Djulbegovic B, Lacey M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635–638. [PubMed: 10968436]
11. Djulbegovic B. Acknowledgment of uncertainty: a fundamental means to ensure scientific and ethical validity in clinical research. *Curr Oncol Rep* 2001;3:389–395. [PubMed: 11489238]
12. Djulbegovic B, Cantor A, Clarke M. The importance of preservation of the ethical principle of equipoise in the design of clinical trials: relative impact of the methodological quality domains on the treatment effect in randomized controlled trials. *Account Res* 2003;10:301–315. [PubMed: 14989285]
13. Kelahan AMCR, Marinucci, D. The history, structure, and achievements of the cancer cooperative groups. *Managed Care Cancer* 2001:28–33.
14. Procedure Manual RTOG. Radiation Therapy Oncology Group; Philadelphia, Pa: 2003.
15. Djulbegovic B. Association between competing interests and conclusions: denominator problem needs to be addressed. *BMJ* 2002;325:1420. [PubMed: 12480865]
16. NCI-CTEP The Investigator's Handbook <http://ctep.cancer.gov/handbook/index.html> Accessed May 3, 2004. Accessibility verified January 18, 2005
17. Alderson P, Green S, Higgins JPT, Cochrane Reviewers' Handbook 4.2.2 [updated December 2003] 2005 John Wiley & Sons Ltd. Chichester, England <http://www.cochrane.dk/cochrane/handbook/hbook.htm> The Cochrane Library, Issue 1, 2004
18. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363–374. [PubMed: 884196]



