Evaluating Patient-Centered Outcomes in the Randomized Controlled Trial and Beyond: Informing the Future with Lessons from the Past

Christopher M. Booth

Abstract

In the era of molecular oncology, patients still define a useful therapy as one that allows them to live longer and helps them to live better. Although patient outcomes have clearly improved as a result of randomized controlled trials (RCT), it is critical that contemporary trials retain the perspective of these fundamental patient-centered outcomes. Trends in study design, results, and interpretation of oncology RCTs from the past provide a useful framework in which to consider how the research community may approach trial design in the future. Although the RCT remains the standard for establishing efficacy, this article also considers how population-based outcome studies can provide insight into effectiveness of new therapies and explores how the results of RCTs translate into benefit in the general population. *Clin Cancer Res; 16(24); 5963–71. ©2010 AACR.*

Although patient outcomes have improved as a result of clinical trials conducted since the 1970s, in the contemporary era, patients still define a useful therapy as one that improves the quality, or increases the quantity, of survival. Despite considerable change in the treatment and outcome of patients with cancer, it is critical that investigators, clinicians, and policy makers not lose sight of these fundamental patient-centered outcomes. Furthermore, although the randomized controlled trial (RCT) remains the gold standard for establishing efficacy of new therapies, increased effort is required to understand how the results of RCTs influence clinical practice and to assess the benefit of new treatments in the general population. This article explores how the RCT in oncology has evolved over time and proposes issues that require consideration in the design and interpretation of future trials. The article also considers how population-based outcome studies can provide insight into the uptake of new medical therapies and explores how the results of RCTs translate into benefit in the “real world.”

Evolution of Design and Outcomes of Oncology Randomized Controlled Trials

A recent overview of 321 RCTs of systemic therapy in breast, colorectal, and non–small cell lung cancer (NSCLC), published between 1975 and 2004, provides insight into the methods, results, and outcomes of oncology trials over time (Table 1; ref. 1). RCTs in oncology have become larger and increasingly involve multiple international centers. Although the median sample size of RCTs has increased from 100 to 446 patients per study, the total duration of study accrual for these trials has remained relatively stable since the 1970s. During the study period, encouraging changes occurred in the statistical methodology and endpoints in RCTs. The primary endpoint of oncology RCTs has shifted away from response rate to survival and other time-to-event endpoints. The proportion of trials using intention-to-treat (ITT) analyses has also increased. Despite these improvements, in the most recent decade (1995 to 2004), one third of systemic therapy RCTs published in high-impact and widely read journals did not explicitly state the primary endpoint of the study and only half of trials report ITT analyses of all randomized patients.

Evolution of Design and Outcomes of Oncology Randomized Controlled Trials

Although patient outcomes have improved as a result of clinical trials conducted since the 1970s, in the contemporary era, patients still define a useful therapy as one that improves the quality, or increases the quantity, of survival. Despite considerable change in the treatment and outcome of patients with cancer, it is critical that investigators, clinicians, and policy makers not lose sight of these fundamental patient-centered outcomes. Furthermore, although the randomized controlled trial (RCT) remains the gold standard for establishing efficacy of new therapies, increased effort is required to understand how the results of RCTs influence clinical practice and to assess the benefit of new treatments in the general population. This article explores how the RCT in oncology has evolved over time and proposes issues that require consideration in the design and interpretation of future trials. The article also considers how population-based outcome studies can provide insight into the uptake of new medical therapies and explores how the results of RCTs translate into benefit in the “real world.”

Evolution of Design and Outcomes of Oncology Randomized Controlled Trials

A recent overview of 321 RCTs of systemic therapy in breast, colorectal, and non–small cell lung cancer (NSCLC), published between 1975 and 2004, provides insight into the methods, results, and outcomes of oncology trials over time (Table 1; ref. 1). RCTs in oncology have become larger and increasingly involve multiple international centers. Although the median sample size of RCTs has increased from 100 to 446 patients per study, the total duration of study accrual for these trials has remained relatively stable since the 1970s. During the study period, encouraging changes occurred in the statistical methodology and endpoints in RCTs. The primary endpoint of oncology RCTs has shifted away from response rate to survival and other time-to-event endpoints. The proportion of trials using intention-to-treat (ITT) analyses has also increased. Despite these improvements, in the most recent decade (1995 to 2004), one third of systemic therapy RCTs published in high-impact and widely read journals did not explicitly state the primary endpoint of the study and only half of trials report ITT analyses of all randomized patients.

