Randomized controlled trials in the era of molecular oncology: methodology, biomarkers, and end points

A. Kay¹,², J. Higgins¹, A. G. Day², R. M. Meyer¹,² & C. M. Booth¹,²*
¹Division of Cancer Care and Epidemiology, Queen’s University Cancer Research Institute, Kingston; ²Kingston General Hospital, Kingston, Canada

Background: We previously reported metrics of systemic therapy randomized controlled trials (RCTs) in breast cancer, colorectal cancer (CRC), and non-small-cell lung cancer (NSCLC) published 1975–2004. To evaluate trends in the era of targeted therapies (TT), we have repeated a similar analysis of RCTs published 2005–2009.

Methods: A search for phase III RCTs of systemic agents published in five major journals 2005–2009 was carried out. Trials were classified as TT if they involved any non-hormonal targeted agent. We extracted data regarding biomarker use. Integral biomarkers were defined as tests used to determine eligibility, stratification, or allocation. Descriptive statistics were used to analyze trends over time.

Results: One hundred and thirty-seven eligible RCTs were evaluated. Compared with 1995–2004, the number (17–27 RCTs/year) and size (median sample size 446–722, P < 0.001) of RCTs increased. The proportion of RCTs evaluating TT increased from 4% (7/167) to 29% (40/137) (P < 0.001). There was an increase in the proportion of trials with financial support from industry [57% (95/167) to 78% (107/137), P = 0.001]. Biomarkers were included in 58% (80/137) of RCTs; integral biomarkers were included in 36% (49/137) of trials. Among the 49 RCTs using integral biomarkers, 40 (82%) used HER2 and/or ER/PR status in studies of breast cancer.

Conclusions: RCTs published in 2005–2009 are larger, more likely to evaluate TT, and be supported by industry. Biomarkers may be increasingly used, but the most common use relates to traditional use of ER/PR and evolving use of HER2 in breast cancer RCTs.

Key words: biomarkers, chemotherapy, clinical trials, end points, oncology, outcomes

introduction

We have previously reported the metrics of randomized controlled trials (RCTs) testing systemic therapies for patients with breast cancer, colorectal cancer (CRC), and non-small-cell lung cancer (NSCLC) published in major journals between 1975 and 2004 [1]. Over the study period of three decades, RCTs became larger with an increase in time-to-event measures as the primary end point and a parallel reduction in the use of response rate. The proportion of RCTs considered positive (defined as strong endorsement of the experimental arm by study authors) also increased. Independent predictors of a positive interpretation included: significant P-value for primary end point, time-to-event end point, funding support from industry, and effect size. Only 4% of these trials included targeted therapies (TT).

With the advent of agents directed at specific molecular targets, we hypothesized that RCTs in oncology would become more complex with increasing use of biomarkers and surrogate end points for survival. To address these issues, we designed the current study using similar methodology to our prior overview to provide a comprehensive review of RCTs in breast cancer, CRC, and NSCLC published between 2005 and 2009. Our objectives were to describe trends in: (i) trial methodology and reporting, (ii) selection of primary end points, and (iii) use of targeted agents and biomarkers. From this overview, we expect to gain insight about the interpretations and directions of clinical trials in the era of molecular oncology.

methods

search strategy

A search was undertaken for all RCTs of systemic therapy in breast, CRC, and NSCLC published between 1 January 2004 and 31 December 2009 in the following journals: Journal of Clinical Oncology, Journal of the National Cancer Institute, New England Journal of Medicine, Lancet, and Journal of the American Medical Association. These journals were selected because they were considered to contain a high proportion of widely read and practice-changing clinical trials in oncology. Furthermore, RCTs published in these journals formed the basis of our previous work [1] and were therefore included in the current study to describe trends over time. As with our previous study, indexes and tables of contents of these journals were reviewed electronically to find relevant articles. Exclusion criteria included: studies of a radiation and/or surgical intervention; studies of cancer screening and prevention; articles presenting data from multiple RCTs; studies comparing the same drug(s) given by different dose, route or
schedule; multiple reports of the same study (the first final report in a journal we reviewed was included); phase II or ‘pilot’ studies; and articles that did not report efficacy results or presented results only for a subgroup of the original study population.

data abstraction
The previously designed data abstraction form and data manual were used to capture information regarding study methodology, industry funding support, and results. Sections related to TT and use of biomarkers were incorporated into the revised forms. These tools were piloted on 10 RCTs by two authors (AK and CMB); results were compared and the forms then underwent final revisions. To ensure consistency of the abstraction process, all eligible studies were reviewed by a single author (AK), using a data manual as a guide.