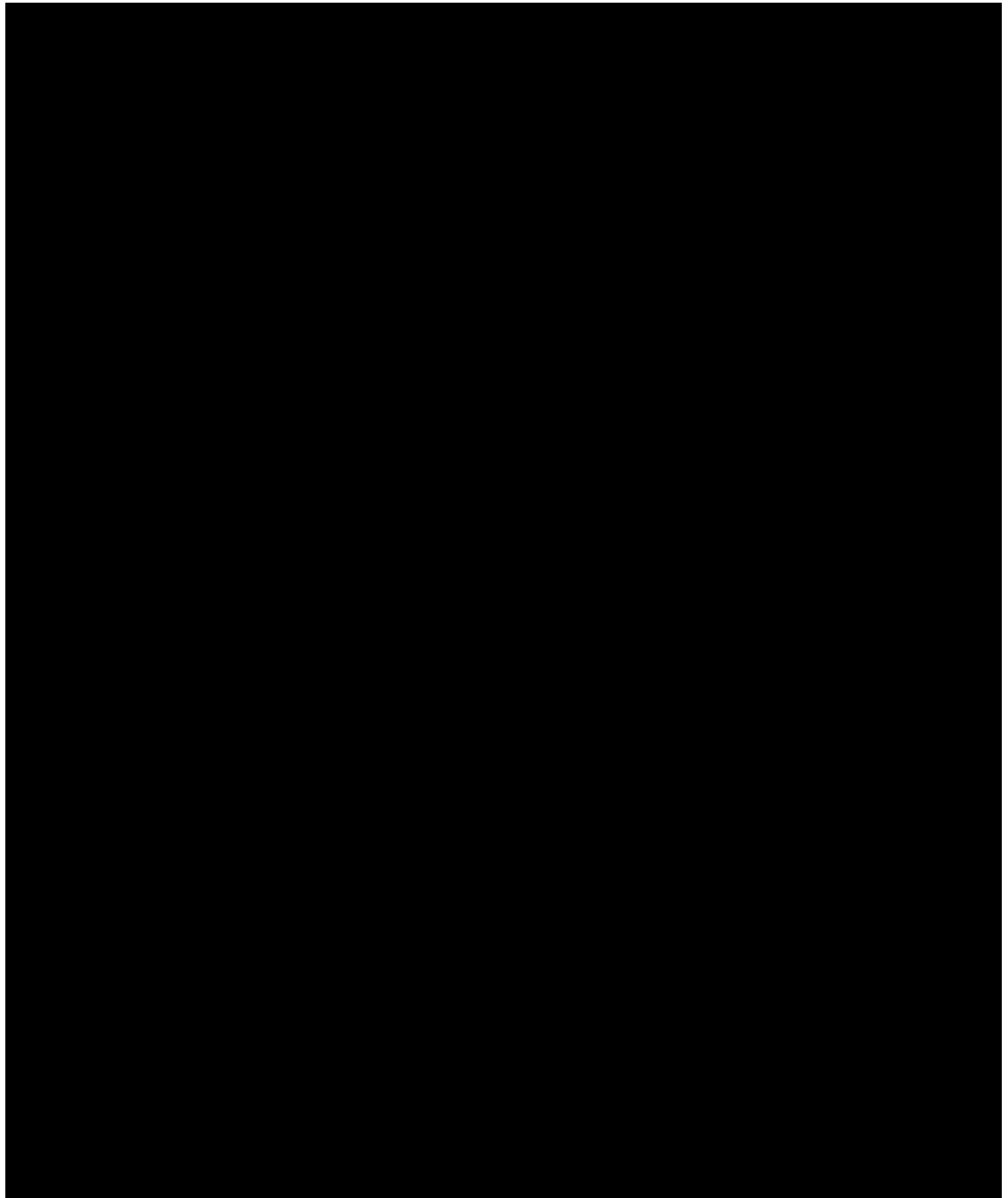
19. Moher D, Schulz KF, Altman D, Group CONSORT. (Consolidated Standards of Reporting Trials). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987–1991. [PubMed: 11308435]
20. Moher D, Jones A, Lepage L, CONSORT group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–1995. [PubMed: 11308436]
21. Egger, M.; Smith, DS.; Altman, D. *Systematic Reviews in Health Care: Meta-analysis in Context*. 2nd ed.. BMJ; London, England: 2001.
22. Soares HP, Daniels S, Kumar A, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 2004;328:22–24. [PubMed: 14703540]
23. Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA* 2001;285:1206–1208. [PubMed: 11231752]
24. Gilbert JP, McPeck B, Mosteller F. Statistics and ethics in surgery and anesthesia. *Science* 1977;198:684–689. [PubMed: 333585]
25. Colditz GA, Miller JN, Mosteller F. Measuring gain in the evaluation of medical technology: the probability of a better outcome. *Int J Technol Assess Health Care* 1988;4:637–642. [PubMed: 10291102]
26. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy, I: medical. *Stat Med* 1989;8:441–454. [PubMed: 2727468]
27. Machin D, Stenning S, Parmar M, et al. Thirty years of medical research council randomized trials in solid tumours. *Clin Oncol (R Coll Radiol)* 1997;9:100–114. [PubMed: 9135895]
28. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998;17:2815–2834. [PubMed: 9921604]
29. Early Breast Cancer Trialists Collaborative Group Introduction and Methods Section Reproduced From: *Treatment of Early Breast Cancer: Worldwide Evidence 1985-1990* Oxford University Press Oxford, England <http://www.ctsu.ox.ac.uk/publications/brecan90/> Accessed January 18, 2005
30. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995;48:557–560 [PubMed: 7853047]