The funding of clinical trials has also changed remarkably over the past four decades. In the 1970s and early 1980s, government funding provided support for most RCTs in oncology, and industry was responsible for funding approximately 4%. In the most recent decade, more than half of RCTs (57%) were funded by industry, whereas the proportion of studies funded by government support has decreased substantially (1).

The concept of “benefit” and what constitutes a “positive” clinical trial is complex and involves variables from several interrelated domains. However, at least three important variables merit specific comment. The first is
the magnitude of the benefit conferred by the new therapy: To what degree is the new drug better than the old drug, and what is that benefit (i.e., effect size)? The second concept is the level of statistical significance one can assign to the observed difference in outcome between treatment arms. The third variable that may influence whether clinicians consider a trial “positive” or “negative” is what the study authors say in their summary of the study results (i.e., author conclusions). Author conclusions themselves are highly complex but may take into account variables such as the effect size, P-value, quality of life data, treatment toxicity, and financial costs. Furthermore, other more subtle influences may lead to distorted presentation of study results or “spin.” In a recent overview of 72 RCTs with statistically nonsignificant results for the primary endpoint, Boutron and colleagues found evidence of “spin” in more than 40% of articles (2).

As shown in Table 2, several insightful findings have emerged from tracking the results and interpretation of oncology RCTs since the 1970s. First, the effect size (i.e., the relative benefit of new drug versus old drug) across individual trials has remained stable over time. Second, a...
considerably greater proportion of RCTs has results that are statistically significant, likely directly related to the large increase in sample size. Third, despite effect size remaining stable since the 1970s, authors of modern RCTs are considerably more likely to call their trial "positive" and strongly endorse the experimental arm. Although sponsorship status is not associated with increased effect size, industry funded RCTs are more likely to strongly endorse novel treatments. Multivariate analyses suggest that independent predictors of endorsing the experimental therapy are significant $P$-values for primary endpoint, time-to-event primary endpoint, industry sponsorship, and effect size (1).

**How Can We Improve Randomized Controlled Trials?**

Trends in RCTs over the past four decades provide a framework for discussing aspects of trial design and interpretation that may be improved as we look to the next generation of clinical trials. Although an exhaustive overview of specific issues is beyond the scope of this article, this section explores common themes that emerge from related literature (Fig. 1).

**Appropriate design of clinical trials**

As described elsewhere in this CCR FOCUS series by LoRusso and colleagues (3), one aspect that continues to plague oncology drug development is the inefficiency of the phase II to phase III transition. In an overview of 100 phase II trials with promising results presented at the American Society of Clinical Oncology (ASCO) 1995 to 1996, Berthold and colleagues found that one decade later, only 13 phase II studies had moved to the phase III setting (4). In a survey of promising contemporary phase II trials, the primary reasons for not moving forward related to limited financial and patient resources and inability to access the drug. In the design and implementation of oncology trials, careful consideration must be given to how valuable resources for clinical research (i.e., patients, drug, and financial) are used.

In the era of targeted cancer therapy, an important challenge to trial design is to identify which subset of patients is likely to benefit from a novel therapy on the basis of molecular characteristics at the tumor level. Given the vast number of molecular agents that have been studied in RCTs, the list of therapies that have conferred major benefit to patients is remarkably short (5, 6). It is likely that one of the reasons we have not observed more substantial improvements in outcome is that the treatment effect is being diluted by patients who do not express the molecular target of interest. RCTs in breast cancer have successfully included specific subsets of patients on the basis of tumor characteristics for many years, dating back to early trials of hormonal therapy (7–9), and more recently for trastuzumab (10). On the contrary, the development of bevacizumab for advanced colorectal and breast cancers provides a telling example of what may happen when an appropriate drug and tumor target relationship is not considered in determining the study population of interest. Although bevacizumab may offer meaningful benefit to a molecular subtype of patients with these diseases, to date the pivotal trials of this agent in breast and colorectal cancer have included unselected patients and have been associated with modest improvements in patient outcomes (11–13). The benefit of these therapies in unselected populations is further reduced when one considers the substantial financial costs and toxicities associated with treatment. As discussed elsewhere in this CCR FOCUS series by Dalton and colleagues, it is very likely that