Date of publication, disease site, intervention, and study setting were recorded. Country of study origin was assigned based on the institutional affiliation of the first author. Consistent with our previous work [1], source of financial support was determined based on explicit statements in the article. In the absence of an explicit statement regarding funding support, RCTs with author(s) whose affiliation was a pharmaceutical company were classified as industry supported. We captured use of biomarkers in each trial. A biomarker was defined as ‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention’ [2]. Biomarkers were further classified as integral if the tests used were to determine eligibility, stratification, or allocation and integrated if they tested hypotheses [3].

The primary end point of each study was identified; if there was no explicit statement, the end point implied to be of primary importance was recorded. We evaluated use of intention-to-treat analysis based on what was stated in the manuscript.

statistical analysis
Trends in RCT characteristics were compared across study periods (2005–2009 versus 1995–2004). Study size was compared between periods by the Wilcoxon–Mann–Whitney test; all other characteristics were categorical and compared by Fisher’s exact test. Multiple logistic regression using backward stepwise selection with entry and exit criteria of $P < 0.15$ was used to identify factors that were independently associated with the use of overall survival (OS) as the primary end point. Factors that were considered in this model included: breast cancer versus CRC/NSCLC, palliative versus adjuvant therapy, cooperative group trial, industry support, and use of targeted agent. All $P$-values are two-sided without adjustment for multiple testing. All analyses were performed in SAS 9.2 (SAS Institute inc., Cary NC).

results
The initial search yielded 285 studies, of which 137 were eligible (Figure 1). The final cohort of trials comprised 162 675 randomized patients.

RCT design
Characteristics of the study cohort are shown in Table 1. For comparative purposes, we also show results from our previously published overview of RCTs testing systemic therapies from the same journals in the preceding decade (1995–2004) [1]. In the 2005–2009 study cohort, 50% (69/137) of trials included patients with breast cancer; RCTs of CRC and NSCLC comprised 23% (32/137) and 26% (36/137) of the

cohort, respectively. Most trials were multicentre (99%, 135/137) and international in scope (61%, 84/137); 55% (75/137) of studies involved cooperative trials groups. Seventy-eight percent (107/137) of trials were supported by industry.

Oversight by a Data Monitoring Committee (DMC) was reported in 52% (71/137) of trials and 21% (29/137) were terminated early. Reasons for early study closure included: results of interim analyses (59%, 17/29), inadequate accrual (31%, 9/29), and release of results from a related trial (10%, 3/29).

Compared with the prior decade (1995–2004), the mean number of RCTs reported per year has increased from 17 to 27. While median sample size has increased markedly (446–722 patients, $P < 0.001$), there has been no substantial change in the time of study accrual. Use of targeted agents has increased from 4% to 29% (7/167 versus 40/137, $P < 0.001$). The proportion of RCTs supported by industry has increased from 57% to 78% (95/167 versus 107/137, $P < 0.001$). Fifty of the 107 RCTs supported by industry were cooperative group clinical trials. Finally, there was a trend toward more trials in 2005–2009 reporting a statistically significant result for the primary end point compared with 1995–2004 [53% (73/137) versus 42% (70/167), $P = 0.051$].

primary end point of RCTs
Time-to-event primary end points were used in 90% (123/137) of RCTs in 2005–2009 compared with 75% (125/167) of trials in the preceding decade ($P = 0.001$) (Table 2). Among these trials, contemporary trials were less likely to use OS [41% (50/123) versus 66% (82/125), $P < 0.001$] and more likely to use disease-free survival [21% (26/123) versus 13% (16/125), $P = 0.092$]. Trials using OS as the primary end point were less likely to report a statistically significant result compared with trials using other time-to-event end points [32% (16/50) versus 70% (51/73), $P < 0.001$].

Among the 123 trials with a time-to-event primary end point, we found that breast cancer [odds ratio (OR) 0.03, 95% confidence interval (CI) 0.01–0.12, $P < 0.001$] and industry supported (OR 0.10, 95% CI 0.02–0.38, $P < 0.001$) trials were significantly less likely to report OS as the primary end point.