31. Eddy D. Comparing benefits and harms: the balance sheet. *JAMA* 1990;263:2493–2505. [PubMed: 2329639]
32. SYSTAT [computer program]. Version 10.2. SPSS, Inc; Chicago, Ill: 2002.
33. *Review Manager* [computer program]. Version 4.2.2003 <http://www.cochrane.org/cochrane/revman.htm> The Cochrane Collaboration-available on the world wide web at <http://www.cochrane.org/cochrane/revman.htm>
34. Altman DG, Schulz KF, Moher D, CONSORT group. The revised CONSORT statement: explanation and elaboration. *Ann Intern Med* 2001;134:663–694. [PubMed: 11304107]
35. Mann H, Djulbegovic B. Choosing a control intervention for a randomised clinical trial. *BMC Med Res Methodol* 2003;3:7. [PubMed: 12709266]
36. Mann HD, Djulbegovic B. Comparator bias: why comparisons must address genuine uncertainties. [jameslindlibrary.org](http://www.jameslindlibrary.org/essays/bias/comparator_bias.html) [http://www.jameslindlibrary.org/essays/bias/comparator\\_bias.html](http://www.jameslindlibrary.org/essays/bias/comparator_bias.html) Accessed May 3, 2004. Accessibility verified January 24, 2005
37. Djulbegovic B, Clarke M. Informing patients about uncertainty in clinical trials [letter]. *JAMA* 2001;285:2713–2714.
38. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med* 1994;331:394–398. [PubMed: 8028622]
39. Alderson P. Absence of evidence is not evidence of absence. *BMJ* 2004;328:476–477. [PubMed: 14988165]
40. Chalmers I. Well informed uncertainties about the effects of treatments. *BMJ* 2004;328:475–476. [PubMed: 14988164]
41. Edwards SJ, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J. Ethical issues in the design and conduct of randomized controlled trials. *Health Technol Assess* 1998;2:1–130.
42. Peto R, Baigent C. Trials: the next 50 years. *BMJ* 1998;317:1170–1171. [PubMed: 9794846]

43. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141–145. [PubMed: 3600702]
44. Lilford RJ, Djulbegovic B. Equipoise and “the uncertainty principle” are not mutually exclusive. *BMJ* 2001;322:795. [PubMed: 11322146]
45. Juni, P.; Altman, DG.; Egger, M. Assessing the quality of randomised controlled trials. In: Egger, M.; Smith, GD.; Altman, DG., editors. *Systematic Reviews in Health Care: Meta-analysis in Context*. BMJ; London, England: 2001. p. 87-108.
46. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921–928. [PubMed: 12928469]
47. Kaptchuk TJ. Effect of interpretive bias on research evidence. *BMJ* 2003;326:1453–1455. [PubMed: 12829562]
48. Kuhn, T. *The Structure of Scientific Revolutions*. 2nd ed.. University of Chicago Press; Chicago, Ill: 1970.
49. Kuhn, T. Objectivity, value judgement, and theory choice. In: Klemke, ED.; Hollinger, R.; Rudge, DW., editors. *Introductory Readings in the Philosophy of Science*. Prometheus Books; New York, NY: 1998. p. 435-450.
50. Ioannidis JP, Evans SJ, Gotzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781–788. [PubMed: 15545678]

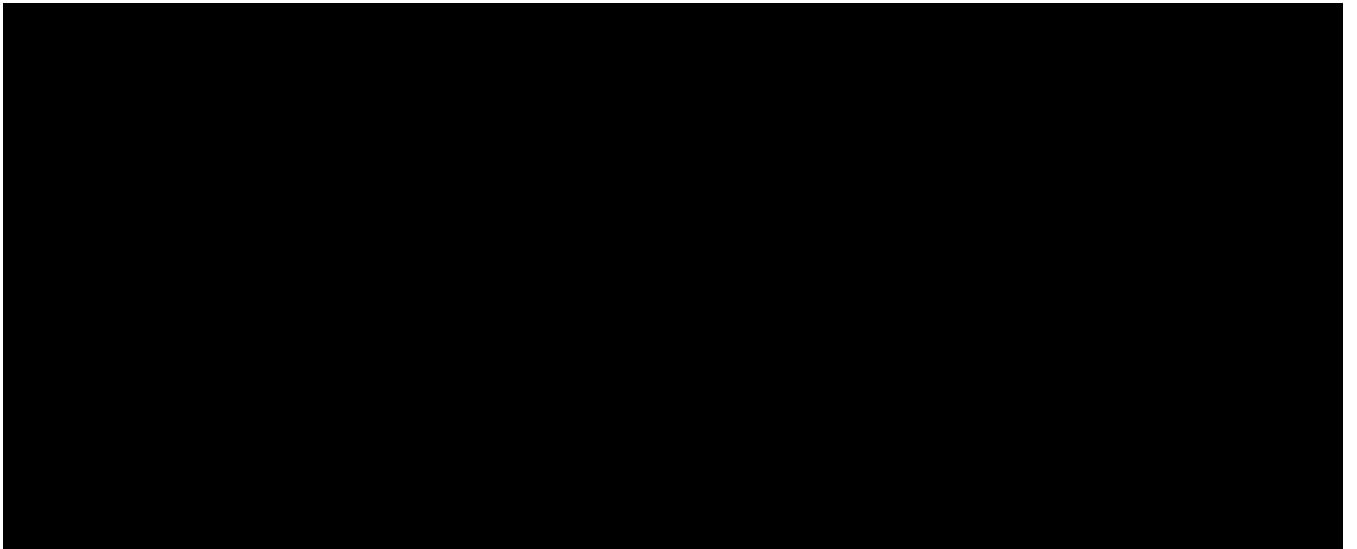


One trial that compared 2 experimental treatments (without inclusion of a standard treatment) was excluded from our analysis. Out of 3 trials that were unpublished, we were able to evaluate the outcomes from 1 of them. Therefore, overall analysis was performed using 57 trials.

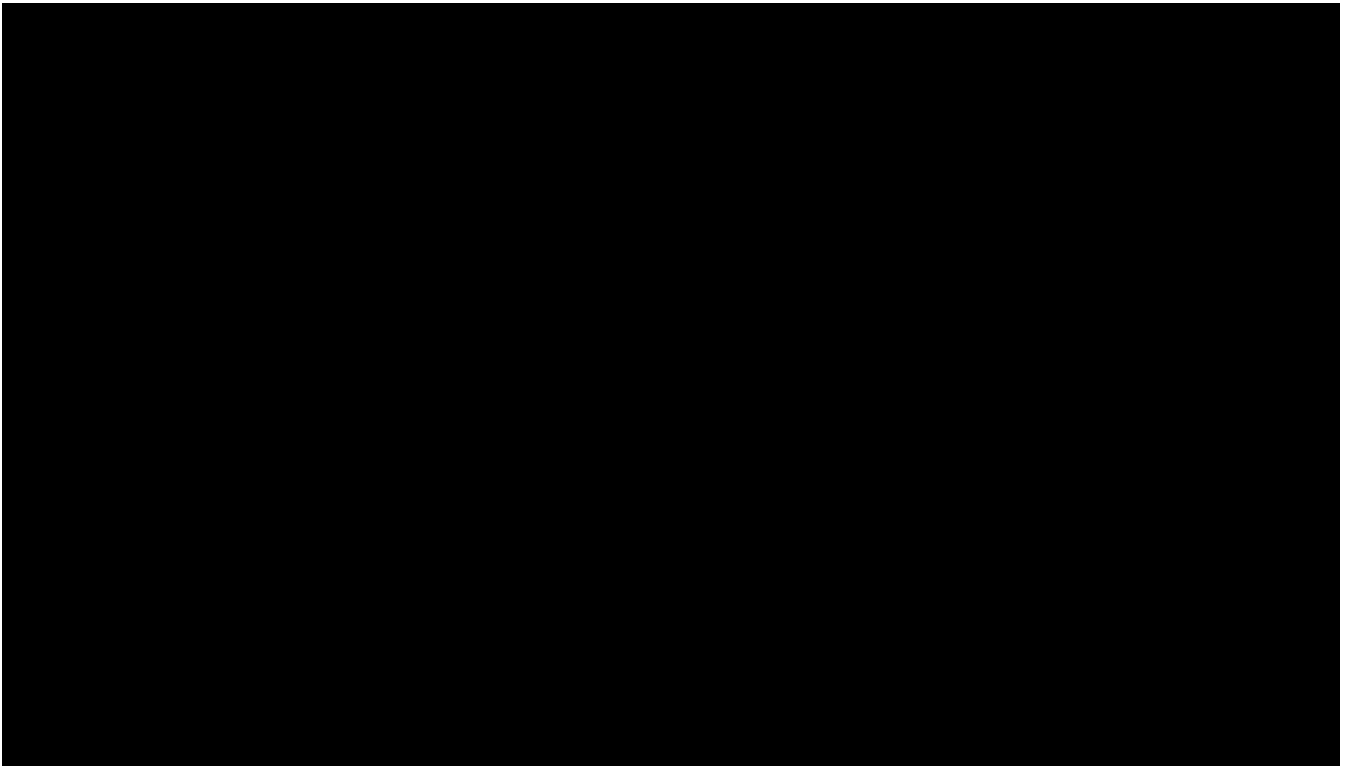
**Figure 1.** Publication Rate of the Completed Phase 3 Trials Performed by the Radiation Therapy Oncology Group