<table>
<thead>
<tr>
<th>Table 2. What constitutes a &quot;positive&quot; randomized controlled trial?</th>
<th>trends over time in results and interpretation of benefit in oncology randomized controlled trials 1975 to 2004</th>
<th>Results and Interpretation of RCTs</th>
<th>1975 to 1984</th>
<th>1985 to 1994</th>
<th>1995 to 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Effect size: to what degree is the new therapy better than the standard of care?</td>
<td>Median HR (95% CI)</td>
<td>1.4 (1.0–2.3)</td>
<td>1.2 (1.0–2.4)</td>
<td>1.2 (1.1–1.3)</td>
<td></td>
</tr>
<tr>
<td>2. Statistical significance: is the observed magnitude of difference between arms statistically significant?</td>
<td>$P &lt; 0.05$ for primary endpoint</td>
<td>23%</td>
<td>30%</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>3. Author conclusions: what do the authors say in their summary of the trial results?</td>
<td>Author strong endorsement of experimental arm*</td>
<td>31%</td>
<td>39%</td>
<td>49%</td>
<td></td>
</tr>
</tbody>
</table>

*Author conclusions were scored on the basis of a statement in the study abstract using a modified 7-point Likert Scale as described by Ridker and colleagues (63).

Data from Booth et al (1).
Biomarkers will play an increasingly important role in the era of personalized medicine (16).

The increasing sample size of clinical trials (1) puts further strain on a system in which patient accrual is already suboptimal in many environments. As discussed in a recent report from the Institute of Medicine (17), no more than 5% of patients with cancer are entered onto a clinical trial and a considerable portion of RCTs do not meet accrual success (18). Although many of these trials may close for reasons other than accrual, it is possible that large sample sizes required by RCTs may impair the ability of investigators to answer clinically important questions.

Clinically relevant endpoints

The ultimate goal of all cancer research is to improve outcomes for our patients. Although outcomes have been traditionally evaluated using endpoints such as tumor size and length of overall survival, increasing emphasis is being placed on quality of life and other patient-reported outcomes. In recent years, important changes in the management of some diseases have been based primarily on improvements in these domains (19, 20). For this reason, it is critical that patient-centered endpoints be evaluated and interpreted with the same level of rigor as traditional endpoints.

In their overview of 112 RCTs in advanced cancer in which quality of life was evaluated (21), Joly and colleagues found that only 19% (21/112) established an a priori hypothesis with respect to quality of life and that only 21% (24/112) found a minimal difference in quality-of-life change score that showed any clinical benefit. Parallel to this is the increasing use of surrogates of overall survival such as progression-free or disease-free survival. Although these endpoints have been shown to correlate with improvement in quantity and/or quality of life for patients in select diseases (22–25), it is of concern that surrogate endpoints are increasingly used in diseases for which little validation exists (26–28).