Figure 1. Identification of randomized controlled trials of systemic therapy in breast, colorectal, and non-small-cell lung cancer published 2004–2009.

| Initial search results | N= 285 |
| Did not meet inclusion criteria: | N= 42 |
| • Prevention trials = 21 |
| • Not systemic therapy = 21 |
| RCTs meeting inclusion criteria | N= 243 |
| Excluded: | N= 106 |
| • Second report of same trial n= 5 |
| • Phase 2/3 trial n= 3 |
| • Data from multiple trials n= 7 |
| • Studies of dose/sequence/route n= 28 |
| • No efficacy results presented n= 7 |
| • Secondary analyses n= 45 |
| • Study of radiotherapy n= 11 |

| Study Cohort of RCTs | N= 137 |
| Date of publication, disease site, intervention, and study setting were recorded. Country of study origin was assigned based on the institutional affiliation of the first author. Consistent with our previous work [1], source of financial support was determined based on explicit statements in the article. In the absence of an explicit statement regarding funding support, RCTs with author(s) whose affiliation was a pharmaceutical company were classified as industry supported. We captured use of biomarkers in each trial. A biomarker was defined as ‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention’ [2]. Biomarkers were further classified as integral if the tests used were to determine eligibility, stratification, or allocation and integrated if they tested hypotheses [3]. The primary end point of each study was identified; if there was no explicit statement, the end point implied to be of primary importance was recorded. We evaluated use of intention-to-treat analysis based on what was stated in the manuscript. |
Although palliative trials were more likely than adjuvant trials to use OS, this association did not persist after controlling for disease site. Neither cooperative group status nor use of targeted agents was associated with use of OS before or after adjustment for other factors (all $P > 0.15$).

**targeted agents and biomarkers**

Among the 40 trials involving TT, the following agents were used: bevacizumab ($n = 8$), trastuzumab ($n = 7$), cetuximab ($n = 6$), gefitinib ($n = 5$), erlotinib ($n = 3$), lapatinib ($n = 3$), panitumumab ($n = 2$), bexarotene ($n = 2$), prinomastat ($n = 1$), doxefquidar ($n = 1$), edrecolomab ($n = 1$), aprinocarsen ($n = 1$), and BMS-275291 (a metalloproteinase inhibitor, $n = 1$). Compared with RCTs of traditional systemic therapy, studies of targeted agents were more likely to be conducted in the palliative setting [88%, (35/40) versus 43% (42/97), $P < 0.001$], involve patients with NSCLC [43% (17/40) versus 23% (22/97), $P = 0.023$], and be supported by industry [95% (38/40) versus 71% (69/97), $P = 0.001$]. They were also more likely to incorporate biomarkers in eligibility criteria [43% (17/40) versus 20% (19/97), $P = 0.010$].

Biomarkers were included in 58% (80/137) of RCTs [63% (25/40) of TT trials versus 57% (55/97) of non-TT trials, $P = 0.57$]. Integral biomarkers were included in 36% (49/137) of RCTs; biomarkers were used in eligibility, stratification, and treatment allocation in 36, 21, and 0 RCTs, respectively. Among the 49 RCTs using integral biomarkers, 40 (82%) used HER2 and/or ER/PR status in studies of breast cancer. Integrated biomarkers were explored in subset analyses in 67 (49%) RCTs. Specific biomarkers and their incorporation into trial design/analysis are shown in Table 3.

**discussion**

In this study, we have explored trends in study design, end points, and use of biomarkers among a cohort of RCTs published in five high-impact journals in the molecular era of oncology. Several important findings have emerged. First, compared with the preceding decade, there was a substantial increase in the number and size of RCTs reported. Modern RCTs are more likely to be international and multicentre in scope which likely accounts for the increased rate of patient accrual. Second, a greater proportion of systemic therapy RCTs are now evaluating TT and incorporating biomarkers into study design and analysis. Third, while biomarkers may be increasingly used, the most common use of integral biomarkers relates to traditional use of ER/PR and evolving use of HER2 in breast cancer RCTs. Fourth, while contemporary RCTs are more likely to have a time-to-event primary end point compared with the previous decade, there is increasing use of end points other than OS. Finally, industry involvement is increasingly common and now provides financial support for the vast majority of trials.