**Figure 2.**  
Overall Survival in All Radiation Therapy Oncology Group Trials That Examined Survival



**Figure 3.**  
Evaluation of Innovative Treatments in Radiation Therapy Oncology Group Trials: Main Outcomes



**Figure 4.**  
Sensitivity Analysis of Impact of Methodological Quality on Survival

**Table 1.**  
 Characteristics of the Randomized Controlled Trials Included in the Analysis

Variable	No.(%)of Trials(n = 57)
Disease	
Head and neck	16 (28)
Lung	12 (21)
Prostate	4 (7)
Metastases	9 (16)
Other	16 (28)
Type of treatment	
Adjuvant	6 (11)
Neoadjuvant	2 (3)
Palliative	13 (23)
Definitive	34 (60)
Other	2 (3)
Type of intervention	
Type or dose of radiation	34 (60)
Use of radiosensitizer	9 (16)
Radiotherapy + chemotherapy	8 (14)
Other	6 (10)
Study design	
Parallel	56 (98)
Crossover	1 (2)
Masking	
Double-blinding	2 (4)
Single-blinding	2 (4)
None	53 (92)
Type of comparison	
Active treatment	54 (95)
Placebo or no active treatment	3 (5)
Primary end point	
Survival	32 (56)
Disease-free survival or locoregional control	13 (23)
Best response	5 (9)
Other	7 (12)

**Table 2.**  
Methodological Quality of the Radiation Therapy Oncology Group Randomized Controlled Trials\*

Variable	No.(%) of Trials(n = 57)	
	Actual	Reported
Expected difference in outcomes prespecified		
Yes	45 (79)	7 (12)
No	8 (14)	...
Unclear or missing data	4 (7)	50 (88)
Expected $\alpha$ error prespecified		
Yes	45 (79)	6 (11)
No	8 (14)	...
Unclear or missing	4 (7)	51 (89)
$\beta$ Error prespecified		
Yes	44 (77)	6 (11)
No	9 (16)	...
Unclear or missing data	4 (7)	...
Sample size prespecified		
Yes	45 (79)	9 (16)
No	8 (14)	...
Unclear or missing data	4 (7)	48 (84)
Adequate allocation concealment (randomization done through central office)		
Yes	57 (100)	24 (42)
Unclear or missing	...	33 (58)
Dropouts described		
Yes	52 (91)	52 (91)
No	5 (9)	5 (9)
Intention-to-treat analysis		
Yes	48 (84)	41 (72)
No	7 (12)	6 (11)
Unclear	2 (4)	10 (17)

\* Quality of reporting vs actual methodological quality is based on information from research protocols and verification by the radiation therapy oncology group statistical headquarter.