One of the greatest challenges in interpreting the results of oncology RCTs is distinguishing results that are clinically important from those that are statistically significant (29, 30). With the remarkable increase in RCT sample size and the resulting change in statistical power, it has become possible to detect very small differences in outcomes.
between treatment arms, which are rendered appealing by virtue of a *significant* P-value. In fact, it is quite likely that therapies evaluated in older trials associated with effect sizes that we now would consider adequate [i.e., hazard ratio (HR) = 0.80] may have been abandoned because of greater expectations in the past. An illustrative example of this is shown in Fig. 2, which depicts survival curves from two trials of patients with advanced NSCLC. The older trial by Cellerino and colleagues (31) accrued patients in 1985 to 1988, whereas the more recent Eastern Cooperative Oncology Group (ECOG) 4599 trial by Sandler and colleagues (32) accrued from 2001 to 2004. Taking into account the different scales of the horizontal axis, on first review, the difference between treatment arms in each study appears comparable. Due to a substantially smaller sample size, the observed difference in the Cellerino trial was not significant (HR 0.78, P = 0.153), whereas the same effect size in ECOG 4599 was significant (HR 0.79, P = 0.003) and led to regulatory approval of bevacizumab for advanced NSCLC. However, if the Cellerino trial had been repeated with a sample size comparable to ECOG 4599, it is very likely that the observed difference would have been statistically significant. A, the Cellerino trial compared alternating chemotherapy versus best supportive care (BSC) and was powered to show an improvement in survival rate from 25 to 50% at 1 year (HR 0.5). To detect this improvement with 5% two-sided α and 80% power required 65 deaths. The trial enrolled 128 patients. Observed median survival was 8 months in the experimental group and 5 months in the control group (P = 0.153). In describing the survival curve, the authors state that “the graphic comparison of the two intervals clearly indicates no significant difference in survival between the treatments.” The observed HR for 1-year survival (estimated from the curve to be 33% and 25% for experimental and control groups, respectively) is 0.78. To have adequate power to detect this magnitude of difference would have required 492 deaths. Reprinted with permission from Cellerino et al. (31). Copyright © 2008 American Society of Clinical Oncology. All rights reserved. B, the Sandler trial (ECOG 4599) compared paclitaxel-carboplatin with bevacizumab (BPC) to paclitaxel-carboplatin (PC) alone and was powered to detect a HR for death of 0.80 with a one-sided α of 2.5% and 81% power. To detect this improvement required 650 deaths. The trial enrolled 878 patients. Observed median survival in the experimental group was 12 months and 10 months in the control group (HR 0.79, P = 0.003). Reprinted with permission from Sandler et al (32). Copyright © 2006 Massachusetts Medical Society. All rights reserved.

Although small improvements should not by themselves change patient care, in some settings, these small steps may lead to substantial cumulative benefits over time. The sequential use of systemic therapy regimens in metastatic colon cancer is one such example in which results from cumulative RCTs have led to a major improvement in the outcomes of patients (33). In light of this paradox, Sobrero and Bruzzi have proposed a novel drug approval process in which drugs associated with modest benefit are granted provisional approval for further study and become eligible for full approval should a series of further studies show a cumulative and meaningful impact on patient outcomes (5). This new process may address some of the challenges of conducting definitive RCTs when the drug of interest has already received regulatory approval on “lesser” evidence. Recognition is growing for the need to distinguish RCTs that are explanatory in nature from those that are pragmatic (34). Explanatory RCTs are conducted in tightly controlled settings with highly selected patients and are designed to explore the efficacy or biologic proof of concept of a novel therapy, often using short-term surrogate outcome measures. Results from these trials are generally used to inform further research rather than directly influence clinical practice. The initial randomized phase II trial of bevacizumab in advanced colorectal cancer (33) is one example of an explanatory trial. Despite the results showing improved survival in the experimental arm, the trial was designed and interpreted in the context of how to best inform subsequent
studies of bevacizumab rather than directly influencing clinical practice. The pivotal RCT of erlotinib in advanced pancreatic cancer (15) is an example of an explanatory phase III trial; although the study met its primary endpoint, the results were presented in the context of informing further study rather than directly guiding patient care.

At the other end of the continuum, pragmatic trials test the effectiveness of an intervention in a setting that more closely resembles routine practice, with outcomes that have direct relevance to clinicians, patients, and policy makers. The results of pragmatic trials are to be used directly by those involved in providing patient care. MRC CR07/ NCIC-CTG CO16 RCT (36) was a large pragmatic trial designed to evaluate the impact of preoperative radiotherapy on outcomes of patients with resectable rectal cancer. Patients were enrolled from 80 centers across five countries. Surgery was guided by local policy without central training. The study results showed a clear role for short-course radiotherapy among patients with rectal cancer in the general population. Identifying whether a RCT is pragmatic versus explanatory in nature may provide a framework for stakeholders in determining how the results of the study can and/or should be used.