These trends in study design, setting, and intervention may influence how RCTs are interpreted by clinicians and policy makers. Larger sample size and the resultant increase in statistical power allow investigators to detect increasingly smaller differences between treatment arms. Additionally,

---

**Table 1.** Design of oncology RCTs 2005–2009 and 1995–2004

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Median sample size</td>
<td>722 (137)</td>
<td>446 (167)</td>
</tr>
<tr>
<td>Median time for accrual (mos)</td>
<td>35.5 (127)</td>
<td>33 (153)</td>
</tr>
<tr>
<td>Median follow-up (mos)</td>
<td>37 (104)</td>
<td>47 (105)</td>
</tr>
<tr>
<td>Interventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any chemotherapy</td>
<td>115 (84%)</td>
<td>124 (74%)</td>
</tr>
<tr>
<td>Any placebo/observation</td>
<td>19 (14%)</td>
<td>25 (15%)</td>
</tr>
<tr>
<td>Any hormonal agent</td>
<td>23 (17%)</td>
<td>18 (11%)</td>
</tr>
<tr>
<td>Any targeted agent</td>
<td>40 (29%)</td>
<td>7 (4%)</td>
</tr>
<tr>
<td>Intention-to-treat analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any ITT analysis</td>
<td>133 (97%)</td>
<td>155 (93%)</td>
</tr>
<tr>
<td>Primary end point</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-to-event end points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS</td>
<td>123 (90%)</td>
<td>125 (75%)</td>
</tr>
<tr>
<td>DFS</td>
<td>50 (36%)</td>
<td>82 (49%)</td>
</tr>
<tr>
<td>RFS</td>
<td>26 (19%)</td>
<td>16 (10%)</td>
</tr>
<tr>
<td>TTP</td>
<td>9 (7%)</td>
<td>8 (5%)</td>
</tr>
<tr>
<td>PFS</td>
<td>10 (7%)</td>
<td>19 (11%)</td>
</tr>
<tr>
<td>EFS</td>
<td>25 (18%)</td>
<td>N/A</td>
</tr>
<tr>
<td>RR</td>
<td>3 (2%)</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>8 (6%)</td>
<td>23 (14%)</td>
</tr>
<tr>
<td>Source of RCT funding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td>43 (31%)</td>
<td>51 (31%)</td>
</tr>
<tr>
<td>Industry</td>
<td>107 (78%)</td>
<td>95 (57%)</td>
</tr>
<tr>
<td>Foundation</td>
<td>17 (12%)</td>
<td>24 (14%)</td>
</tr>
<tr>
<td>Cooperative group</td>
<td>75 (55%)</td>
<td>75 (45%)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1%)</td>
<td>2 (1%)</td>
</tr>
</tbody>
</table>

*Number of studies for which these data were available is shown in parentheses.

*Data shown are for studies in which a primary end point was either explicitly identified or implied in the article. Time-to-events end points include overall survival (OS), disease-free survival (DFS), relapse-free survival (RFS), time to progression (TTP), progression-free survival (PFS), and event-free survival (EFS).

*Use of PFS and EFS was not captured in the previously reported study (Booth et al. [1]).

RCT, randomized controlled trial; NSCLC, non-small-cell lung cancer; CRC, colorectal cancer; ITT, intention-to-treat; RR, response rate; mos, months; N/A, not available.
RCTs without TT (n = growth factor receptor; VEGF, vascular endothelial growth factor. RCT, randomized controlled trial; TT, targeted therapy; EGFR, epidermal

points

Other time-to-event end points

- OS = 50 (5A, 2P)
- 12 (6A, 6P)
- 31 (3A, 27P, 1N)

was observed in our earlier work (Figure 2)[1]. This trend is

industry funding is consistent with trends since the 1970s that

having more trials with statistically significant results being

increasing use of primary end points other than OS may risk

have more trials with statistically significant results being

The increasing proportion of RCTs supported through

industry funding is consistent with trends since the 1970s that

was observed in our earlier work (Figure 2) [1]. This trend is

not surprising given our study criteria limiting trial eligibility to

those testing systemic therapy and an evolution from the

previous emphasis of hypotheses that tested already-approved

agents according to various doses (e.g. the dose intensity

hypothesis) to current prioritization of hypotheses associated

with altering a molecular target. The expense of these agents

[6], regulatory, and data management complexities associated

with using clinical trials data for new drug licensing

submissions [7] and perceptions that industry might best meet

these regulatory requirements [8] may contribute to the high

proportion of trials being conducted or otherwise supported by

industry. Systematic risks associated with reductions in

academically led clinical trials have been described [9] and

include impediments in addressing objectives associated with

comparative effectiveness research. Our findings further inform

the need for broad discussions that address the future of cancer

clinical trials [10].

Despite important advances in cancer therapy over the past

several decades, many patients would still define a useful

therapy as one that improves the quantity of survival and/or

leads to improved quality of life. However, there may be

substantial limitations to using OS as a primary trial end point,

including the need for sufficient sample sizes and lengthy

follow-up, and the confounding effects of subsequent lines of

therapy. Surrogate end points can be very helpful in addressing

these limitations but must first be validated by satisfying

statistical criteria [11–15]. Within our study cohort, the vast

majority of NSCLC trials used OS as the primary end point.