Reporting of trials and avoidance of bias

As we strive to translate the results of oncology RCTs into clinical practice, it is essential that language be used that will allow clinicians and patients to clearly understand the potential benefit of a specific therapy. Elsewhere in this CCR FOCUS series, Smith and Hillner (37) provide a framework for having honest discussions with our patients. Saltz has recently proposed changes to common terminology used in the reporting of clinical trials (30). He emphasized the need to distinguish results that are only statistically significant from those that are clinically meaningful. He also suggested that commonly used phrases such as progression-free "survival," reduction of the "risk of death," and "tolerable" toxicity be reconsidered in an effort to improve clarity. Our own group recently described use of "clinical benefit" as an endpoint since it was originally described as a composite measure of pain, performance status, and weight among patients with pancreas cancer (38). Since the pivotal trial of gemcitabine in advanced pancreas cancer reported this outcome in 1997 (20), 71 trials reported in the Journal of Clinical Oncology have included a clinical benefit endpoint. Among these articles, 72% used "clinical benefit" to describe changes in objective tumor measurements (38). Whether such patients experience true clinical benefit depends on whether they also have improvement in the duration and/or quality of survival.

How and when the results of RCTs are disseminated is receiving increasing attention. It has become common for RCTs to terminate early on the basis of the results of interim analyses. Korn and colleagues recently reviewed 27 such trials conducted by National Cancer Institute (NCI) cooperative groups (39): Follow-up reports of these 27 trials suggest that the initial results and conclusions are preserved with longer follow-up. Termination based on a planned interim analysis is quite different to informal reporting of preliminary analyses at major cancer meetings. Preliminary data should not be made public for epidemiologic and statistical reasons (40, 41), which primarily relate to the statistical instability of early results and the potential for nonfinal analyses to misinform clinical practice. In a recent overview of 138 oncology RCTs published between 2000 and 2004, 44% of related conference abstracts used language to imply that the data were nonfinal and 63% of abstracts contained important data discordance when compared with the final publication (42). Furthermore, when compared with published articles, authors’ conclusions were substantively discordant in 10% of abstracts. Accordingly, presentation of nonfinal results should be discouraged and clinicians should be cautious when interpreting immature data from clinical trials.

In the past decade, two sources of potential bias in oncology clinical trials have been described: publication bias (the preferential publication of RCTs with "positive results"; ref. 43) and sponsorship bias. It is hoped that publication bias will diminish with mandatory trial registration (44) and an increasing trend among major journals to publish negative trials. Although several groups have found that industry-sponsored studies are more likely to be positive than those not sponsored by industry, the reasons for this are likely to be complex and multifactorial (1, 45, 46). Possible explanations for this observation include: industry having a different threshold for determining whether to take a novel therapy to phase III study; differences in trial design and selection of the control arm; publication bias; and selective interpretation of trial results (2, 47). As consumers of medical literature, both clinicians and policy makers should bear publication and sponsorship biases in mind when appraising the evidence.

Assessing Outcomes of New Cancer Therapies in the "Real World"

To better evaluate the impact of cancer therapies in the general population, it is necessary to look beyond the traditional clinical trial paradigm. Although the RCT offers excellent internal validity, because patients, physicians, and concurrent care in the general population may be very different from the controlled context of a clinical trial (48), it is essential that we not assume that the benefits shown in RCTs will be fully realized at the population level.

Recently, innovative study designs and access to large electronic databases have allowed investigators to evaluate the impact of new medical therapies in the "real world" (49, 50). These population-based outcome studies may provide insight into: uptake of new therapy, the impact of a change in practice and/or policy on outcomes, and the societal benefit of medical research. Furthermore, as discussed elsewhere in this CCR FOCUS series by Schnipper and colleagues, consideration of economic aspects of cancer care from a population perspective are urgently needed (51). Population-based studies also provide a mechanism to identify
gaps in care following publication of a pivotal RCT and can facilitate targeted efforts in knowledge translation. Investigators can also learn about toxicities and outcomes associated with therapies when they are given to patient groups that are often underrepresented in RCTs (i.e., the elderly and those with other medical conditions). Important limitations to this study design include the possibility that changes in outcome at the population level may reflect confounding by other changes in treatment and/or case mix. However, given the unique insights they provide outside of the controlled environment of clinical trials, population-based outcome studies are becoming increasingly recognized as important in the generation and dissemination of medical evidence (50, 52–54). Two recent examples of such studies are described below.