This is consistent with the lack of a validated surrogate end

point associated with OS for this disease [16]. Conversely, the

greater use of disease-free survival (DFS) and progression-free

survival (PFS) (in the adjuvant and palliative settings,

respectively) as primary end points in RCTs of colorectal is

supported by evidence showing these end points to be valid

surrogates for OS in this setting [17–20]. Although the validity

of surrogate end points in breast cancer is less certain, 88% of

the 59 breast cancer RCTs with a time-to-event primary end

point used end points other than OS. While improvements in

DFS led to regulatory approval of aromatase inhibitors in early

stage breast cancer, the relationship between DFS and OS in

this setting has not been definitively established [11, 21].

Likewise, the frequent use of PFS and TTP as surrogate end

points for survival in the advanced setting is not yet supported

by the existing literature [22–25].

In the molecular era, there is growing interest in biomarkers.

Overall, 58% (80/137) of trials in our cohort used biomarkers.

However, the vast majority of biomarker use relates to

traditional use of HER2 and ER/PR status in breast cancer. This

suggests that despite growing interest in personalized medicine,

biomarkers related to novel TT are not yet being used in

a substantial way in oncology RCTs.

An example of the important role biomarkers may play was

provided in the development of cetuximab for advanced CRC.

The pivotal RCT enrolled patients with chemo-refractory CRC

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Eligibility criteria</th>
<th>Stratification factor</th>
<th>Subset analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCTs with TT (n = 40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>7</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>ER/PR</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>HER2</td>
<td>9</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>K-ras</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>VEGF</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>RCTs without TT (n = 97)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER/PR</td>
<td>18</td>
<td>13</td>
<td>34</td>
</tr>
<tr>
<td>HER2</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>CEA</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Ras mutation</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>RRM1 and ERCC1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

RCT, randomized controlled trial; TT, targeted therapy; EGFR, epidermal growth factor receptor; VEGF, vascular endothelial growth factor.
expressing immunohistochemically detectable epidermal growth factor receptor (EGFR). Subjects were randomized to cetuximab or best supportive care. While there was a statistically significant but clinically modest 1.5 month improvement in median survival among the complete population [26], an exploratory biomarker analysis revealed that patients with wild-type KRAS tumors derived a 4.7 month improvement in median survival while there was no detectable benefit among patients with mutant KRAS [27]. This exploratory biomarker analysis had a direct influence on clinical care as most authorities now state that patients with mutant KRAS should not be treated with EGFR inhibitors [28].

This report builds on previous work by our group in this field [1]. Our results should be interpreted in the context of study limitations. A potential weakness is that by including only RCTs of systemic therapy in breast cancer, NSCLC, and CRC, our findings may not be generalizable to other disease sites. Also, by limiting our search to five journals, we did not capture every RCT published during the study period. Publication bias has been well described [29] and we recognize that our cohort of trials does not represent the entire body of RCTs in oncology. However, we were most interested in methodology, funding source, and outcomes of practice-changing RCTs; a high proportion of which are published in the journals we included. Furthermore, we chose these journals to allow for a comparison of our results with our previously reported overview of oncology RCTs that were published 1975–2004. Finally, although our results describe trends in design and results of RCTs published 2005–2009, it remains uncertain whether our findings extend to RCTs that are currently accruing patients in the molecular era.

In summary, the size and number of systemic therapy RCTs published in major journals continue to increase over time as does the focus on TT. Industry has an increasing role in the funding of oncology RCTs. Investigators, especially those studying targeted agents, should consider how to make better use of biomarkers to identify those patients who have the greatest probability of responding to therapy. Attempts should continue to improve methodology of RCTs in oncology with particular emphasis on using primary end points that are clinically meaningful and valid. Consideration of aspects of design of oncology RCTs will ensure clinical trials of the future have the greatest impact on improving patient outcomes.

acknowledgements
Dr Booth is supported as a Cancer Care Ontario Chair in Health Services Research.

disclosure
The authors declare no conflict of interest.

references


9. DeMets DL., Califf RM. A historical perspective on clinical trials innovation and leadership: where have the academics gone? JAMA 2011; 305(7): 713–714.


11. Gill S, Sargent D. End points for adjuvant therapy trials: has the time come to accept disease-free survival as a surrogate end point for overall survival? Oncologist 2006; 11(6): 624–629.