In the past decade, pivotal RCTs have established clear roles for chemotherapy given concurrently with radiation in cervical cancer (55) and adjuvant chemotherapy in NSCLC (56, 57). Given the age and comorbidity of patients enrolled in clinical trials, concern was expressed about whether the RCT results were generalizable to the general population.

Following the NCI announcement recommending that "strong consideration should be given to adding chemotherapy to radiation therapy in the treatment of invasive cervical cancer" (55), Pearcey and colleagues undertook a population-based cohort study of the management and outcome of cervical cancer in Ontario (58). Using linked administrative and cancer registry databases, Pearcey and colleagues identified all cases of cervical cancer diagnosed in Ontario between 1992 and 2001 (N = 4,069). Prior to the NCI announcement, less than 10% of women treated with chemotherapy increased rapidly, and from 1999 to 2001, more than 60% of such women received chemotherapy. The authors found that adoption of concurrent chemotherapy was associated with an 11% improvement in 3-year overall survival (P < 0.01). This survival benefit was virtually identical to that which would have been expected on the basis of the results of the clinical trials. Furthermore, the rate of hospitalizations during treatment did not increase, suggesting that the new treatment did not add substantial toxicity.

Using similar methodology, our group has recently explored uptake of adjuvant chemotherapy for NSCLC (59). Using the Ontario Cancer Registry linked to electronic records of treatment, we identified all 6,304 cases of NSCLC in Ontario that underwent surgical resection. During the study period, the proportion of surgical cases receiving adjuvant chemotherapy increased from 7 (2001 to 2003) to 31% (2004 to 2006). We also found a substantial improvement in survival among patients with resected NSCLC at the population level, consistent with the relevant clinical trials. Over the study period, there was no suggestion of major chemotherapy-related toxicity as measured by rates of hospitalization in the first 6 months after surgery.

These examples highlight how population-based outcome studies may complement knowledge gained from RCTs and serve to identify whether advances from clinical trials translate into meaningful benefit in the "real world"; this synergistic relationship is very much aligned with the principles of comparative effectiveness research (60). Furthermore, in an era when investment in health research is being increasingly scrutinized, these studies provide critical information related to the "research payback" of investments made in clinical trials (61).

Conclusion

Since the 1970s, large improvements have been made in the management and outcome of patients with cancer (62). The design and conduct of RCTs are responsible for many of these advances. As we look forward to the next generation of clinical trials, it will be critical to keep patient-centered outcomes at the forefront of trial design. Clear communication of study methodology and benefit of therapy is essential to convey to patients and other stakeholders the benefit of novel therapies. Finally, as we look to build on the advances made through RCTs, population-based outcome studies are needed to provide complementary data to understand how practice changes after publication of landmark RCTs and whether new medical therapies benefit patients in the "real world."

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Drs. Ralph Meyer, Dongsheng Tu, Ian Tannock, Patricia LoRusso, and Steven Averbuch are gratefully acknowledged for thoughtful comments provided on an earlier version of this article.

Grant Support

C.M. Booth is supported as a Cancer Care Ontario Chair in Health Services Research.

Received 08/16/2010; revised 09/17/2010; accepted 10/13/2010; published online 12/15/2010.

References


Evaluating Patient-Centered Outcomes in the Randomized Controlled Trial and Beyond: Informing the Future with Lessons from the Past

Christopher M. Booth

*Clin Cancer Res* 2010;16:5963-5971.

**Updated version**
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/16/24/5963

**Cited articles**
This article cites 59 articles, 31 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/16/24/5963.full.html#ref-list-1

**Citing articles**
This article has been cited by 12 HighWire-hosted articles. Access the articles at:
/content/16/24/5963.full.html#related-urls

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.